From the Institute of Experimental Dermatology
of the University of Lübeck
Director: Prof. Dr. Hauke Busch

Multi-Omics Analysis Uncovers Split Formation and Keratinocyte Detachment as Key Drivers of Long-Lasting Cellular Effects in Pemphigus Vulgaris

Dissertation
for Fulfillment of
Requirements
for the Doctoral Degree
of the University of Lübeck

from the Department of Computer Sciences and Engineering or
from the Department of Natural Sciences

Submitted by

Sen Guo
from Henan, China

Lübeck 2025

First referee: Professor Dr. rer. nat. Hauke Busch

Second referee: Professor Dr. rer. nat. Karsten Seeger

Date of oral examination: 25/07/2025

Approved for printing. Lübeck, 11/08/2025

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne die Benutzung anderer als der angegebenen Hilfsmittel selbständig verfasst habe; die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe des Literaturzitats gekennzeichnet.

Lübeck, 09.10.2024

Abstract

Pemphigus Vulgaris (PV) is an autoimmune disease in which antibodies mistakenly target the adhesion proteins Desmoglein 1 (DSG1) and/or Desmoglein 3 (DSG3) on skin cells, leading to the loss of cell-cell adhesion and causing blistering. While the molecular changes following antibody binding in PV remain poorly understood, we investigated these downstream effects by analyzing gene and protein responses in two experimental models: human primary epidermal keratinocytes (HPEKs) and human skin organ culture (HSOC).

Samples were treated with either PX43, a human-derived antibody fragment targeting DSG1 and DSG3, or AK23, a mouse-derived DSG3 antibody, alongside control treatments. In the HPEK model, PV antibody treatments did not trigger notable changes compared to controls at 5, 10, or 24 hours. However, in the HSOC model, only PX43 induced tissue splitting and significant changes in gene and protein expression, particularly in pathways linked to inflammation and immune signaling (e.g., TNFα, Interferon α/γ, IL2-STAT5, IL5-STAT3).

Importantly, these molecular changes resembled those seen in wounded or inflamed skin, suggesting that physical damage from blister formation—not the antibody binding itself—is the primary driver of downstream cellular responses. This study reveals that tissue injury may be the main trigger for disease progression in PV, pointing toward new therapeutic targets that focus on modulating the wound response and inflammation, rather than the antibodies alone.

Zusammenfassung

Pemphigus Vulgaris (PV) ist eine Autoimmunerkrankung, bei der Antikörper fälschlicherweise die Adhäsionsproteine Desmoglein 1 (DSG1) und/oder Desmoglein 3 (DSG3) auf Hautzellen angreifen, was zum Verlust der Zell-Zell-Adhäsion führt und Blasenbildung verursacht. Während die molekularen Veränderungen nach der Bindung von Antikörpern in der PV nach wie vor nur unzureichend verstanden werden, haben wir diese nachgeschalteten Effekte durch die Analyse von Gen- und Proteinreaktionen in zwei experimentellen Modellen untersucht: menschliche primäre epidermale Keratinozyten (HPEKs) und menschliche Hautorgane (HSOC).

Die Proben wurden entweder mit PX43, einem vom Menschen stammenden Antikörperfragment, das auf DSG1 und DSG3 abzielt, oder mit AK23, einem von der Maus stammenden DSG3-Antikörper, sowie mit Kontrollbehandlungen behandelt. Im HPEK-Modell lösten die PV-Antikörperbehandlungen im Vergleich zu den Kontrollen nach 5, 10 oder 24 Stunden keine nennenswerten Veränderungen aus. Im HSOC-Modell führte jedoch nur PX43 zu einer Aufspaltung des Gewebes und zu signifikanten Veränderungen in der Gen- und Proteinexpression, insbesondere in Signalwegen, die mit Entzündungen und Immunsignalen verbunden sind (z. B. TNFα, Interferon α/γ, IL2-STAT5, IL5-STAT3).

Wichtig ist, dass diese molekularen Veränderungen denen ähnelten, die bei verletzter oder entzündeter Haut zu beobachten sind, was darauf hindeutet, dass die physische Schädigung durch die Blasenbildung - und nicht die Antikörperbindung selbst - der primäre Auslöser für die nachgeschalteten zellulären Reaktionen ist. Diese Studie zeigt, dass die Gewebeschädigung der Hauptauslöser für das Fortschreiten der Erkrankung bei PV sein kann, was auf neue therapeutische Ziele hindeutet, die sich auf die Modulation der Wundreaktion und der Entzündung konzentrieren und nicht auf die Antikörper allein.

Table of Contents

Abbreviations

AAbs	Autoantibodies
ACOD1	Aconitate Decarboxylase 1
AD	Activation Domain
ADAMTS4	ADAM Metallopeptidase with Thrombospondin Type 1 Motif 4
ADGRG3	Adhesion G Protein-Coupled Receptor G3
ADORA3	Adenosine A3 Receptor
AGR2	Anterior Gradient 2
AK23	Anti-Desmoglein 3
APOB	Apolipoprotein B
ARHGEF38	Rho Guanine Nucleotide Exchange Factor 38
ARL11	ADP Ribosylation Factor Like GTPase 11
ATP	Adenosine triphosphate
BAMBI	BMP And Activin Membrane Bound Inhibitor
BDKRB1	Bradykinin Receptor B1
BHLHE40	Basic Helix-Loop-Helix Family Member E40
BIVM-ERCC5	BIVM-ERCC5 readthrough
BME	β-mercaptoethanol
BP	Bullous Pemphigoid
CAAR-T	Chimeric Autoantibody Receptor T cell
CAD	Cutaneous Autoimmune Disease
CAND2	Cullin Associated and Neddylation Dissociated 2
CCDC187	Coiled-Coil Domain Containing 187
CCL3	C-C motif chemokine Ligand 3
CCL3L3	C-C motif Chemokine Ligand 3 Like 3
CCL3L3	C-C Motif Chemokine Ligand 3 Like 3
CCL4	C-C motif chemokine Ligand 4
CCL4	C-C motif chemokine ligand 4
CCL4L2	C-C Motif Chemokine Ligand 4 Like 2
CEBPB	CCAAT/enhancer-binding Protein Beta
CFAP44	Cilia and Flagella Associated Protein 44
CFAP73	Cilia and Flagella Associated Protein 73
CFB	Complement Factor B
CHAD	Chondroadherin
CMPK2	Cytidine/uridine Monophosphate Kinase 2
COL22A1	Collagen type XXII Alpha 1 chain
CPB2-AS1	CPB2 antisense RNA 1
CSF3	Colony Stimulating Factor 3
CXCL10	C-X-C Motif Chemokine Ligand 10
CXCL11	C-X-C Motif Chemokine Ligand 11
CXCL2	C-X-C motif chemokine Ligand 2

CXCL3	C-X-C motif chemokine Ligand 3
CXCL8	C-X-C motif chemokine Ligand 8
CYP1A1	Cytochrome P450 family 1 subfamily A member 1
CYP1A1	Cytochrome P450 family 1 subfamily A member 1
DAPs	Differentially Abundant Proteins
DBD	DNA-binding Domain
DC	Dendritic Cells
DEGs	Differentially Expressed Genes
DMSO	Dimethyl Sulfoxide
DNM1	Dynamin 1
DSG1	Desmoglein 1
DSG3	Desmoglein 3
DSP-AS1	DSP Antisense RNA 1
E2F4	E2F transcription factor 4
EBF1	Early B-cell Factor 1
EGFR	Epidermal Growth Factor Receptor
ELF3	E74 Like ETS Transcripiton Factor 3
ELISA	Enzyme Linked Immunosorbent Assay
ERICH6-AS1	ERICH6 antisense RNA 1
EtOH	Ethanol
FABP3	Fatty Acid Binding Protein 3
FABP7	Fatty Acid Binding Protein 7
FCGR1A	Fc Gamma Receptor Ia
FDA	Food and Drug Administration
FDR	False Discovery Rate
FOS	Fos proto-oncogene
FOSB	FosB Proto-Oncogene
FOXL2	Forkhead Box L2
FOXM1	Forkhead box protein M1
FOXP1	Forhead Box P1
FZD2	Frizzled Class Receptor 2
G0S2	G0/G1 switch 2
GABRB3	Gamma-Aminobutyric Acid type A Receptor Subunit beta3
GABRB3	Gamma-Aminobutyric Acid Type A Receptor Subunit Beta3
GATA6	GATA binding protein 6
GDF15	Growth Differentiation Factor 15
GEM	Gel Bead-in-emulsion
GLM	Generalized Linear Model
GO	Gene Ontology
HaCaT	Human Adult low Calcium Temperature
HEV	High Endothelial Venules
hIgG	Human Immunoglobulin G

HLA	Human Leukocyte Antigen
HMCN1	Hemicentin 1
HPEK	Human primary epidermal keratinocyte
HPLC	High Performance Liquid Chromatography
HSOC	Human skin organ culture
IFIT1	Interferon induced protein with Tetratricopeptide repeats 1
IFIT2	Interferon Induced Protein with Tetratricopeptide Repeats 2
IFIT3	Interferon Induced Protein with Tetratricopeptide Repeats 3
IFNB1	Interferon Beta 1
IGF1	Insulin like Growth Factor 1
IGFBP5	Insulin-like Growth Factor Binding Protein 44
IGHV3-33	Immunoglobulin Heavy Variable 3-33
IL2	Interleukin 2
IMAC	Immobilized Meta Affinity Chromatography
IRF1	Interferon Regulatory Factor 1
ISG15	ISG15 ubiquitin like modifier
IVIG	Intravenous Immunoglobulin
KEGG	Kyoto Encyclopedia of Genes and Genomes
KLF15	KLF Transcription Factor 15
KLK6	Kallikrein Related Peptidase 6
LAL	Limulus Amebocyte Lysate
LBHD1	LBH Domain Containing 1
log2FC	Log2 Fold-Change
LPS	Lipopolysaccharide
LRT	Likelihood Ratio Test
LRT	Likelihood Ratio Test
MAGED4B	MAGE Family Member D4B
MAP2K6	Mitogen-Activated Protein Kinase Kinase 6
MEF2C	Myocyte Enhancer Factor 2C
MHC	Major Histocompatibility Complex
mIgG	Mouse Immunoglobulin
MoR	Mode of Regulation
MPV17L	MPV17 Mitochondrial Inner Membrane Protein Like
MS	Mass Spectrometry
MTARC1	Mitochondrial Amidoxime Reducing Component 1
MYC	MYC proto-oncogene
NB	Negative Binomial
NEGR1	Neuronal Growth Regulator 1
NFATC1	Nuclear Factor of Activated T-cells, Cytoplasmic 1
NFKB1	Nuclear Factor Kappa B subunit 1
NINJ1	Ninjurin 1
NKD1	NKD inhibitor of WNT signaling pathway 1

NPIPB6	Nuclear Pore Complex Interacting Protein Family Member B6
NR1H2	Nuclear Receptor Subfamily 1 Group H Member 2
OASL	2'-5'-Oligoadenylate Synthetase Like
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCMTD1-DT	PCMTD1 Divergent Transcript
PD	Population doublings
PDAI	Pemphigus Disease Area Index
PEDS1-UBE2V1	PEDS1-UBE2V1 Readthrough
PF	Pemphigus Foliaceus
PKC	Protein Kinase C
PPARA	Peroxisome Proliferator Activated Receptor Alpha
PPDPFL	Pancreatic Progenitor Cell Differentiation and Proliferation Factor Like
PSMC1P1	Proteasome 26S subunit, ATPase 1 Pseudogene 1
PV	Pemphigus Vulgaris
RCOR3	REST Corepressor 3
REL	REL proto-oncogene
RELA	RELA proto-oncogene
RLT	RNA Lysis Buffer
RSAD2	Radical S-adenosyl methionine domain containing 2
RSAD2	Radical S-adenosyl methionine Domain containing 2
S100A8	S100 calcium binding protein A8
S100A9	S100 calcium binding protein A9
SAA2	Serum Amyloid A2
scFv	Single-chain variable fragment
SCX	Strong Cation Exchange
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SERPINB4	Serpin Family B Member 4
SINHCAFP3	SINHCAF pseudogene 3
SLC28A3	Solute Carrier Family 28 Member 3
SMIM11	Small Integral Membrane Protein 11
SNORD3D	Small Nucleolar RNA, C/D box 3D
SP1	Specific Protein 1
SRF	Serum Response Factor
SSD	Optional Signal-sensing Domain
STAT3	Signal Transducer and Activator of Transcription 3
STAT5	Signal Transducer and Activator of Transcription 5
TF	Transcription Factor
TLS	Tertiary Lymphoid Structure
TMEM265	Transmembrane Protein 265
TMIGD3	Transmembrane and Immunoglobulin Domain containing 3

TNFα	Tumor necrosis factor
ULM	Univariate Linear Model
UMAP	Uniform Manifold Approximation and Projection
UNC13D	Unc-13 homolog D
VSIG4	V-set and Immunoglobulin domain containing 4
VST	Variance Stabilizing Transformation
WDR17	WD repeat domain 17
ZIC2	Zic Family Member 2

1. Introduction and Background

1.1 Autoimmune Disease – A General Overview

Autoimmune disease is a condition that results from an abnormal response of the adaptive immune system, which mounted against self-antigens, targets and attacks healthy, functioning parts of the body as if they were foreign organisms, thus causing tissue damage [1]. It can be organ-specific like Type I Diabetes which affects the pancreatic islets and systemic like Systemic Lupus Erythematosus which can cause inflammation in the skin, kidneys and brain [2,3]. Susceptibility to autoimmune disease is controlled by environmental and genetic factors [4]. More than 100 recognized autoimmune diseases have been identified, according to the Autoimmune Disease List from the Global Autoimmune Institute website [5]. Various body parts can be affected [6].

Components of the Immune System

The immune system is a complex network of biological processes that defend an organism against diseases. It identifies and reacts to various pathogens, including viruses, parasitic worms, cancer cells, and foreign objects like wood splinters, while differentiating them from the organism's own healthy tissues. In many species, the immune system is divided into two primary subsystems. The innate immune system offers a rapid, pre-programmed response to general threats, while the adaptive immune system develops a customized response by learning to recognize specific molecules from previous encounters. The innate immune system is the body's first line of defense against infections and harmful agents, providing a rapid and non-specific response. It is evolutionarily ancient and present in all multicellular organisms [7]. Unlike the adaptive immune system, it does not have memory and does not improve its response upon repeated exposure to the same pathogen. Both subsystems rely on various molecules and cells to carry out their roles [8].

The immune system defends its host from infections through increasingly specific layers of protection. Physical barriers act as the first defense, blocking pathogens like bacteria and viruses from entering the body. If these barriers are breached, the innate immune system provides an immediate but generalized response. This innate system is found in all animals [9]. In vertebrates, if the pathogen manages to evade the innate defense, a second layer of

protection, the adaptive immune system, is triggered by the innate response. During an infection, the adaptive immune system fine-tunes its response to better recognize the invading pathogen. Once the pathogen is eliminated, this enhanced response is stored as immunological memory, enabling the system to mount faster and stronger defences during future encounters with the same pathogen [10,11].

Both innate and adaptive immunity rely on the immune system's ability to differentiate between self and non-self molecules. In immunology, self molecules refer to the components of an organism's body that the immune system recognizes as its own. In contrast, non-self molecules are identified as foreign. A specific category of non-self molecules is called antigens—named for their role in generating antibodies—which are defined as substances that bind to specific immune receptors and trigger an immune response. Table 1 below provides a summary of the two types of immune systems.

Table 1: Components of the immune system

Innate immune system	Adaptive immune system
Response is non-specific	Pathogen and antigen specific response
Exposure leads to immediate maximal response	Lag time between exposure and maximal response
Cell-mediated and humoral components	Cell-mediated and humoral components
No immunological memory	Exposure leads to immunological memory
Found in nearly all forms of life	Found only in jawed vertebrates

Adaptive Immune System

The adaptive immune system first evolved in early vertebrates, enabling a more robust immune response and the development of immunological memory, where each pathogen is "remembered" by its unique antigen [12]. This system's response is antigen-specific, requiring the recognition of distinct "non-self" antigens through a process known as antigen presentation. Antigen specificity allows the immune system to create responses precisely targeted to specific pathogens or infected cells. The ability to generate these targeted responses is preserved by "memory cells" in the body. If the same pathogen infects the body again, these memory cells rapidly respond to eliminate it.

The adaptive immune system consists of specialized white blood cells known as lymphocytes-primarily B and T cells, derived from hematopoietic stem cells in the bone marrow [13]. B cells

are responsible for the humoral immune response, whereas T cells are involved in the cell-mediated immune response.

T cells can be further categorized into different types based on their functions. Cytotoxic T cells (CD8+) recognize antigens bound to Class I MHC molecules, while helper ($CD4^+$) and regulatory T cells ($CD4^+$, $FoxP3^+$) identify antigens presented by Class II MHC molecules. This distinction in antigen presentation corresponds to the various roles of T cell types. Additionally, a third, less common type, γδ T cells, can recognize antigens independently of MHC binding [14].

During thymus development, double-positive T cells encounter self-antigens, and iodine plays a crucial role in thymus development and function [15]. In contrast, B cells possess antigen-specific receptors in the form of antibodies on their surface. These antibodies can directly recognize native (unprocessed) antigens, which may include large molecules on pathogens or smaller haptens, such as penicillin, when attached to carrier molecules.

Each lineage of B cells produces a unique antibody, representing the diverse range of antibodies the body can generate. When B or T cells encounter their specific antigens, they undergo rapid proliferation, creating numerous clones that target the same antigen. This phenomenon is known as clonal selection.

Once activated, killer T cells seek out cells displaying the antigen and release cytotoxins, such as perforin, which form pores in the membranes of target cells, leading to apoptosis [16]. This process is vital for halting viral replication and is tightly regulated, often requiring strong activation signals, sometimes facilitated by helper T cells. Figure 1 below illustrates a schematic representation of the adaptive immune system, detailing its key components and interactions.

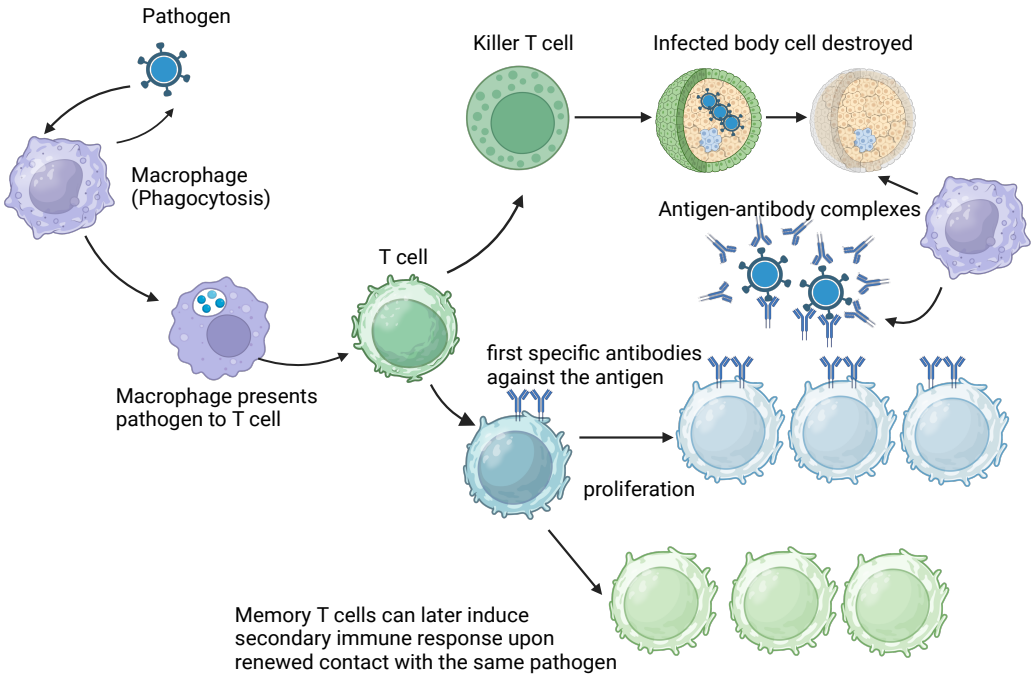


Figure 1: Summary of the adaptive immune response. The adaptive immune response begins when antigen-presenting cells, such as macrophages, display pathogen fragments to T lymphocytes, activating cytotoxic T cells to kill infected cells and stimulating B lymphocytes to produce specific antibodies. This coordinated response not only eliminates the pathogen but also generates memory cells that ensure a faster and more efficient response during future infections with the same pathogen. The figure is created by biorender.com.

In the case of Pemphigus Vulgaris (PV), an organ-specific autoimmune skin disorder, the adaptive immune system generates autoantibodies against desmogleins—key proteins involved in cell-cell adhesion in the epidermis. This targeted immune response disrupts epidermal integrity and leads to characteristic blistering, making PV a prototypical example of antibody-mediated autoimmunity.

1.2 Pemphigus Vulgaris: Pathophysiology and Clinical Significance

Pemphigus Vulgaris is a rare, chronic blistering skin disorder and the most prevalent form of pemphigus. The term "pemphigus" is derived from the Greek word *pemphix*, which means "blister" [17]. This condition is classified as a type II hypersensitivity reaction, characterized by the production of antibodies against desmosomes, structures within the skin that help maintain adhesion between certain skin layers. As these desmosomes are targeted by the immune response, the layers of skin begin to separate, resulting in a clinical presentation that resembles blister formation (Figure 2). The blisters arise from acantholysis, which involves the

disruption of intercellular connections mediated by autoantibodies[18]. Without treatment, the disease progressively worsens over time, leading to larger and more widespread lesions across the body that functionally mimic the effects of a severe burn.

Before the development of modern treatments, the mortality rate for pemphigus vulgaris was nearly 90%. However, with the introduction of corticosteroids as the primary treatment, the current mortality rate has decreased to approximately 5% to 15% [19]. Despite this improvement, pemphigus vulgaris was ranked as the fourth leading cause of death associated with skin disorders in 1998. Consequently, it remains classified as a potentially fatal condition. The disease predominantly impacts middle-aged and older adults, particularly those aged between 50 and 60 years. Historically, there has been a higher incidence in women [20].

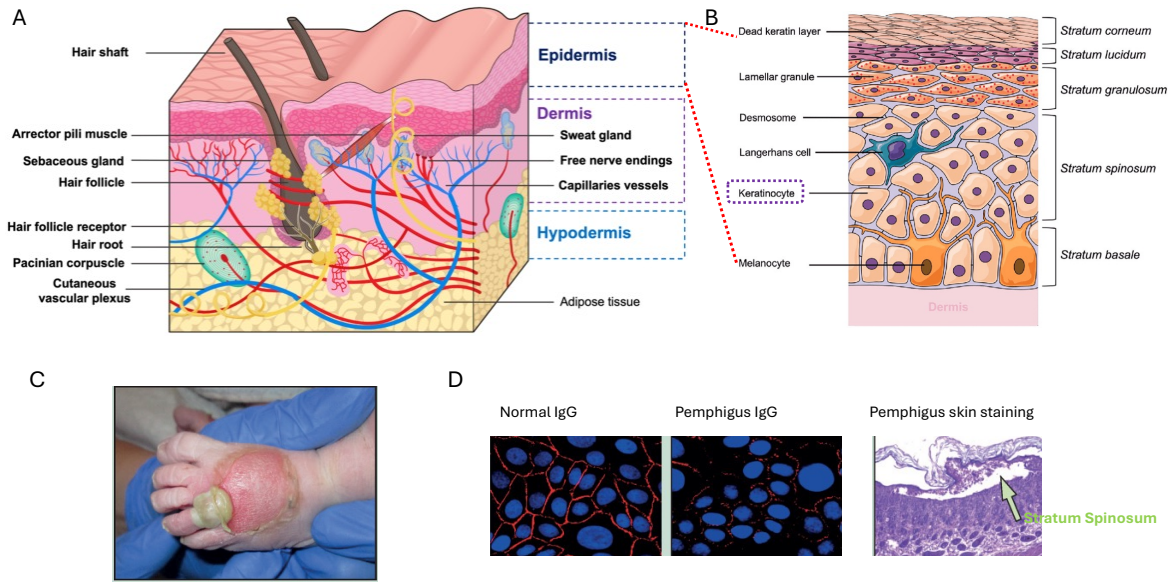


Figure 2: Schematic view of skin layers and the blister in Pemphigus. (A) Human skin is composed of three primary layers: the epidermis, the dermis, and the hypodermis (subcutaneous layer), which is rich in adipose tissue. Each layer serves distinct structural and functional roles, with the epidermis providing a protective barrier, the dermis housing connective tissue, blood vessels, and sensory receptors, and the hypodermis acting as insulation and energy storage. Figure from Ramadon et al. (2022) [21] (B) Keratinocytes, the predominant cell type in the epidermis, are the primary targets in pemphigus vulgaris, where autoantibodies disrupt their adhesion, leading to characteristic blistering. Figure from Ramadon et al. (2022) [21] (C) Maternal autoantibodies from mothers with pemphigus can cross the placenta and lead to temporary blistering in newborns. Figure from Schmidt et al. (2019) [22] (D) Degradation of desmosomes (red stain) occurs in cultured keratinocytes after incubation with IgG from patients; separation occurs within the epidermis, primarily in the stratum spinosum. Figure from Schmidt et al. (2019) [22].

Blisters in Pemphigus Vulgaris

Pemphigus Vulgaris primarily manifests as oral blisters, particularly affecting the buccal and palatine mucosa, though cutaneous blisters are also common. Other mucosal surfaces,

including the conjunctiva, nasal passages, esophagus, and genitals (penis, vulva, vagina, cervix), as well as the anus, may also be involved. Flaccid blisters typically appear on the skin, with the palms and soles often spared. These blisters frequently erode, leading to ulcerated lesions and erosions. A hallmark of the disease is a positive Nikolsky sign, where blistering is induced either on normal-appearing skin or at the periphery of an existing blister. In severe cases, significant pain during chewing can result in weight loss and malnutrition [23].

Desmoglein 3 (DSG3) and Desmoglein 1 (DSG1)

Desmosomes are specialized cellular structures responsible for cell-to-cell adhesion. As a type of junctional complex, they appear as spot-like adhesions distributed randomly along the lateral sides of plasma membranes. Known for their strong adhesive properties, desmosomes are commonly found in tissues subjected to significant mechanical stress, such as cardiac muscle, bladder tissue, gastrointestinal mucosa, and epithelial layers [24].

Desmoglein 3 (DSG3) and desmoglein 1 (DSG1) are calcium-binding transmembrane glycoproteins, key components of desmosomes in vertebrate epithelial cells. Although both DSG1 and DSG3 belong to the same desmosomal cadherin family, they have distinct expression patterns and functional roles. DSG1 is primarily found in the superficial layers of the epidermis, contributing to skin barrier function and the strength of the outermost layers. In contrast, DSG3 is more abundantly expressed in the basal and lower suprabasal layers (Figure 3), where it helps maintain the integrity and cohesion of deeper cell layers [25]. Their differential expression is crucial for epithelial tissue stratification and proper function.

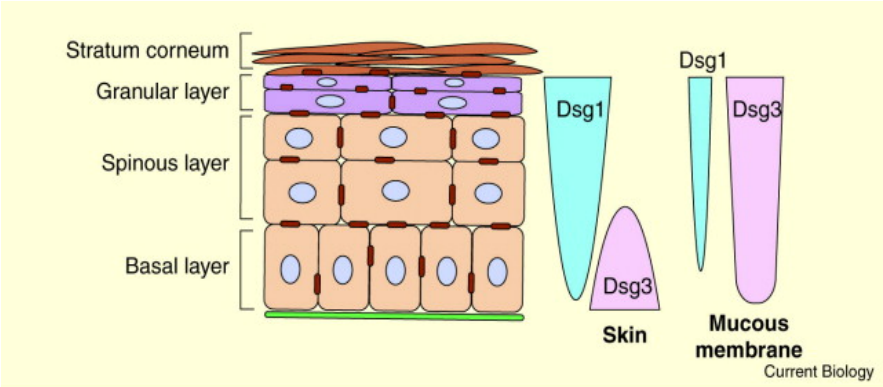


Figure 3: The distribution of DSG1 and DSG3 in the skin and mucous membrane. In the skin, DSG3 is localized in the deeper basal and immediate suprabasal layers, while DSG1 is predominantly present in the more differentiated upper layers. Conversely, in the oral mucosa, both DSG1 and DSG3 are expressed across all layers. Figure 3 from Ishii et al. (2001) [25].

Desmosomes are dynamic structures that adapt to environmental and mechanical stress, particularly in tissues like the skin and heart. Beyond their adhesive function, they regulate key processes such as proliferation, differentiation, inflammation, and barrier integrity. Desmosomal proteins interact with signaling pathways like EGFR, IGF1, and Hippo, both influencing and being regulated by these pathways. Their dysfunction is implicated in skin and heart diseases, highlighting their role as signaling hubs that modulate cellular processes, including wound healing and tissue regeneration [26].

The Unresolved Mechanisms Behind Pemphigus Vulgaris: A Continuing Debate

The very early hypothesis proposes that the anti-DSG3 antibody-dependent steric hindrance interferes with intercellular adhesive functions of DSG3, leading directly to desmosomal dissociation [27], however *Yasuo et al.* reported that no inhibition of Ca^{2+}-induced desmosome formation by PV-IgG binding to surface PV-antigens, suggesting that PV-IgG does not directly inhibit desmosome formation, even though the antibodies may cause steric hindrance between homophilic DGS3 interactions [28], thus this hypothesis is limited that steric blockade alone can not explain acantholysis.

Another hypothesis is that the PV-IgG-induced intracellular signaling events could lead to desmosomal dissociation, PV-IgG induces phosphorylation of DSG3, leading to dissociation of DSG3 from plakoglobin in cultured keratinocytes [29]. Other reports also show that apoptosis signaling is involved in PV [30], and p38 MAPK signaling pathways [31]. However, this hypothesis is also limited that Energy-dependent signaling does not account for immediate intercellular splits seen at non-desmosomal regions, and protease inhibition fails to fully prevent acantholysis, suggesting signaling is necessary but not sufficient [32].

Later Yasuo et al. (2023) using a keratinocyte dissociation assay, demonstrated that PV-IgG depletes desmosomes of DSG3 by approximately 80% in cultured keratinocytes, resulting in a 40% reduction in adhesive strength. Based on these findings, they proposed that pemphigus represents a desmosome-remodeling disease, where autoantibodies targeting DSG3 induce its depletion through endocytosis [33,34]. This process, regulated by protein kinase C (PKC), compromises cell adhesion and leads to epidermal blistering (Figure 4). However, initiates endocytosis remains unclear whether it is direct antibody crosslinking or downstream signaling and in vivo evidence for this sequence is still lacking.

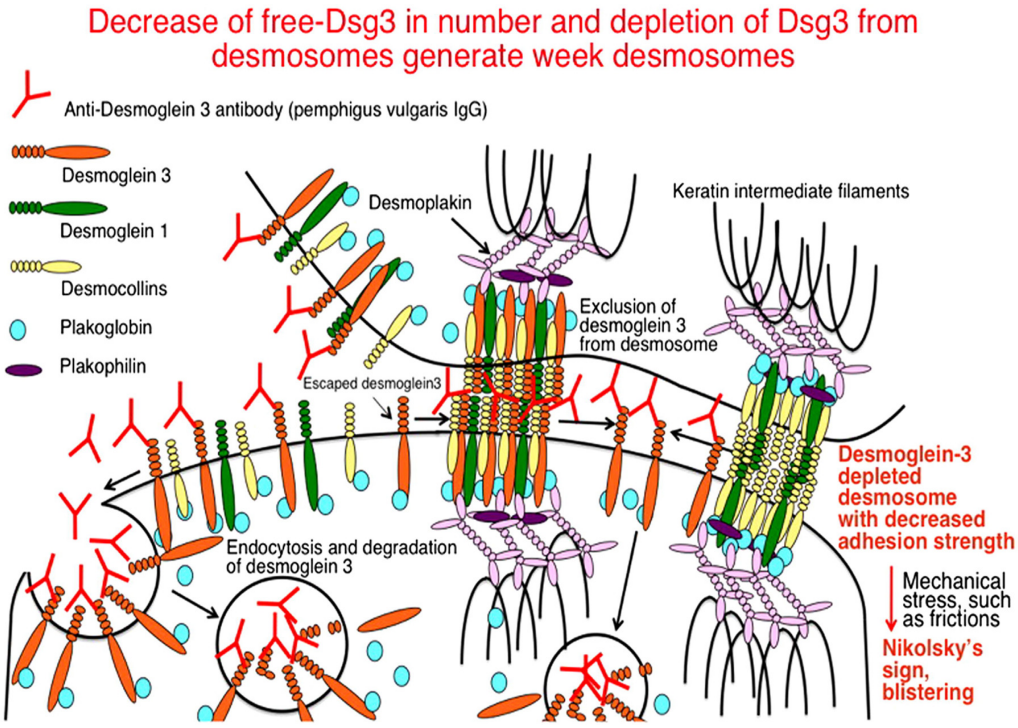


Figure 4: Proposed mechanisms for Dsg3 depletion in desmosomes. The binding of autoantibodies to the DSG3 leading to weakened adhesion and blistering in pemphigus vulgaris (PV), involves the binding of anti-Dsg3 antibodies (present in PV-IgG) to free Dsg3 proteins on the cell surface. This triggers the internalization of Dsg3 into endosomes. Additionally, Dsg3 proteins within the core of desmosomes are excluded and may also undergo endocytosis. The resulting shortage of Dsg3 on the cell membrane causes the formation of Dsg3-depleted desmosomes, reducing their adhesion strength. These processes are likely regulated by multiple signaling pathways. Figure 4 from Kitajima Yasuo (2013) [33].

The Dysregulated Autoimmune Response Hypothesis, supported by multiple studies, suggests that immune system dysfunction—particularly involving T cells, B cells, and cytokine imbalances—plays a significant role in disease development. This perspective extends beyond the direct effects of autoantibodies or intracellular signaling pathways, highlighting the role of systemic immune dysregulation in pemphigus pathogenesis [35–38]. To be noted, cytokine blockade alone does not fully abrogate blistering, and complement-deficient models still develop acantholysis—implying immune factors modulate, but do not drive, the core adhesion loss [39].

Despite significant progress, the exact mechanisms underlying pemphigus vulgaris (PV) remain incompletely understood. Further studies are needed to untangle the multiple interrelated processes contributing to disease development.

Diagnosis of Pemphigus Vulgaris

Due to its rarity, the diagnosis of pemphigus vulgaris is often challenging and can be delayed. Early symptoms may include erosions in the mouth or blisters on the skin, which can cause itching or pain. Theoretically, these blisters exhibit a positive Nikolsky's sign, where slight rubbing causes the skin to slough off, this sign is not always consistently reliable. The gold standard for diagnosis is a punch biopsy from the perilesional area, which is examined using direct immunofluorescence. This method typically reveals acantholysis, the loss of intercellular connections between keratinocytes. Acantholytic cells, which can also be observed using a Tzanck smear, are rounded, nucleated keratinocytes. Their formation is a result of antibody-mediated damage to the desmosomal protein desmoglein, disrupting cell adhesion.

Treatment of Pemphigus Vulgaris

Corticosteroids and other immunosuppressive drugs have traditionally been used to manage pemphigus symptoms. However, due to the significant and long-term side effects associated with steroid use, their duration and dosage should be minimized. Other treatments, such as intravenous immunoglobulin (IVIG), mycophenolate mofetil, methotrexate, azathioprine, and cyclophosphamide, have also been utilized with varying levels of effectiveness.

Monoclonal antibodies, such as rituximab, have emerged as a well-established alternative to corticosteroids and are increasingly utilized as a first-line treatment for pemphigus. In the summer of 2018, the Food and Drug Administration (FDA) granted full approval for rituximab for this indication following a successful fast-track evaluation [40]. Multiple case series have demonstrated that many patients can achieve remission after just one cycle of rituximab, with earlier treatment—possibly even at the time of diagnosis—associated with greater success. Furthermore, combining rituximab with monthly intravenous immunoglobulin (IVIG) infusions has led to long-term remission, with no recurrence of disease observed in patients even 10 years after discontinuation of treatment. This finding is based on a small trial involving 11 patients, of whom 10 were followed to completion [41].

However, current treatments remain suboptimal: corticosteroids are frequently limited by severe adverse effects, and biologics such as rituximab show variable efficacy and incomplete durability of response. In this context, our study’s identification of novel molecular targets

may pave the way for more effective—and better tolerated—therapeutic strategies in clinical practice.

1.3 Overview of Bulk RNA Sequencing, Single-Cell RNA Sequencing, and Proteomics: Techniques and Key Analytical Approaches

Advancements in omics technologies have significantly enhanced our understanding of autoimmune diseases by enabling comprehensive analyses of molecular changes at various biological levels. These technologies allow researchers to investigate gene expression, protein composition, and post-translational modifications, shedding light on disease mechanisms and potential therapeutic targets. The following subsections introduce key omics approaches—bulk RNA sequencing, shotgun proteomics, phosphoproteomics, and single-cell RNA sequencing—that are widely used in autoimmune disease research.

1.3.1 Bulk RNA-seq

RNA-seq is a next-generation sequencing method used to quantify RNA molecules in a biological sample, offering insights into gene expression (the transcriptome) [42,43]. It enables the detection of alternative splicing, post-transcriptional modifications, gene fusions, mutations/SNPs, and differential expression across conditions or over time [44]. RNA-Seq can detect various RNA types, including the total RNA, miRNA, and ribosomal RNA [45].

Before the advent of RNA-seq, gene expression analysis relied on microarrays. However, limitations like poor quantification and cross-hybridization artifacts [46].

Advanced applications of RNA-seq include single-cell sequencing, 3' mRNA-seq, and real-time native RNA sequencing [47,48]. Additionally, emerging uses—such as detecting copy number variations, microbial contamination, and neoantigens—are facilitated by advancements in bioinformatics [49]. A summary of the RNA-seq is shown in Figure 5.

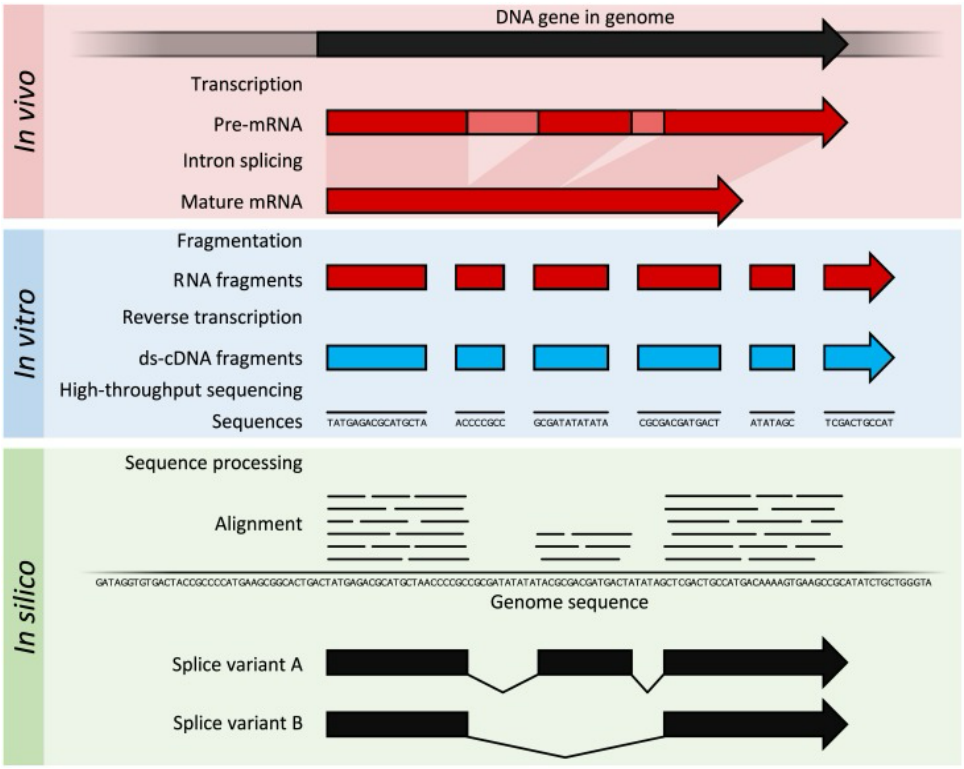


Figure 5: Summary of RNA-seq. In eukaryotic organisms, genes are transcribed and spliced to generate mature mRNA transcripts (depicted in red). The mRNA is then extracted, fragmented, and reverse-transcribed into stable double-stranded complementary DNA (ds-cDNA), represented in blue. This ds-cDNA undergoes high-throughput short-read sequencing. The resulting sequences are aligned to a reference genome, enabling the reconstruction of the transcribed regions. This analysis provides insights into gene expression localization, quantifies relative expression levels, and identifies alternative splice variants. Figure is extracted from Lowe et al. (2017) [50].

As illustrated in Figure 6, the workflow for bulk RNA-seq data analysis encompasses several critical steps: processing and quantification of raw sequencing reads to derive gene expression values for each sample, dimensionality reduction for exploratory study to assess sample clustering, identification of differentially expressed genes (DEGs) with associated log2 fold-change and statistical significance metrics, including p-values and adjusted p-values. Functional annotation of enriched pathways is performed using databases such as HALLMARK, KEGG, GO, and Reactome, with HALLMARK selected for representative results due to its non-redundancy and widespread use. Additionally, transcription factor (TF) activation is predicted to elucidate regulatory mechanisms governing gene expression.

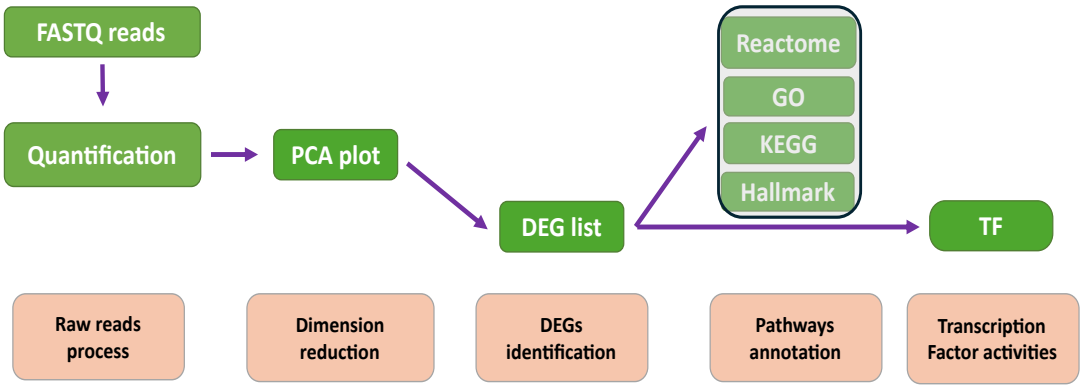


Figure 6: Summary of bulk RNA-seq data analysis. The analysis of the bulk RNA-seq data mainly includes the steps of the raw reads process, dimension reduction, DEGs identification, pathways annotation, and prediction of transcription activities.

Kallisto is an RNA-seq quantification tool that significantly accelerates processing times by up to two orders of magnitude while maintaining accuracy comparable to traditional methods [51]. This efficiency is achieved through a novel approach called pseudo alignment, which rapidly determines the set of transcripts compatible with each read without performing the base-level alignment. The detailed methodology is illustrated in Supplementary Figure 16 and explained in Supplementary Method 7.3.1.

Table 2 presents a representative excerpt from the abundance.tsv output file. The first column lists the transcript names, while the subsequent columns provide the transcript length, effective length, estimated read counts, and TPM (Transcripts Per Million) values.

Table 2: Representative output of transcript abundance quantification by kallisto

Target_id	Length	Eff_length	Est_counts	TPM
Transcript 1	xxx	xxx	xxx	xxx
…	xxx	xxx	xxx	xxx
Transcript N	xxx	xxx	xxx	xxx

The transcript-level quantification results from Kallisto were subsequently imported and aggregated to gene-level expression values using the tximport package. This conversion process leverages transcript-to-gene mapping annotations to summarize transcript abundance estimates, such as TPM (Transcripts Per Million) or estimated counts, into corresponding gene-level metrics. The tximport tool ensures accurate aggregation by accounting for transcript length and effective length differences, providing gene-level expression values suitable for downstream differential expression analysis and other transcriptomics workflows.

Principal Component Analysis

Principal Component Analysis (PCA) is a fundamental linear dimensionality reduction technique widely used in data analysis and machine learning. It transforms data into a new coordinate system to identify directions, known as principal components, that capture the largest variations in the dataset. These principal components are orthogonal unit vectors forming an orthonormal basis, ensuring the transformed dimensions are linearly uncorrelated.

The primary objective of PCA is to sequentially identify these principal components such that the variance explained by each successive component decreases while maintaining orthogonality. PCA simplifies complex datasets by fitting a dimensional ellipsoid to the data, where each axis of the ellipsoid corresponds to a principal component. The axis lengths reflect variance: shorter axes represent directions of lower variance and lesser significance in

describing the data's structure. This transformation facilitates exploratory data analysis, particularly for visualizing patterns or clusters by projecting data onto the first few principal components.

The proportion of variance explained by each principal component is determined by dividing its eigenvalue by the sum of all eigenvalues. This metric guides the selection of the most significant components. Visualizing data in two or three dimensions using the top principal components often reveals clusters or patterns that are challenging to discern in the original high-dimensional space.

An example of PCA plot can be seen in Figure 7.

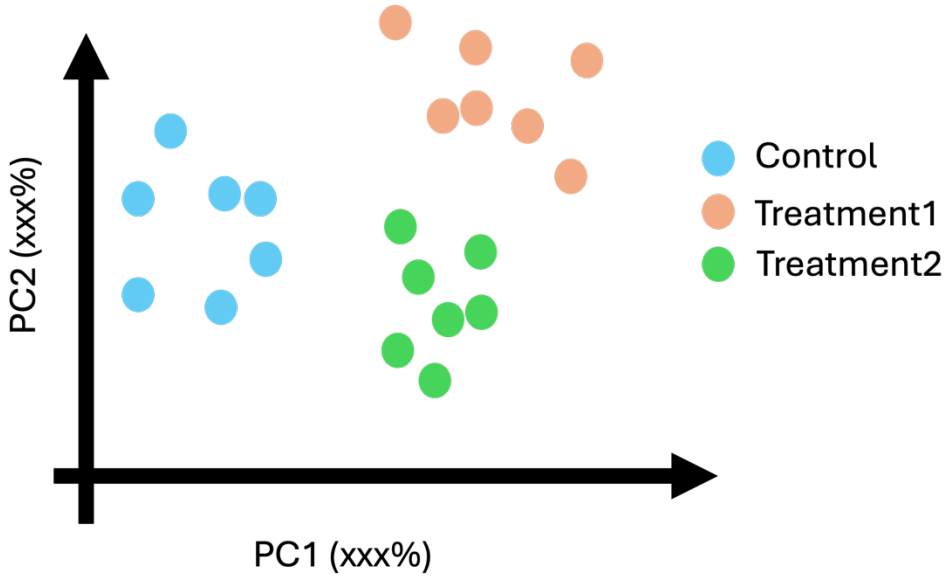


Figure 7: An example PCA plot. The PCA plot visualizes the variance captured by the first two principal components (PC1 and PC2), with the percentage values indicating the proportion of total variance explained by each component. Samples are color-coded based on their respective groups, allowing for the identification of clustering patterns and providing an initial assessment of differences among groups.

In RNA-seq data analysis, principal component analysis (PCA) is commonly used to visualize the variance in gene expression across samples. A PCA plot can reveal the degree of separation between treatment and control groups, providing an initial assessment of overall transcriptional differences. The step-by-step process for performing PCA in detailed in Supplementary Method 7.3.2.

Differentially Expressed Genes

Identifying differentially expressed genes (DEGs) is fundamental in understanding the molecular mechanisms underlying biological processes and disease conditions. By comparing gene expression profiles across different conditions, treatments, or time points, researchers can uncover genes that are activated or suppressed in response to specific stimuli. Statistical methods are essential in this process, as they enable the quantification of gene expression levels and the identification of significant changes in expression, accounting for biological variation and technical noise. These analyses provide valuable insights into the functional roles of genes, facilitating the discovery of potential biomarkers and therapeutic targets.

Differential gene expression analysis involves the application of statistical methods to identify changes in gene expression levels between experimental groups, using replicated samples for accurate estimation. As illustrated in Figure 8, the comparison between two groups (Group A and Group B) is based on the relative positioning of their mean expression values with respect to the global mean expression. When the means of Group A and Group B are close to the global mean, no significant difference in gene expression is observed. However, when the means of Group A and Group B exhibit substantial deviations from the global mean, a significant difference in gene expression between the two groups is likely to be present.

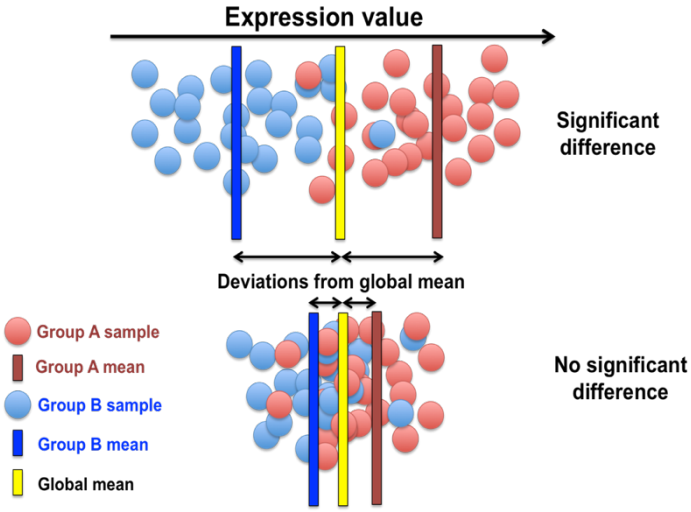


Figure 8: Example for comparing expression levels between groups. The figure is extracted from the online tutorial by Sarah Bonnin (2020) [52].

Gene expression data derived from RNA-seq experiments typically consist of count data, where the number of reads mapped to each gene reflects the gene's expression level. A key challenge in analyzing such data is accounting for both the mean expression and the variability (or overdispersion) often observed in RNA-seq data, where the variance exceeds the mean.

Differentially expressed genes were quantified using a likelihood ratio test (LRT) implemented in DESeq2 [53]. This approach is based on a generalized linear model (GLM) with a negative binomial (NB) distribution, which is well-suited for modeling RNA-seq data due to its ability to account for overdispersion—where the variance exceeds the mean. By applying this statistical framework, we can more accurately detect genes with significant expression changes across conditions, providing insights into key biological mechanisms underlying the studied phenotype. GLM is detailed in the Supplementary Method 7.3.3.

In DESeq2, a negative binomial model is applied to estimate gene-specific dispersion values, with shrinkage techniques employed to improve the estimation, particularly for low-expression genes. These adjustments lead to more stable and accurate dispersion estimates across all genes. The primary goal is to identify differentially expressed genes (DEGs) by comparing models that include treatment effects to those that do not. This comparison uses likelihood ratio tests (LRT), where the null hypothesis assumes no treatment effect, and the alternative hypothesis suggests a treatment effect. The LRT then quantifies the significance of gene expression changes due to the treatment, enabling reliable identification of DEGs.

The primary steps in DESeq2 are illustrated in Figure 9, adapted from online tutorial by Mistry et al. (2019) [54].

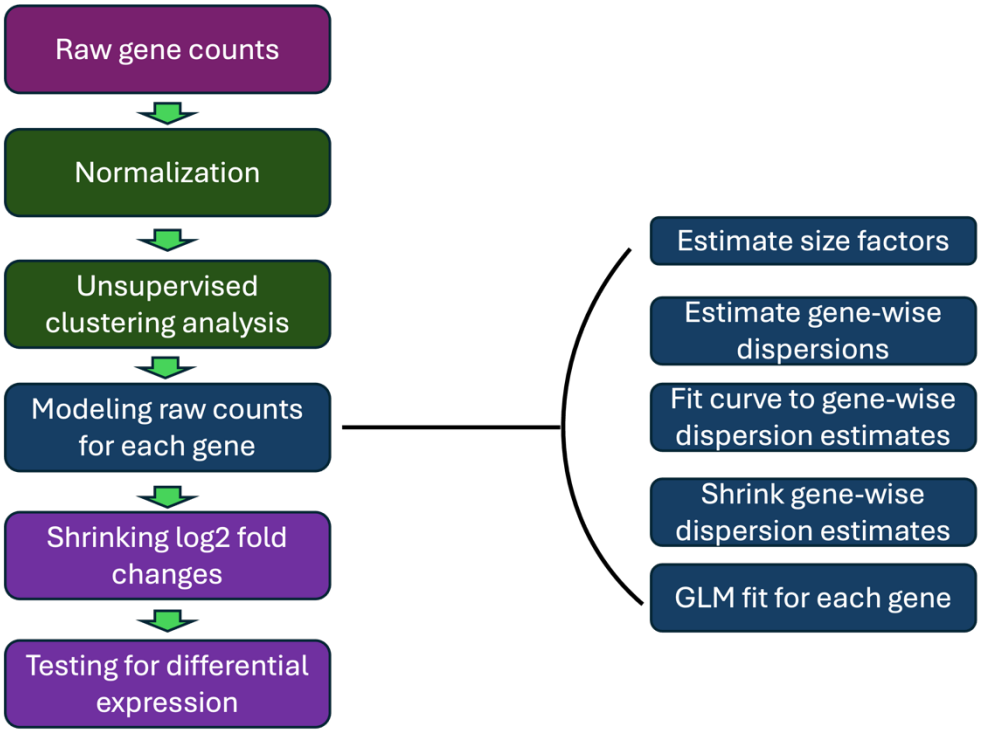


Figure 9: Main steps in DESeq2.The differential gene expression analysis workflow begins with raw gene counts and includes normalization, unsupervised clustering, and modeling raw counts for each gene. Subsequent steps involve shrinking log2 fold changes and testing for differential expression. Key computational steps include estimating size factors, estimating gene-wise dispersions, fitting a curve to dispersion estimates, shrinking these estimates, and performing a generalized linear model (GLM) fit for each gene. The Figure is adapted from the online tutorial by Mistry et al. (2019) [54].

Table 3 shows a representative output of the differentially expressed genes (DEGs) from DESeq2, which includes columns for Gene, BaseMean, Log2FoldChange, LfcSE, Stat, Pvalue, and Padj.

- Gene: Names of genes.
- BaseMean: The average expression level of a gene across all samples, indicating its baseline activity. Higher values suggest higher expression.
- Log2FoldChange: The log-transformed ratio of gene expression between two conditions (e.g., treated vs. control). Positive values indicate upregulation, negative values downregulation, and 0 indicates no change.
- LfcSE (Log2 Fold Change Standard Error): The standard error of the Log2FoldChange estimate, reflecting its reliability. Smaller values indicate more reliable estimates.
- Stat (Statistic): A statistical measure (e.g., t-statistic) used to test whether the observed Log2FoldChange is significantly different from zero. Larger values suggest stronger evidence of differential expression.
- Pvalue: The probability that the observed results occurred by chance under the null hypothesis (no differential expression). P-values below 0.05 indicate statistical significance.
- Padj (Adjusted P-value): The p-value adjusted for multiple comparisons to reduce the risk of false positives. A Padj below 0.05 indicates statistical significance after correction.

Table 3: Representative output of DEGs from DESeq2

Gene	BaseMean	Log2FoldChange	LfcSE	Stat	Pvalue	Padj
Gene 1	xxx	xxx	xxx	xxx	xxx	xxx
...	xxx	xxx	xxx	xxx	xxx	xxx
Gene N	xxx	xxx	xxx	xxx	xxx	xxx

The DEGs can be visualized using methods such as heatmaps or volcano plots, as demonstrated in Figure 10.

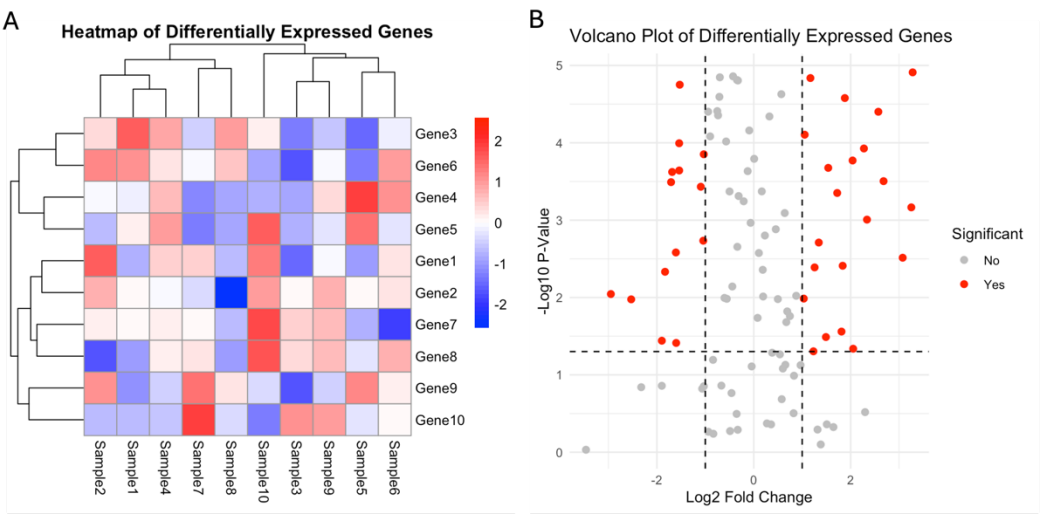


Figure 10: Example of heatmap and volcano plots for displaying differentially expressed genes. The left heatmap plot displays the top differentially expressed genes that meet a specified threshold, while the right volcano plot illustrates all genes, with significant ones highlighted in color.

After identifying differentially expressed genes, it becomes more feasible to associate them with specific phenotypes and identify key biomarkers correlated with the disease.

Pathway enrichment analysis

Pathway enrichment analysis enables researchers to uncover mechanistic insights from gene lists derived from omics experiments including RNA-seq. This approach identifies biological pathways overrepresented in a gene list compared to what would be expected by chance.

Large gene lists often necessitate extensive manual literature review, making their interpretation impractical. Pathway enrichment analysis is a commonly employed solution to this challenge, which condenses the extensive gene list into a more concise and interpretable set of biological pathways.

Pathway enrichment analysis can be conducted using three main approaches, as illustrated in Figure 11:

- Overrepresentation-based. This method requires a gene list of interest and tests whether any pathways are observed in this list more than expected by chance against a predefined background gene set [55].
- Functional scoring system (ranking)-based. Ranking-based methods consider functional information generated by different omics datasets, first ranks the total gene

set based on detected signals in omics studies, such as transcript abundance, then tests whether genes annotated to the same pathway tend to cluster together at the top (or bottom) of the ranked list [56].

- Pathway topology-based methods. Topology-based methods sim to account for additional information that impacts pathway activity by integrating scores measuring gene positions within a pathway and gene-gene interactions into the enrichment tests [57].

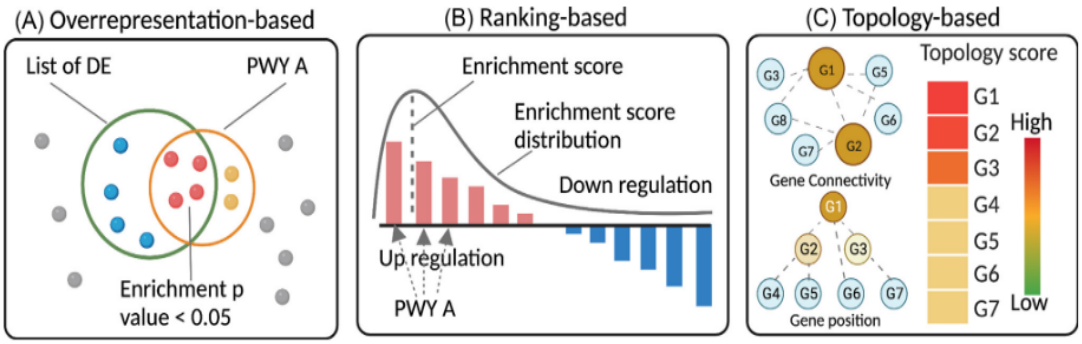


Figure 11: Overview of three types of methods for pathway enrichment analysis. (A) Over-representation-based methods examine whether any pathways are observed in a gene list of interest more than expected by chance compared with a background set. (B) Ranking-based methods first rank the total gene set based on detected signals, such as change of gene expression, then test whether genes annotated to the same pathway tend to cluster together at the top (or bottom) of the ranked list. (C) Topology-based methods integrate scores measuring gene positions within a pathway and gene-gene interactions into the enrichment tests. DE represents differentially expressed genes. PWY represents a pathway. G1 to G7 represents gene 1 to gene 7. Figure is extracted from the Zhao et al. (2023) [58].

Ranking-based methods are widely utilized, as over-representation-based approaches have several limitations, including the assumption of gene independence and the need for an arbitrary cutoff to define differentially expressed gene sets. In contrast, topology-based methods rely on experimental evidence for pathway structures and gene-gene interactions, which remain largely unavailable for many organisms.

GAGE employs a rank-based approach to enrichment analysis and incorporates methodological enhancements that make it particularly suitable for datasets with heterogeneous sample sizes and complex experimental designs [59]. It is recognized for its high consistency, sensitivity, specificity, and strong biological interpretability. Therefore, in this study, GAGE is used to do the pathway enrichment analysis. GAGE algorithm is detailed in Supplementary Figure 17 and Supplemental Method 7.3.4.

Pathway Reference Databases

Several pathway reference databases have been utilized, including HALLMARK, which represents well-defined biological states and processes; Gene Ontology (GO), which provides comprehensive gene annotations; Kyoto Encyclopedia of Genes and Genomes (KEGG), which describes metabolic and signaling pathways; and Reactome, a curated database of biological pathways that enables a detailed and multifaceted analysis of biological functions and pathways [60–63]. HALLMARK was ultimately selected due to its non-redundancy and widespread use.

Each HALLMARK gene set is a 'refined' collection derived from multiple 'founder' sets. It represents a specific biological state or process with coherent expression patterns. By summarizing key information from the original founder sets while minimizing variation and redundancy, the HALLMARK gene sets offer more precise and concise inputs for gene set enrichment analysis. The process of generating HALLMARK gene sets is illustrated in Figure 12.

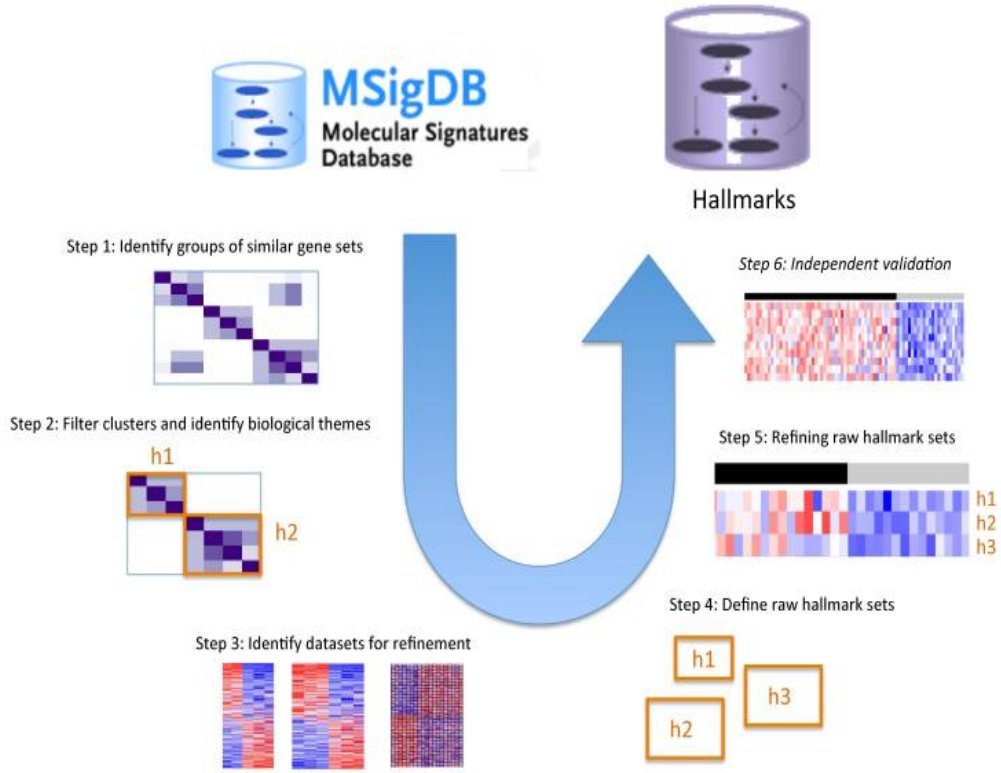


Figure 12: Process of generating HALLMARK gene sets.First, groups of similar gene sets are identified and clustered based on biological themes (Steps 1–2). Next, relevant datasets are selected for refinement (Step 3), and raw hallmark sets are defined (Step 4). These sets undergo further refinement (Step 5) before undergoing independent validation to ensure robustness and biological relevance (Step 6). The figure is extracted from Liberzon et al. (2015) [60].

A summary of the HALLMARK gene sets is provided in Supplementary Table 1. Including the name, process category, description, number of founder sets and number of gene it contains.

An example for the display of the top enriched pathways for the RNA-seq results can be seen in Figure 13.

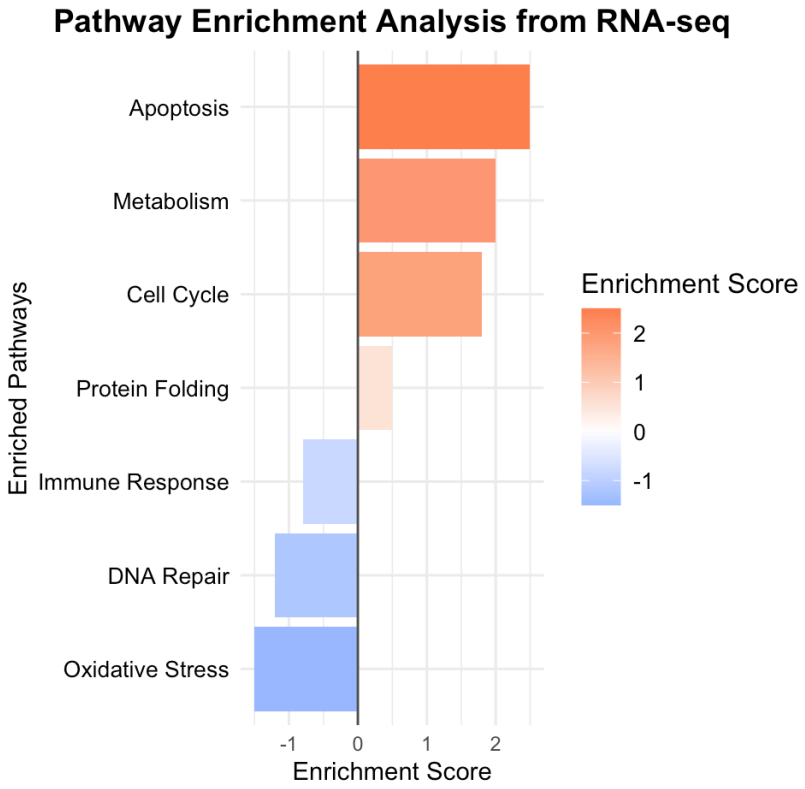


Figure 13: Example bar plot displaying the top enriched pathways form RNA-seq analysis. Pathway enrichment analysis can identify the most differentially regulated pathways across conditions, based on specified p-values or adjusted p-values.

Transcriptions Factors

A transcription factor (TF), also known as a sequence-specific DNA-binding factor, is a protein that regulates the transcription of genetic information from DNA to messenger RNA by binding to specific DNA sequences [64].

Transcription factors (TFs) regulate gene expression by activating or repressing specific genes, ensuring that they are expressed in the appropriate cells, at the correct times, and in the right quantities throughout both cellular and organismal life. TFs function in coordinated networks to govern fundamental cellular processes such as cell division, growth, and apoptosis, as well as more complex events such as cell migration and tissue organization during embryogenesis. Furthermore, TFs modulate gene expression in response to extracellular signals, including

hormones. The human genome encodes approximately 1,600 distinct transcription factors, many transcription factors exhibit tissue-specific expression [65].

Figure 14 provides an example of the domain architecture associated with transcription factor binding. Figure 15 provides a simplified representation of transcription factor (TF) activity and its regulation.

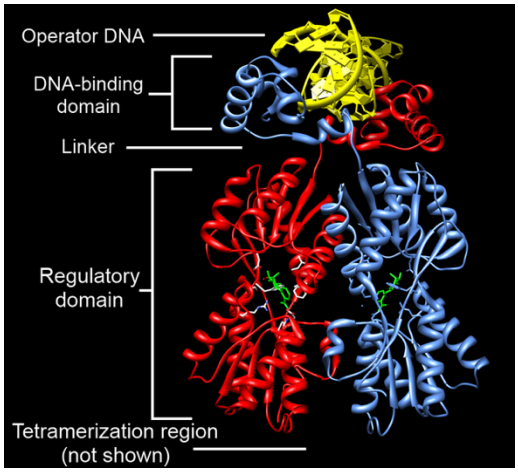


Figure 14: An example of domain architecture associated with transcription factor binding. Lactose Repressor (LacI): The N-terminal DNA-binding domain (labeled) of the lac repressor interacts with its target DNA sequence (gold) in the major groove through a helix-turn-helix motif. Binding of the effector molecule (green) in the regulatory domain (labeled) induces an allosteric response, facilitated by the linker region (labeled). The figure is extracted from the WIKIPEDIA that introduce the lac repressor [66].

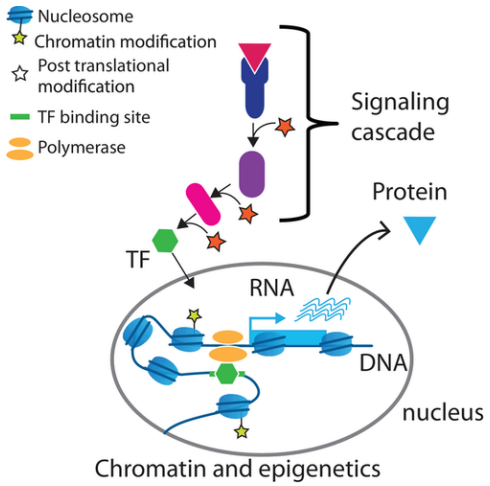


Figure 15: Simplified view of transcription factor (TF) activity and its regulation. TFs can be activated by signaling cascades, then bind to the DNA, where they can regulate transcription, resulting in altered RNA and protein expression. Figure is extracted from Weidemüller et al. (2021) [67].

The decoupleR package was utilized to infer transcription factor (TF) activities using $\log_2$ fold-change (log2FC) values of differentially expressed genes as input, the inference was based on

prior knowledge of transcription factor-target interactions, generating activity scores that reflect the regulatory influence of transcription factors within the dataset [68]. Algorithm of decoupleR is detailed in Supplementary Figure 18 and Supplementary Method 7.3.5.

CollecTRI is a comprehensive database that integrates a curated collection of transcription factors (TFs) and their transcriptional targets from 12 distinct resources [69]. This compilation provides a robust foundation for inferring TF activities, as demonstrated in Figure 16.

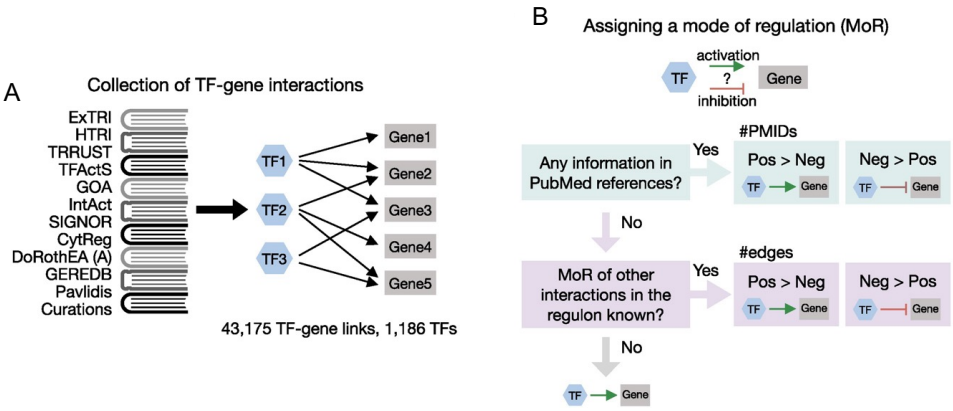


Figure 16: Description of transcription factor (TF)–gene interactions in the CollecTRI-derived regulons. (A) Collecting transcription factor (TF)–gene links to construct regulons from CollecTRI. Depicting prior knowledge resources used to collect links, which were aggregated within CollecTRI. (B) Flow chart describing how the mode of regulation (MoR) was assigned to each TF–gene link. The MoR, indicating the direction of transcriptional regulation from the TF to its target gene, was determined for each TF–gene link, based on factors such as PubMed references (PMIDs) and the MoR of other genes in the regulon. The figure is extracted from the Dott et al. (2023) [69].

The top differentially expressed transcription factors can also be displayed in the bar plot as shown in Figure 17.

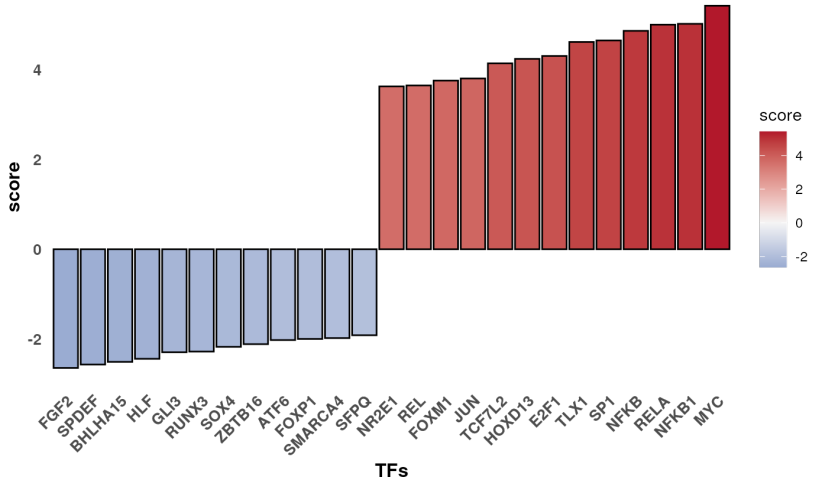


Figure 17: Example bar plot displaying the differentially expressed transcription factors. The top differentially expressed transcription factors, including both upregulated and downregulated ones, along with their predicted activity scores, can be visualized in a bar plot. The figure was extracted from the decoupleR manual website by Pau Mompel (2024) [70].

1.3.2 Single-Cell RNA-seq (scRNA-seq)

Single-cell RNA sequencing (scRNA-seq) enables the measurement of gene expression at the single-cell level, representing a transformative technology that significantly advances our understanding of biological processes [71]. It has been used to evaluate transcriptional variability within individual cells [72,73], across cell groups [74,75], and to identify novel cell types and states across tissues such as blood [76–78], spleen [79], brain [80–85], intestine [86,87], and pancreas [88]. Additionally, scRNA-seq enables the reconstruction of differentiation pathways through pseudotime trajectories, capturing cells at various stages of development [89–91]. This method also uncovers gene expression covariation within cells, using natural variations and engineered perturbations to elucidate molecular networks and mechanisms [92–94].

The single-cell RNA sequencing (scRNA-seq) process primarily involves sample input, encapsulation of single cells with barcoded beads in droplets, library construction, sequencing, and data analysis. Figure 18 presents an overview of the 10X Genomics Gel Bead-in-emulsion (GEM) technology, while details of GEM generation and barcoding are depicted in Figure 19. Figure 20 provides a GEM-X single-cell 3’ Gel Bead schematic diagram. All figures are sourced from the 10X Genomics website [95].

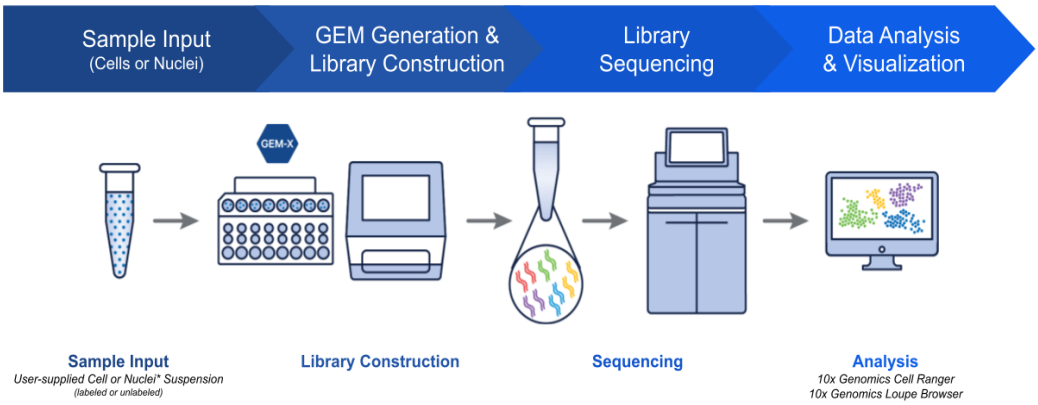


Figure 18: Schematic overview of the GEM-X technology workflow. A typical workflow begins with user-provided cells or nuclei, combined with GEM-X reagents and consumables to enable partitioning, cell lysis, and cellular barcoding. The resulting barcoded transcripts are subsequently amplified and converted into libraries compatible with Illumina or other short-read sequencing platforms. Following sequencing, data is processed with the Cell Ranger pipeline and visualized through the Loupe Browser or other tools. Figure is sourced from the 10X Genomics website by Natalya Ortolano (2024) [95].

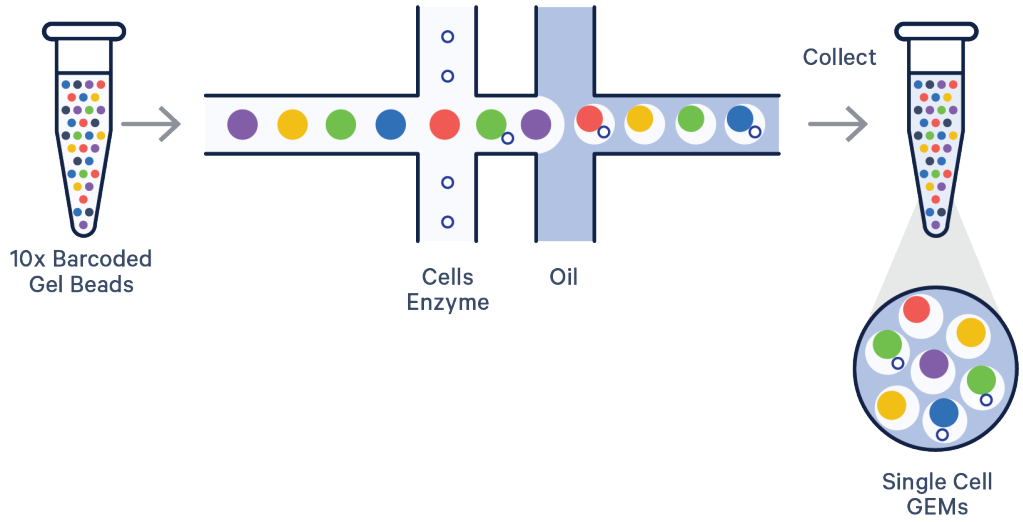


Figure 19: A schematic overview of GEM generation and barcoding with the GEM-X chip workflow. GEMs are created by mixing barcoded Gel Beads, a master mix with cells, and partitioning oil within a GEM-X 3' or 5' Chip. To achieve single-cell resolution, cells are introduced at a limiting dilution so that most (~90–99%) of the resulting GEMs are empty, while the rest predominantly contain a single cell. Figure is sourced from the 10X Genomics website by Natalya Ortolano (2024) [95].

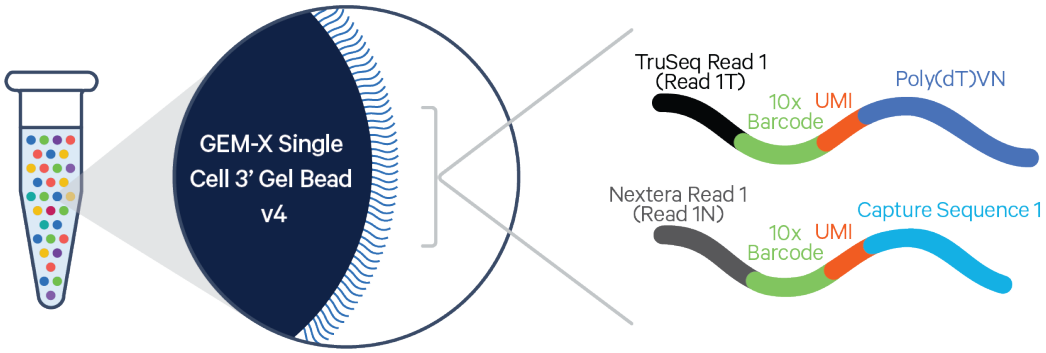


Figure 20: Schematic diagram of a GEM-X Single Cell 3' gel bead. Every Gel Bead is coated with oligos containing an Illumina TruSeq Read 1 (read 1 sequencing primer, Read 1T), 16 nt 10x Barcode, 12 nt unique molecular identifier (UMI), and 30 nt poly(dT)VN. Figure is sourced from the 10X Genomics website by Natalya Ortolano (2024) [95].

Single-cell gene expression measurement is poised to transform our understanding of gene regulation and address longstanding questions in biology. When cells are grouped based on their expression profiles, they cluster according to cell type or developmental stage, enabling unbiased classification and 'reverse engineering' of cell types within any population or tissue, provided that enough cells are sequenced (Figure 21) [96–100]. With extensive, unbiased

sampling, such clustering can reveal all existing cell types, including previously unidentified ones. Moreover, clusters can yield robust, data-driven expression profiles characteristic of each cell type, without requiring prior knowledge of marker genes specific to a tissue or cell type. Consequently, single-cell RNA profiling represents the first method capable of establishing a quantitative, data-driven framework for classifying cell types.

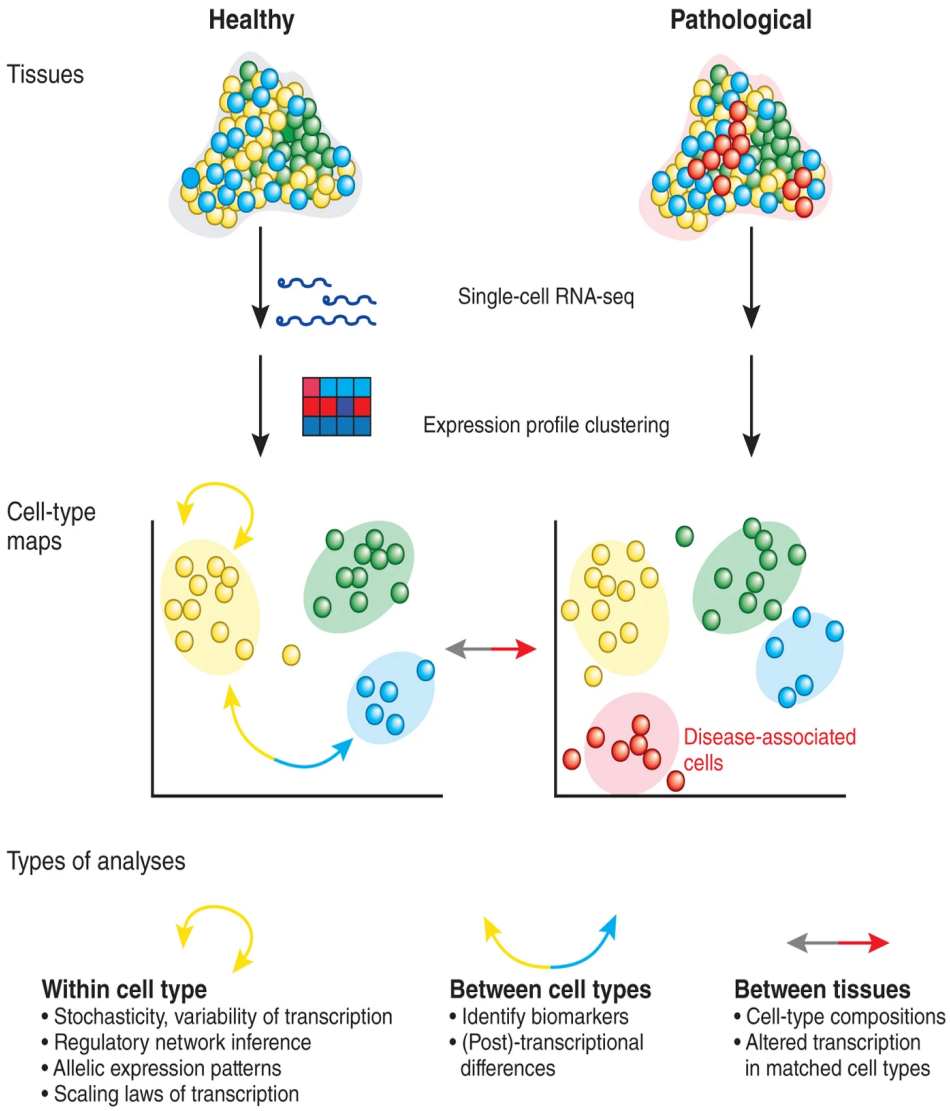


Figure 21: scRNA-seq analysis of tissues and cell types. Based on these expression profiles, cells are grouped into distinct clusters, generating "cell-type maps" that elucidate the cellular composition of each tissue. In healthy tissues, this clustering reflects the intrinsic diversity of cell types. In contrast, clustering in pathological tissues may reveal additional groups, including disease-associated cells with altered gene expression profiles. Several analytical approaches made possible by these cell-type maps: (1) Within-cell-type analysis, which enables the investigation of transcriptional variability, regulatory network inference, allelic expression patterns, and transcriptional scaling laws; (2) Between-cell-type analysis, which facilitates the identification of biomarkers and examination of transcriptional and post-transcriptional differences across cell types; and (3) Between-tissue analysis, which allows for comparisons of matched cell types between healthy and diseased tissues to detect changes in cell composition and transcriptional alterations linked to disease. Overall, this workflow offers a comprehensive framework for exploring cellular diversity and gene regulation in both healthy and pathological states. Figure 12 is extracted from the paper by Rickard Sandberg (2014) [100].

Single-Cell Gene Quantification by Cellranger, Cell Clustering and Differential Gene Expression Analysis with Seurat

Cell Ranger is employed for gene quantification in single-cell gene expression datasets [101]. The detail is illustrated in Supplementary Figure 19 and Supplementary Method 7.3.6.

A streamlined workflow for single-cell RNA sequencing (scRNA-seq) data analysis using Seurat is presented in Figure 22, with each step detailed in Supplementary Method 7.3.7.

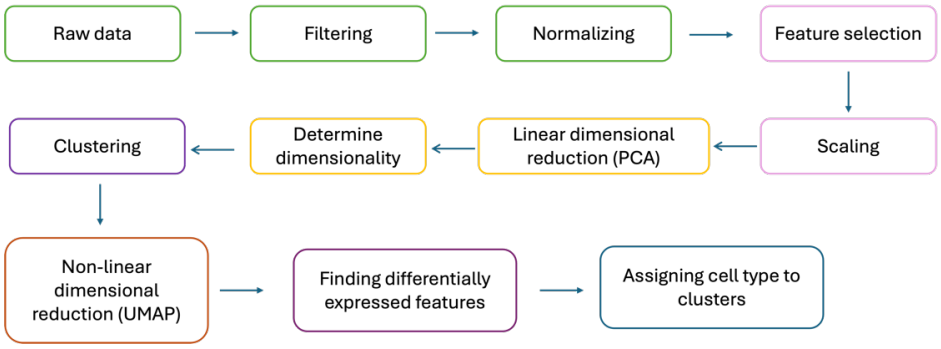


Figure 22: Overview of key steps in Seurat for scRNA-seq data analysis. The analysis begins with raw data preprocessing, including filtering, normalization, and feature selection. Following scaling, dimensionality reduction is performed using principal component analysis (PCA), and the optimal number of dimensions is determined. Cells are then clustered, followed by non-linear dimensionality reduction using UMAP for visualization. Differentially expressed features are identified, and cell types are assigned to clusters based on marker gene expression.

Figure 23 presents a UMAP visualization of the Seurat analysis results, serving as a foundation for downstream differential expression analysis and other advanced computational analyses.

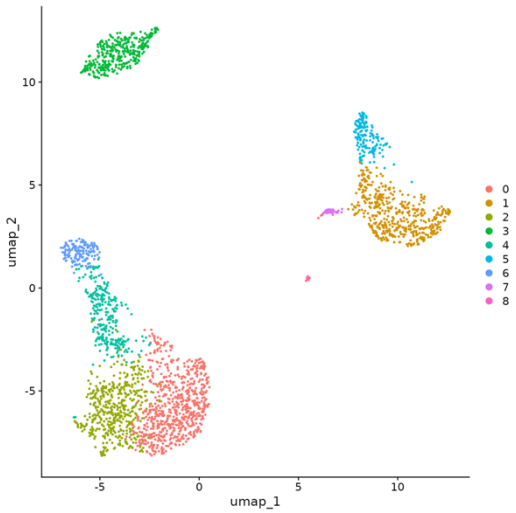


Figure 23: Representative UMAP visualization of the Seurat analysis results. Each cluster represent a distinct cell group. The figure is extracted the Seurat website by Butler et al. (2023) [102].

Cell Type prediction by SingleR

SingleR was employed to predict cell types using its default parameters [103]. It is an easy-to-use tool designed for the automated annotation of single-cell RNA sequencing (scRNA-seq) data by predicting cell types. It operates by correlating single-cell gene expression profiles with reference bulk transcriptomic datasets derived from pure cell populations. The detailed algorithm can be seen in Supplementary Figure 20 and Supplementary Method 7.3.8.

An example output is provided in Figure 24, where normalized scores indicate the confidence of cell type predictions. The colour scale ranges from yellow to red, with higher values representing greater confidence in the assigned cell type.

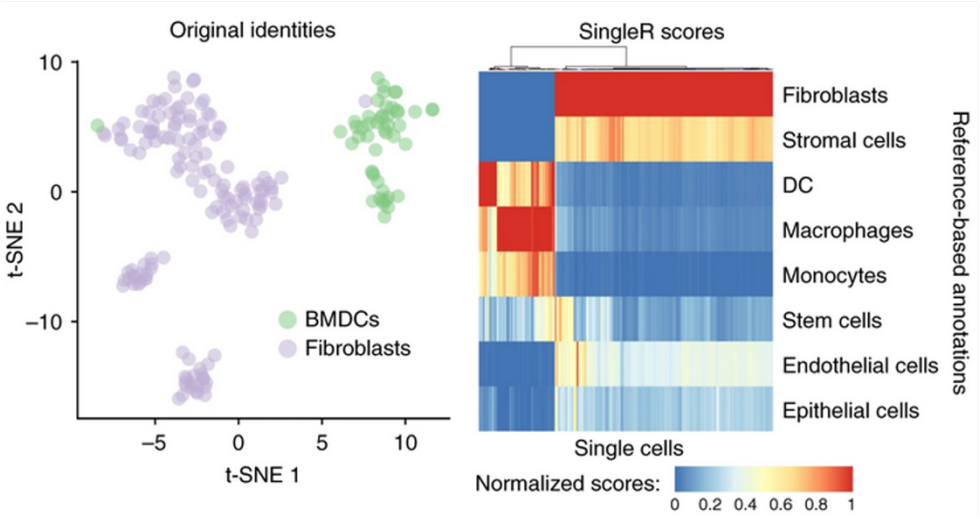


Figure 24: sentative output of cell type predictions generated by SingleR.The left panel depicts the original cell types, which include bone marrow-derived dendritic cells (BMDCs) and fibroblasts. The right panel presents the cell type predictions generated by SingleR, identifying the major cell populations as fibroblasts, stromal cells, dendritic cells (DCs), macrophages, and monocytes. The figure is extracted from Aran et al. (2019) [103].

Differential abundant analysis by DA-seq

Additionally, DA-seq [104] was employed identify variations in the relative abundance of cell populations across experimental conditions. This tool serves as an additional validation method to assess whether there are significant differences between the control and treatment groups. Steps of DA-seq is detailed in Supplementary Method 7.3.9.

1.3.3 Shotgun Proteomics

Shotgun proteomics is a commonly used bottom-up approach for protein identification in complex samples, which integrates high-performance liquid chromatography (HPLC) with mass spectrometry (MS) [105–110]. The proteins in the sample are enzymatically digested into peptides, which are separated via liquid chromatography and then identified using tandem mass spectrometry (MS/MS). This method derives its name from "shotgun" DNA sequencing, reflecting the broad, random selection of peptides for analysis. Shotgun proteomics is widely used to analyze large protein datasets, giving insights into protein expression, post-translational modifications, and protein interactions.

The workflow typically involves four key steps:

1. Protein extraction and digestion: Proteins are extracted from the biological sample and digested into peptides using proteolytic enzymes such as trypsin.
2. Peptide separation: Peptides are separated using liquid chromatography to simplify the mixture before analysis.
3. Tandem mass spectrometry: Peptides are ionized, fragmented, and analyzed using MS/MS to obtain spectra that help identify the peptides.
4. Data analysis: Peptide spectra are matched to known protein databases to infer protein identities and quantify their relative abundances.

Shotgun proteomics offers several advantages that make it a powerful tool for large-scale protein identification. One of its key strengths is its high throughput, allowing researchers to analyze thousands of proteins in a single experiment. Additionally, it provides unbiased and comprehensive coverage, enabling the identification of proteins across a wide dynamic range of abundances. This method is particularly useful for discovering novel proteins, characterizing protein isoforms, and studying post-translational modifications (PTMs). Furthermore, advancements in mass spectrometry technology have significantly improved sensitivity and accuracy, enhancing the reliability of protein identification and quantification.

However, shotgun proteomics also has inherent limitations. One major challenge is incomplete proteome coverage, as the detection of low-abundance proteins may be limited due to sample complexity and ion suppression effects. The reliance on enzymatic digestion

introduces peptide bias, meaning that certain proteins may be underrepresented based on their digestion efficiency. Additionally, the data analysis process is computationally intensive, requiring sophisticated algorithms and extensive database searches to accurately identify proteins. Lastly, reproducibility can be a concern, as variations in sample preparation, chromatography conditions, and instrument performance may impact results across different experiments.

A schematic overview of the shotgun proteomics workflow is illustrated in Figure 25.

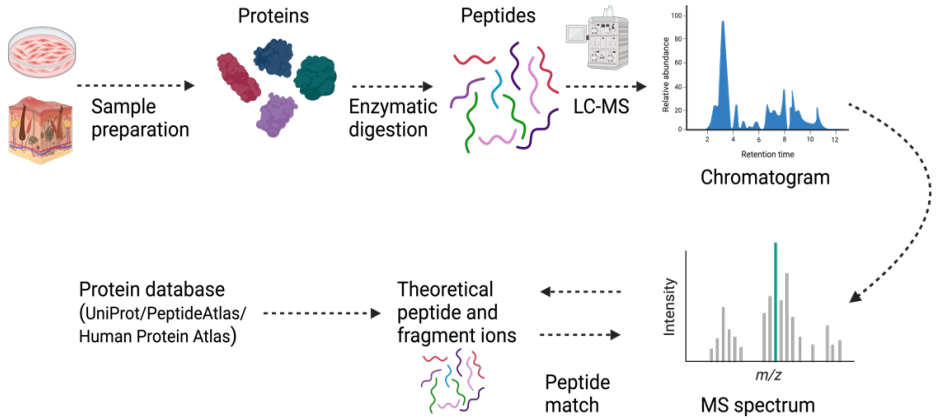


Figure 25: Schematic of a shotgun proteomics experiment. Key steps: (1) extraction of proteins from tear samples, (2) enzymatic digestion to cleave proteins into peptides, (3) separation of peptides using liquid chromatography, (4) analysis of peptide ions via tandem mass spectrometry to produce MS and MS/MS spectra, and (5) comparison of these spectra with predicted spectra derived from peptide sequences in a protein database for identification. Figure 30 is adapted from the Nättinen et al. (2022) [111] and drawn by biorender.com.

The intensity values were first $\log_2$-transformed to meet the normality assumption required by limma [112]. This transformation stabilizes variance and enhances the interpretability of fold changes across the dataset. Limma was then applied to identify differentially abundant proteins across the specified comparisons, as it is well-suited for detecting subtle expression differences in proteomics data when data are approximately normally distributed, which is often achieved through log transformation. The algorithm of Limma is detailed in Supplementary Figure 21 and Supplementary Method 7.3.10.

1.3.4 Phosphoproteomics

Phosphoproteomics, a specialized area of proteomics, focuses on identifying, cataloging, and analyzing proteins that undergo phosphorylation, a common post-translational modification. Phosphorylation is a reversible switch that influences numerous protein attributes, including their activity, cellular location, interactions within protein complexes, and degradation, all of which are central to cell signaling pathways. This modification is so prevalent that an estimated 30–65% of all proteins may be phosphorylated, with some undergoing multiple phosphorylation events [113]. Statistical analyses from extensive datasets predict approximately 230,000 phosphorylation sites in humans, 156,000 in mice, and 40,000 in yeast [114].

Earlier methods to isolate phosphorylated proteins relied on radioactive labeling with 32P-labeled ATP, followed by SDS-PAGE or thin-layer chromatography. These approaches, however, were limited by the inability to produce sufficient protein quantities for thorough phosphorylation analysis. Modern enrichment techniques include affinity purification with phosphospecific antibodies, immobilized metal affinity chromatography (IMAC), strong cation exchange (SCX) chromatography, and titanium dioxide chromatography. Among these, anti-phosphotyrosine antibodies are particularly effective, though antibodies targeting phosphoserine or phosphothreonine are less commonly used. IMAC leverages the phosphate group's affinity for metal ions on the resin, while SCX distinguishes phosphorylated peptides based on their negative charge. Titanium dioxide chromatography, a more recent method, offers faster preparation. Many phosphoproteomic studies now combine these methods to maximize sample purity.

1.4 Knowledge Gaps and Rationale for This Study

Although numerous studies investigating the effects of autoantibodies on keratinocytes have been applied on immortalized HaCaT cell lines [115–117] and non-neoplastic lines, such as skin-derived N/TRET cell lines [118] and oral mucosa-derived cell lines [119]. Studies typically focus either on the normal in vitro differentiation of human primary keratinocyte cells [120] or on the structural analysis of desmoglein in keratinocytes from pemphigus vulgaris patients

using atomic force microscopy [121]. However, current research approaches lack longitudinal studies on keratinocyte responses and fail to fully capture the dynamic transcriptomic and proteomic changes—particularly in primary human keratinocytes or ex vivo skin organ cultures—following exposure to the pathogenic monoclonal antibodies PX43 and AK23 over an extended time course up to 24 hours.

Because primary keratinocyte and skin organ models more faithfully recapitulate in vivo tissue architecture and signaling, a longitudinal multi-omics approach will fill a critical gap: it will reveal the dynamic molecular events that underlie the rapid onset, spatial specificity, and potential reversibility of PV acantholysis. By mapping changes in gene expression and protein abundance in these physiologically relevant systems, our study could uncover novel mechanistic nodes and therapeutic targets that simpler in vitro models cannot expose.

1.5 Objectives of this Study

This study aimed to investigate the molecular effects of pemphigus vulgaris autoantibodies on human primary keratinocyte cultures and human skin organ cultures. Cells were stimulated for up to 24 hours with PX43 scFv and control hIgG, as well as AK23 and its control mIgG. To comprehensively characterize the transcriptional and proteomic changes induced by these autoantibodies, RNA sequencing was performed to identify differentially expressed genes, enriched pathways, and key transcription factors in response to PX43 compared to hIgG and AK23 compared to mIgG. Additionally, proteomic analysis was conducted to examine differentially expressed proteins and associated signaling pathways, providing further insights into the molecular mechanisms underlying pemphigus vulgaris pathogenesis.

2. Overview of Experimental Design and Data Sets

To investigate the molecular impact of pemphigus vulgaris autoantibodies, this study utilized multi-omics datasets derived from both human primary keratinocyte cultures and human skin organ cultures. This section outlines the sources of biological material, the experimental design used for sample preparation, and the quality control measures applied to the transcriptomic and proteomic datasets to ensure robust and reliable analyses.

2.1 Experimental Design

The experimental design consisted of two main components. The first involved human primary epidermal keratinocyte (HPEK) cultures, from which bulk RNA sequencing, proteomic, and single-cell RNA sequencing (scRNA-seq) data were generated. The second component focused on human skin organ cultures (HSOC), for which bulk RNA-seq and proteomic data were obtained.

In designing our time-course, we selected 5h, 10h and 24h post-antibody exposure to capture the key phases of PV acantholysis—from immediate signaling events through to later transcriptional and proteomic remodeling. Prior work has shown that PV-IgG induces rapid kinase activation (e.g. p38 MAPK, ERK) and early desmosome perturbation within 1–5h of treatment, making 5h an optimal point to assay these acute signaling changes [122]. At around 10h, several studies report the emergence of immediate-early gene expression programs and initial protein depletion of extra-desmosomal DSG3, marking the transition from signaling to structural remodeling [39]. Finally, 24h encompasses the full spectrum of downstream transcriptional adaptation and proteome reorganization, including endocytic clearance of cadherins and engagement of stress-response pathways, thus providing a comprehensive view of both the initiation and consolidation phases of antibody-driven keratinocyte detachment [122].

Figure 26 provides an overview of the workflow for Human Primary Epidermal Keratinocyte (HPEK) studies, while Figure 27 outlines the workflow applied in Human Skin Organ Culture (HSOC) experiments.

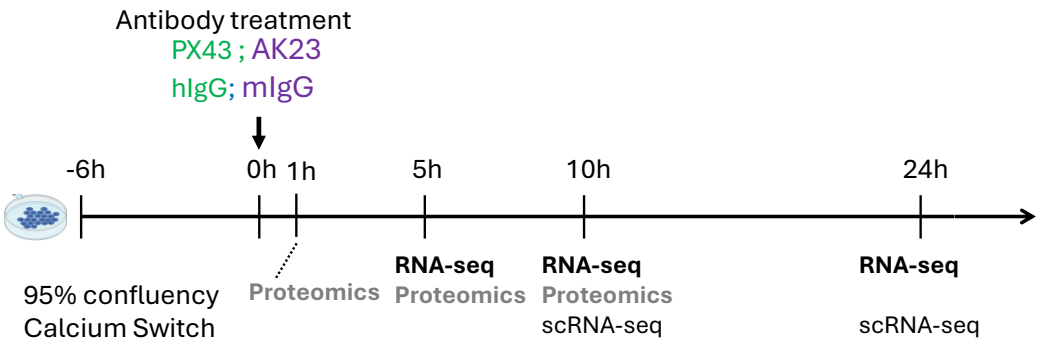


Figure 26: Workflow overview for human primary epidermal keratinocyte (HPEK) studies. Healthy human primary epidermal keratinocyte (HPEK) cells were cultured to 95% confluency, after which the culture medium was switched to a high calcium concentration for 6 hours. Following this treatment, antibodies, including PX43 scFv, complete human immunoglobulin G (hIgG), AK23, and mouse immunoglobulin G (mIgG), were introduced to the HPEK cells, and the cultures were maintained for subsequent time points. Bulk RNA sequencing (RNA-seq) was conducted on samples A, B, and C at 5, 10, and 24 hours (Biological n = 3 for each time point). In contrast, samples X, Y, and Z underwent RNA-seq at 5, 10, 24h (n = 3 for each time point), along with shotgun proteomics and phosphoproteomics analyses at 1, 5, and 10 hours (Biological n = 3 for each time point). Single-cell RNA sequencing (scRNA-seq) was performed for sample Y at 10 and 24 hours (Biological n = 1).

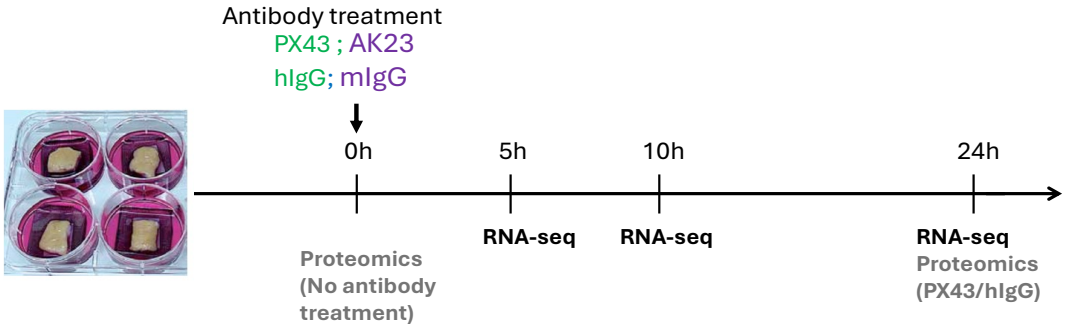


Figure 27: Workflow overview for human skin organ culture (HSOC) studies. Antibodies, including PX43 scFv and complete human IgG, as well as AK23 and mouse IgG, were injected into the epidermal layer of the skin and cultured for 5 hours, 10h hours and 24 hours. Bulk RNA sequencing was performed at 5 hours, 10 hours, and 24 hours post-injection to assess gene expression changes. Shotgun proteomics was conducted at two time points: at 0 hours before antibody treatment and at 24 hours following treatment with PX43 and human IgG.

Dr. William Hariton from Professor Eliane Müller's laboratory at the University of Bern prepared the human primary epidermal keratinocyte (HPEK) and Veronika Hartmann from Professor Jennifer Hundt's laboratory at the University of Lübeck prepared the human skin organ culture (HSOC).

All sequencing, including bulk RNA-seq, scRNA-seq, proteomics, and phosphoproteomics, was carried out by the sequencing facility at the University of Bern.

2.2 Overview of Omics Datasets Used in This Study

For the HPEK samples, bulk RNA-seq data were generated at 5h, 10h, and 24h following treatment. Shotgun proteomics and phosphoproteomics data were collected at 1, 5, and 10 hours. Each time point included treatments with PX43, its control hIgG, AK23, and its control mIgG, with three biological replicates per condition. In addition, single-cell RNA-seq (scRNA-seq) was performed for AK23 and mIgG treatments at 10 and 24 hours, each represented by a single biological sample.

For the HSOC samples, bulk RNA-seq data were generated at 5h, 10h, and 24h following treatment with PX43, hIgG, AK23, and mIgG, with six biological replicates per condition. Shotgun proteomics was performed at the 24-hour time point for PX43 and hIgG treatments, as well as for an untreated baseline control at 0 hours, each with five biological replicates. The transcriptomic data for the HPEK and HSOC samples have been deposited in the GEO database under the accession ID GSE285010.

I conducted the omics data analysis under the supervision of Professor Hauke Busch and Dr. Axel Künstner, with valuable input and suggestions from other members of Professor Hauke Busch's, Professor Jennifer Hundt's, and Professor Ralf Ludwig's lab at the University of Lübeck and Professor Eliane Müller's lab at the University of Bern.

3. Computational Methods and Bioinformatics Analysis

To gain a comprehensive understanding of the molecular mechanisms underlying the biological system under study, we employed an integrative multi-omics data analysis approach. This included bulk RNA sequencing (RNA-seq), shotgun proteomics, phosphoproteomics, and single-cell RNA-seq (scRNA-seq). Standard and custom bioinformatics pipelines were used for quality control, normalization, quantification, statistical analysis, and functional interpretation of each dataset.

The following subsections detail the specific computational steps and analytical methods applied to each data type.

3.1 Bulk RNA-seq Analysis

3.1.1 Quantification of Gene Expression

Raw sequencing reads in FASTQ format [123] were mapped to the human reference transcriptome (GRCh38) using kallisto (v0.45.0) [51], as previously introduced. The first step involved constructing an index of the entire human transcriptome using the kallisto index command, with the reference transcriptome in FASTA format [124], obtained from Ensembl (release 105): https://ftp.ensembl.org/pub/release-105/fasta/homo_sapiens/cdna/. Following index generation, the kallisto quant command was used to pseudoalign raw reads from each sample to the indexed Ensembl human cDNA sequences (version 105), resulting in transcript-level quantification. For each sample, quantification outputs included three files: abundance.tsv, abundance.h5, and run_info.json. These files were generated on a high-performance computing cluster and are available for downstream analyses.

3.1.2 Principal Component Analysis

The top 10% of most variable genes were used to get the PCA (Principal Component Analysis) plot by the PCAtools (v2.14.0) [125], default parameters were used to perform on length-scaled TPM values derived from gene expression data processed by tximport (v1.30.0) [126].

The pca function in the PCAtools package selects the top 10% most variable genes to compute principal components. The biplot function is then used to generate a PCA plot, illustrating the relationships between samples based on the first two principal components (PC1 and PC2).

3.1.3 Identification of differentially Expressed Genes

Differentially expressed genes were quantified using a likelihood ratio test (LRT) implemented in DESeq2 (v1.42.1) [53], as previously introduced. A full model incorporating all experimental variables was compared to a reduced model excluding the treatment effect, to identify genes exhibiting significant expression changes in response to treatment.

The heatmap was generated for the significant differentially expressed genes using the pheatmap (v1.0.12) R package [127] with gene expression values transformed using the regularized logarithm (rlog) transformation. Volcano plots for the significant differentially expressed genes were generated by the tool EnhancedVolcano (v1.20.0) [128], highlighting genes based on log2 fold change and adjusted p-value thresholds.

3.1.4 Pathway Enrichment Analysis

Pathway enrichment analysis was performed using the GAGE (Generally Applicable Gene-set Enrichment) method, implemented via the R package gage (v2.52.2) [59] as previously introduced, with all analyses conducted using the tool's default parameters.

Differentially expressed genes (DEGs) identified by DESeq2, along with their corresponding $\log_2$ fold change values, were used as input for pathway enrichment analysis. The Hallmark gene set collection from the MSigDB database served as the reference for annotation. Enrichment analysis was performed using the gage function from the GAGE package, which computed enrichment scores, p-values, q-values, and other relevant statistical metrics for each comparison.

3.1.5 Transcription Factor Activity Prediction

The decoupleR (v2.8.0) package was utilized to infer transcription factor (TF) activities using $\log_2$ fold-change (log2FC) values of differentially expressed genes as input, the inference was based on prior knowledge of transcription factor-target interactions, generating activity

scores that reflect the regulatory influence of transcription factors within the dataset [68], as previously introduced.

The decoupleR package by reported the top 25 transcription factors (TFs) by default based on the highest absolute predicted activity scores, filtered to include only those with adjusted p-values less than 0.05.

3.2 scRNA-seq Data Analysis

The raw sequencing data obtained from the University of Bern's sequencing facility, was quantified using the count function in Cell Ranger (v7.0.1) [101] as previously introduced. The human reference genome GRCh38 (refdata-gex-GRCh38-2020-A), downloaded from the 10x Genomics website, was utilized for alignment and quantification.

The primary single-cell analysis was conducted using the Seurat package (v4.3.0) [129] as previously introduced, leveraging gene expression quantification results generated by Cell Ranger. Cells with fewer than 200 detected genes or with over 25% of reads mapping to mitochondrial genes were excluded during quality control filtering. Normalization was performed using the global-scaling method LogNormalize, which normalizes gene expression for each cell by the total expression, multiplies by a scaling factor (10,000 by default), and log-transforms the result. The FindVariableFeatures function was used to identify the top 2,000 most variable genes. Data scaling was carried out using the ScaleData function, followed by dimensionality reduction using principal component analysis (PCA) via the RunPCA function. Clustering was performed using the FindClusters function with a resolution parameter of 0.7 to obtain well-resolved clusters. Differential gene expression analysis between clusters was conducted using the default non-parametric Wilcoxon rank sum test implemented in the FindMarkers function.

Cell type annotation was conducted using SingleR, which leverages the gene expression matrix of individual cells to assign cell identities based on reference transcriptomic datasets. Using default parameters, use the filtered and normalized expression matrix from Seurat, SingleR provided unbiased predictions of cell types present in the input data. This analysis enabled a comprehensive assessment of the cellular composition of the sequenced samples, with particular emphasis on confirming the predominance of keratinocytes and evaluating

potential contamination by non-target cell types. Thus support the accuracy and relevance of the dataset for downstream analyses focused on keratinocyte biology.

Additionally, differential abundance analysis (DA-seq) (v1.0.0) [104] was performed using default settings to identify variations in the relative abundance of cell populations of mIgG 24h sample and the AK23 24h sample using the filtered and normalized expression matrix together with the sample information. This analysis provided an extra layer of validation for the clustering and differential expression results, further corroborating the findings derived from these approaches.

3.3 Shotgun Proteomics Data Analysis

The analysis of shotgun proteomics data begins with protein intensity values preprocessed and quantified by the University of Bern's sequencing facility.

Raw protein intensity values were first log2-transformed to stabilize variance across the dynamic range of measurements. To identify differentially abundant proteins across experimental conditions, employed the limma package (v3.58.1) in R [112] as previously introduced. A linear modeling framework was constructed, where the experimental condition (i.e., treatment group and timepoint) was included as a factor (group), and sample batch effects were accounted for (batch). An additive model with interaction terms (group + batch + group:batch) was used to capture potential batch-specific variation across conditions.

Fitting of the linear model was performed using the lmFit function with the default parameter. Variance moderation was applied via empirical Bayes shrinkage using the eBayes function. Specific comparisons of interest were encoded using a contrast matrix, including both treatment-vs-control comparisons (e.g., AK23_1h vs mIgG_1h, PX43_5h vs hIgG_5h) and temporal comparisons within each treatment group (e.g., AK23_10h vs AK23_1h, mIgG_10h vs mIgG_5h). Contrasts were evaluated using contrasts.fit, followed by empirical Bayes moderation with eBayes.

Statistical significance was assessed using moderated t-statistics, and p-values were adjusted for multiple testing using the Benjamini-Hochberg (BH) procedure. Differentially abundant proteins were identified using an adjusted p-value threshold of < 0.05 and the log2FC above

1 or below -1. The topTable function was used to extract ranked lists of proteins for each contrast of interest.

Principal Component Analysis (PCA) plots and heatmaps were generated using the PCAtools (v2.14.0) and pheatmap (v1.0.12) packages, respectively, as in prior bulk RNA-seq analyses.

3.4 Phosphoproteomics Data Analysis

The analysis of phosphoproteomics data begins with protein intensity values pre-processed and quantified by the sequencing facility at the University of Bern. Downstream analysis is similar to the shotgun proteomics.

4. Results: Molecular Insights from the Human Primary Epidermal Keratinocyte (HPEK) and Human Skin Organ Culture (HSOC) Models

Human primary epidermal keratinocyte (HPEK) cells provide a physiologically relevant model to study epidermal biology and cellular responses to external stimuli, making them valuable for investigating skin-related processes at the molecular level.

Human primary epidermal keratinocyte (HPEK) cells and human skin organ culture (HSOC) both serve as physiologically relevant models for studying epidermal biology and cellular responses to external stimuli. These systems are particularly valuable for investigating skin-related processes at the molecular level due to their close approximation to in vivo human skin conditions.

For the HPEK model, two experimental cohorts were established to explore transcriptomic and proteomic dynamics following stimulation.

- Cohort 1 included samples A, B, and C, each comprising pooled keratinocytes from three healthy male donors' foreskins, with three biological replicates per sample.
- Cohort 2 consisted of samples X, Y, and Z, each derived from a single healthy male donor, also with three biological replicates, to minimize potential variability from donor pooling.

  HPEK cells were cultured to 95% confluency and then exposed to high calcium conditions to stabilize cell-cell adhesion and synchronize differentiation. Six hours later, cells were stimulated with PX43 scFv, hIgG, AK23, or mIgG. Samples were harvested at 5h, 10h, and 24h post-stimulation for bulk RNA-sequencing.

- For scRNA-seq, only AK23- and mIgG-treated samples at 10 and 24 hours (without biological replicates) were processed.
- For shotgun and phosphoproteomics, samples X, Y, and Z (with three biological replicates each) were collected at 1h, 5h, and 10h post-stimulation.

For the HSOC (human skin organ culture) model, normal human skin samples were obtained from patients within 24 hours post-surgery, with informed consent provided for research use. Five independent biological replicates (i.e., skin samples from five different donors) were

subjected to stimulation with PX43 scFv, hIgG, AK23, or mIgG antibodies. Following 24 hours of incubation, each skin sample was bisected using a sterile scalpel. One half was processed for histological analysis via haematoxylin-eosin (H&E) staining, while the other half was further subdivided, with a portion used for RNA sequencing.

For proteomic analysis, untreated HSOC samples collected at 0 hours served as negative controls, and samples stimulated with PX43 scFv and hIgG for 24 hours (five biological replicates each) were processed for protein extraction and subsequent mass spectrometry–based proteomics.

4.1 Overall Data Quality

High-quality input data is fundamental to ensuring reliable analytical outcomes and valid biological interpretations. As the well-known adage in data analysis states, “garbage in, garbage out.” Consequently, the first critical step in the sequencing data analysis pipeline involves a thorough assessment of raw data quality to determine whether it meets the required standards. Based on this evaluation, low-quality reads may need to be filtered out, or, in cases of insufficient quality, certain samples may require resequencing.

To assess overall data quality, FastQC was employed to analyze multiple quality metrics, including sequencing depth (read counts), duplication levels, GC content, and base quality scores (e.g., Q30) [130]. These metrics provide a comprehensive overview of the integrity and reliability of the raw data. To facilitate comparison and visualization across all samples, the results from FastQC were aggregated using MultiQC [131].

Figure 28 presents the read count results, while Figure 29 illustrates the mean quality scores across reads for samples X, Y, and Z. Samples A, B, and C are referenced in Appendix A, Supplemental Figure 1-2.

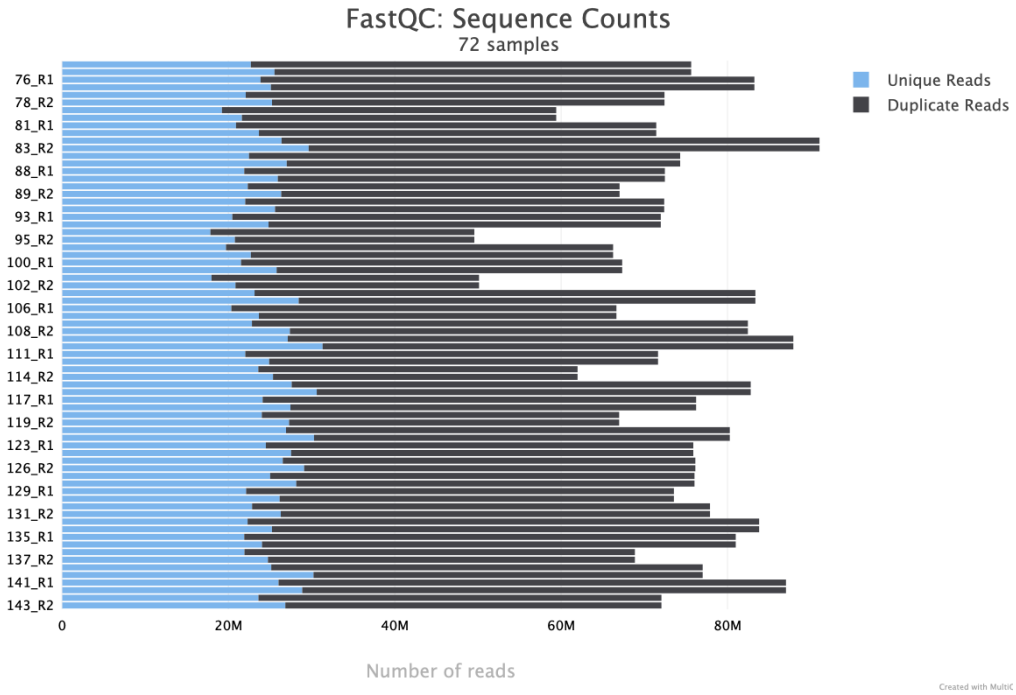


Figure 28: Total read counts from bulk RNA-seq for HPEK samples X, Y, and Z. The quality control results aggregated by MultiQC from FastQC outputs indicate that all samples achieved sequencing depths exceeding 50 million reads, with the majority surpassing 60 million reads. These read counts meet or exceed the recommended thresholds for bulk RNA-seq experiments as outlined by ENCODE guidelines [132], suggesting sufficient sequencing depth for downstream analyses.

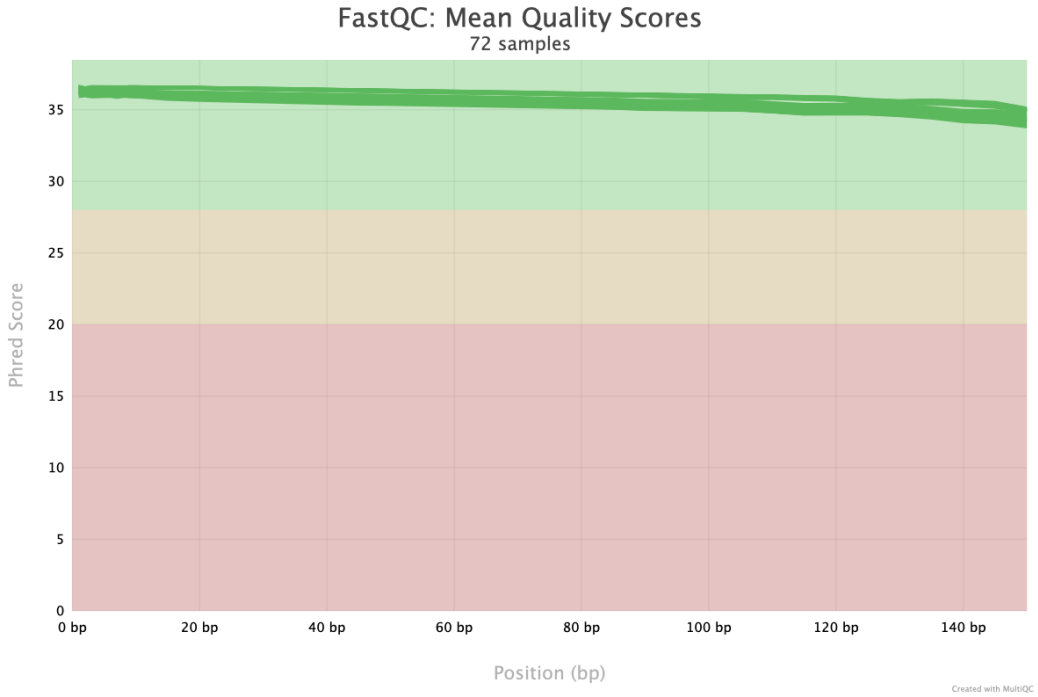


Figure 29: Mean base quality scores across reads from bulk RNA-seq of HPEK samples X, Y, and Z. All samples exhibited mean quality scores above 33.8, as reported by FastQC and summarized by MultiQC. This corresponds to a base call accuracy exceeding 99.9%, indicating that the raw sequencing reads are of exceptionally high quality.

The total read counts for all samples exceed 50 million, with the majority surpassing 60 million reads, aligning with the ENCODE consortium's recommended standards for high-quality RNA-seq data [132]. Additionally, the mean base quality scores across reads are all above 33.8, corresponding to a base call accuracy of over 99.9%. Collectively, these metrics indicate that the raw sequencing data are of high quality.

Following the assessment of samples X, Y, and Z, which demonstrated high sequencing depth and quality, the HSOC raw sequencing reads were similarly evaluated to ensure their suitability for downstream analyses. As shown in Figure 30, the HSOC samples also exhibit robust read counts consistent with high-quality RNA-seq data. Complementing this, Figure 31 presents the distribution of mean base quality scores across reads, further confirming that the HSOC bulk RNA-seq data meet the necessary quality thresholds for reliable analysis.

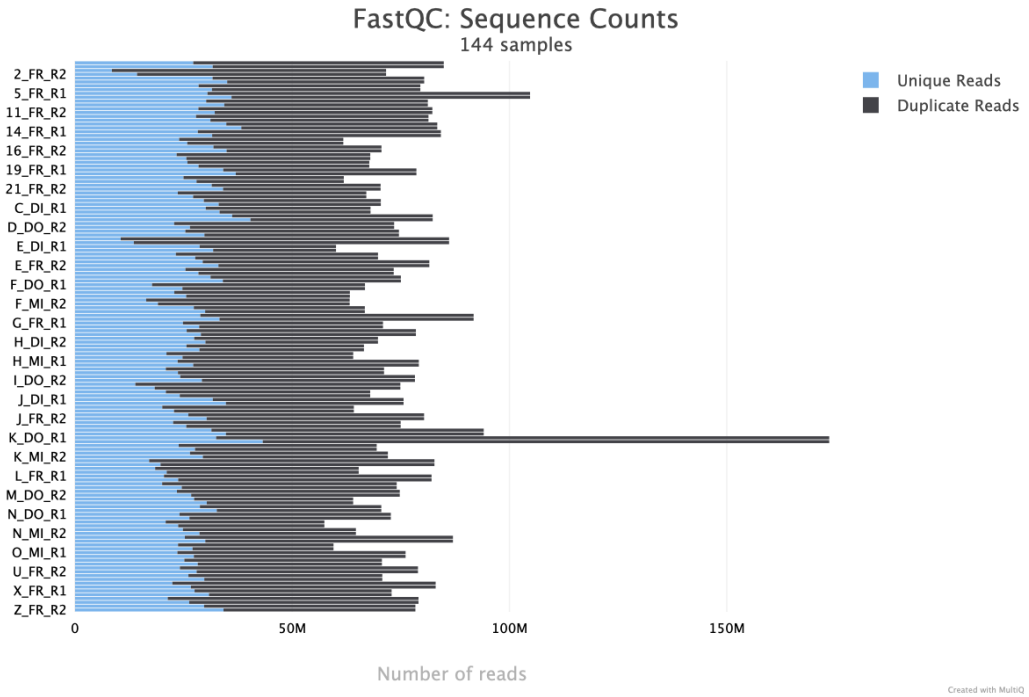


Figure 30: Total read counts from bulk RNA-seq for HSOC samples. The quality control results aggregated by MultiQC from FastQC outputs indicate that all samples achieved sequencing depths exceeding 56 million reads, with the majority surpassing 60 million reads. These read counts meet or exceed the recommended thresholds for bulk RNA-seq experiments as outlined by ENCODE guidelines [132], suggesting sufficient sequencing depth for downstream analyses.

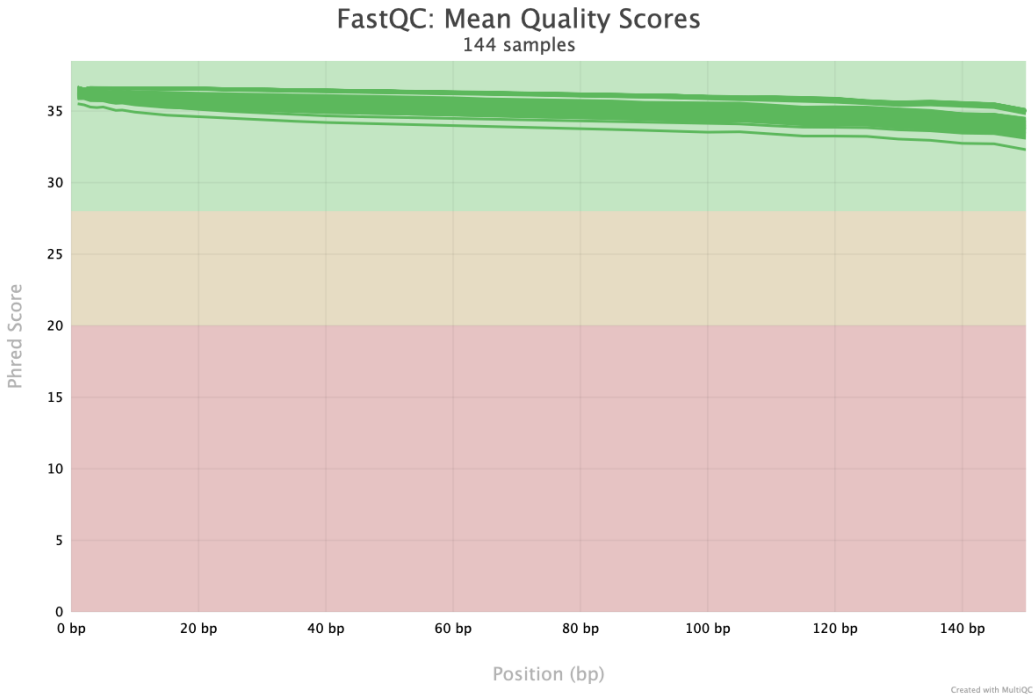


Figure 31: Mean base quality scores across reads from bulk RNA-seq of HSOC samples. All samples exhibited mean quality scores above 32.3, as reported by FastQC and summarized by MultiQC. This corresponds to a base call accuracy exceeding 99.9%, indicating that the raw sequencing reads are of exceptionally high quality.

For the HPEK scRNA-seq dataset, raw sequencing reads were also evaluated and confirmed to be of high quality, ensuring their reliability for downstream single-cell analyses.

Once the high quality of the sequencing reads was confirmed, it is confident to processed with the downstream analyses.

4.2 HPEK Bulk RNA-seq

4.2.1 Donor Effects Observed in HPEK Bulk RNA-seq

To evaluate the effects of AAbs on human primary keratinocyte cells across different treatments (PX43, hIgG, AK23, and mIgG) and time points (5 h, 10 h, and 24 h), a Principal Component Analysis (PCA) plot will first be utilized to assess the overall clustering patterns of the samples. This initial exploratory analysis will provide an overview of the similarities and differences among the sample groups.

Figure 32 shows the PCA plots for samples X, Y, and Z from Experiment 2, at each time point, sample clustering is primarily driven by the first principal component (PC1), accounting for 39–51% of the total variance, depending on the time point. The second principal component (PC2) explains a smaller proportion of variance (19-31%).

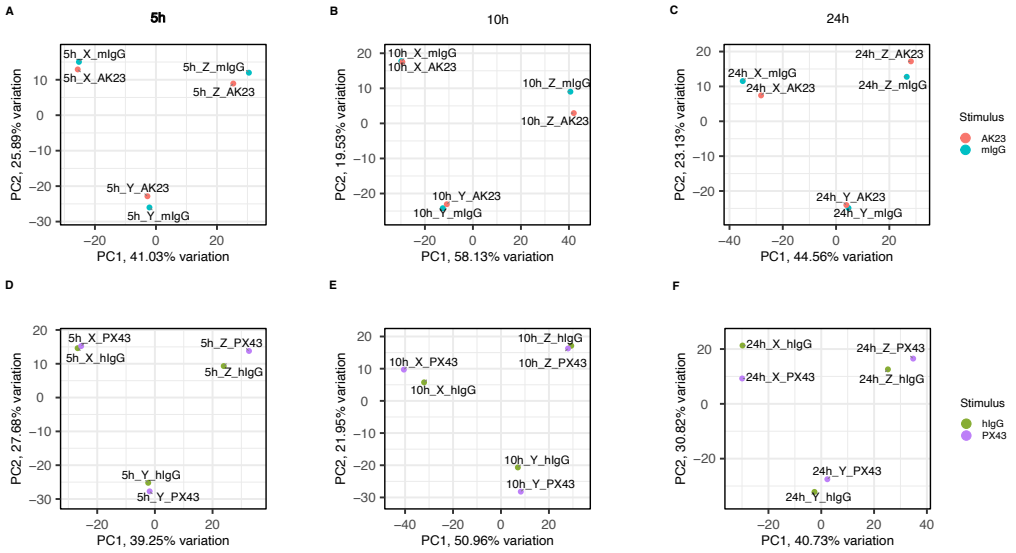


Figure 32: Principal Component Analysis (PCA) plots comparing samples X, Y, and Z from experiment 2. The up three plots show comparisons between AK23 and its control (mIgG), while bottom three plots display comparisons between PX43 and its control (hIgG). Separation between samples occurs mainly along the first principal component (PC1), accounting for 39-51% of the total variance, while variation along the second dimension is minimal.

The treated samples remain relatively close to their respective controls in the PCA plot, suggesting that the overall transcriptional changes induced by treatment are modest. A strong donor-specific effect is evident across the dataset. When samples from different time points (5 h, 10 h, and 24 h) for AK23 and mIgG, or PX43 and hIgG, are analyzed together, a clear temporal clustering is observed, with samples from the same time point grouping more closely. This consistent clustering confirms that time point exerts a substantial effect on the transcriptional profiles. Overall, these patterns suggest that donor and time effects play a more prominent role in sample clustering than the specific effects of autoantibody treatments on human primary keratinocyte cells.

Consistent clustering patterns are observed in the PCA plots of samples A, B, and C (Supplemental Figure 3), as well as in the combined analysis of all samples (A, B, C, X, Y, and Z) shown in Supplemental Figure 4.

Thus, these observations indicate that, under the experimental conditions, antibody treatments elicit only modest transcriptional alterations.

4.2.2 Minimal Transcriptomic Response in HPEK

Following the initial exploratory analysis using PCA to assess overall data structure, the next step involves identifying differentially expressed genes (DEGs). This analysis will provide detailed insights into the molecular differences between treatment and control antibody conditions (PX43 vs. hIgG and AK23 vs. mIgG).

Differentially expressed genes (DEGs) were identified using DESeq2, as described in the Methods section. Genes meeting the criteria of an adjusted p-value (padj) < 0.05 and an absolute $\log_2$ fold change ($|\log_2 FC| > 1$ were considered significantly differentially expressed. The threshold of padj < 0.05 was applied to control the false discovery rate, while the $\log_2$ fold change cutoff was used to capture genes exhibiting biologically meaningful expression changes. These widely accepted criteria provide a balance between statistical stringency and effect size. The resulting DEGs are visualized in volcano plots.

Minimal Differentially Expressed Genes in HPEK Samples X, Y, and Z

Figure 33 presents the DEGs for samples X, Y, and Z, under the same comparative conditions and time points. Results for samples A, B, and C, as well as the combined samples A, B, C, X, Y, and Z, are presented in Supplemental Figures 5 and 6.

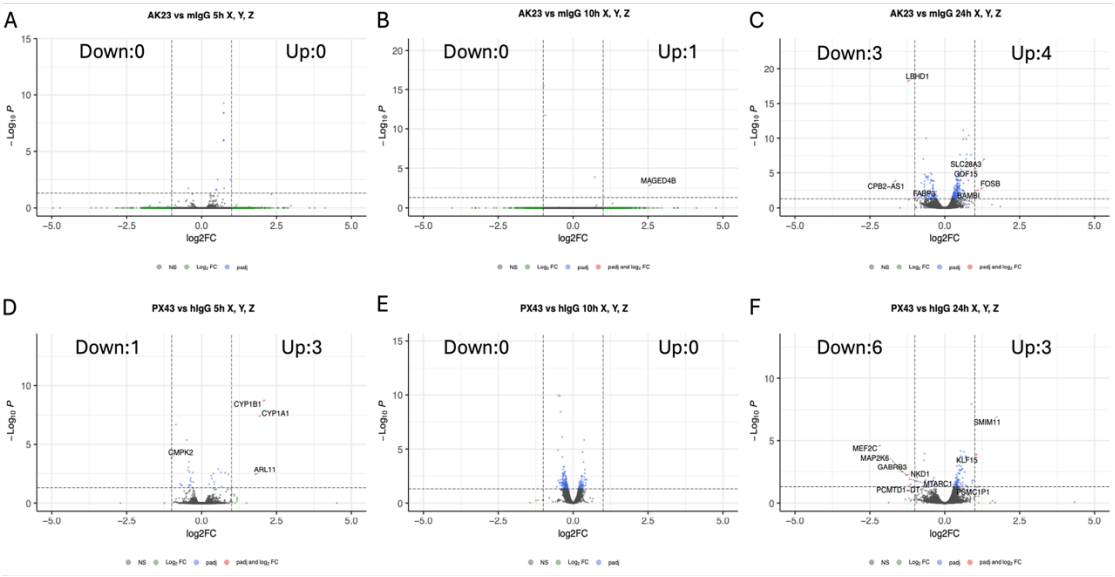


Figure 33: Volcano plots of differentially expressed genes in samples X, Y, and Z comparing PV antibodies with control antibodies. Each subplot displays the $\log_2$ fold-change ($\log_2$FC) on the x-axis and the negative $\log_{10}$ of the adjusted p-value ($-\log_{10}$ P) on the y-axis, with upregulated (Up) and downregulated (Down) genes indicated. Significant genes are labeled in each plot. AK23 vs. mIgG at 24h (Up: *FOSB*, *BAMBI*, *GDF15*, *SLC28A3*, Down: *LBHD1*, *CPB2-AS1*); PX43 vs. hIgG at 5h (Up: *ARL11*, *CYP1A1*, *CYP1B1*, Down: *CMPK2*), 24h (Up: *SMIM11*, *KLF15*, *PSMC1P1*, Down: *MTARC1*, *NKD1*, *PCMTD1-DT*, *GABRB3*, *MEF2C*, *MAP2K6*). Overall, only a few DEGs were identified.

For samples X, Y and Z, AK23 treatment compared to mIgG identified one upregulated gene MAGED4B at 10h, four upregulated genes of *SLC28A3*, *FOSB*, *GDF15* and *BAMBI*, three downregulated genes of *LBHD1*, *CPB2-AS1*, and *FABP3* at 24h. PX43 treatment compared to hIgG identified three upregulated genes of *ARL11*, *CYP1A1*, *CYP1B1*, one downregulated gene *CMPK2* at 5h, three upregulated genes of *KLF15*, *SMIM11*, *PSMC1P1*, six downregulated genes of *MTARC1*, *NKD1*, *PCMTD1-DT*, *GABRB3*, *MEF2C*, *MAP2K6* at 24h.

- *MAGED4B* (MAGE Family Member D4B) correlates with cell migration and growth [133].
- *SLC28A3* (Solute Carrier Family 28 Member 3) is a member of the human solute carrier family transporters and may contribute to drug-induced skin disorders [134].

- *FOSB* (FosB Proto-Oncogene) can be both beneficial and harmful in injured mouse skin depending on the specific context and the dimerization partners involved [135].
- *GDF15* (Growth Differentiation Factor 15) emerges as a novel regulator of mitochondrial homeostasis, mitigating cellular senescence and attenuating age-related phenotypic changes in a 3D human skin mode [136].
- *BAMBI* (BMP And Activin Membrane Bound Inhibitor) encodes a transmembrane glycoprotein related to the type I receptors of the transforming growth factor-beta family and significantly increases its expression in both chronic and acute wounds [137].
- LBHD1 (LBH Domain Containing 1) has been shown to promote the migratory and invasive potential of bladder cancer cells while exerting minimal influence on cellular proliferation [138].
- CPB2-AS1 (CPB2 antisense RNA 1), when upregulated, seems to contribute to impaired wound healing in diabetic conditions [139].
- FABP3 (Fatty Acid Binding Protein 3) is involved in the transport of long-chain fatty acids to various cellular compartments, such as mitochondria for β-oxidation, the endoplasmic reticulum for esterification, or the nucleus for lipid signaling [140].
- ARL11 (ADP Ribosylation Factor Like GTPase 11) has been shown to regulate the activation of macrophages and inflammatory cytokines in a mouse atherosclerosis model [141].
- CYP1A1 (Cytochrome P450 family 1 subfamily A member 1) and CYP1B1 (Cytochrome P450 family 1 subfamily B member 1) belong to the cytochrome P450 enzymes, involved in the metabolism of various endogenous substrates, including fatty acids, steroid hormones, vitamins and have increased enzymatic activity in the inflammatory skin [142–146].
- CMPK2 (Cytidine/uridine Monophosphate Kinase 2) is a mitochondrial nucleotide monophosphate kinase needed for salvage dNTP synthesis that mediates immunomodulatory and antiviral activities through IFN-dependent and IFN-independent pathways [147–151].

- KLF15 (KLF Transcription Factor 15) regulates diverse biological processes that include proliferation, differentiation, growth, development, survival and responses to external stress, also upregulated in the topical glucocorticoid clobetasol propionate (CBP) treated healthy humans [152,153].
- SMIM11 (Small Integral Membrane Protein 11) encodes proteins that are localized in the cytoplasm and mitochondria, as well as focal adhesion points between the cells [154].
- PSMC1P1 (Proteasome 26S subunit, ATPase 1 Pseudogene 1) has been shown to regulate human photoreceptor development and maturation [155].
- MTARC1 (Mitochondrial Amidoxime Reducing Component 1) downregulation in hepatocytes protects against metabolic dysfunction associated with steatohepatitis in multiple murine models [156].
- NKD1 (NKD inhibitor of WNT signaling pathway 1) functions as a passive antagonist of Wnt signaling and affects the homeostasis of stem cells [157].
- PCMTD1-DT (PCMTD1 Divergent Transcript) is a divergent transcript of the PCMTD1 gene expressed at lower levels in the lesion skin [158].
- GABRB3 (Gamma-Aminobutyric Acid type A Receptor Subunit beta3) encodes a ligand-gated ion channel family member and affects the cell-cell junctions between keratinocytes [159].
- MEF2C (Myocyte Enhancer Factor 2C) is a substrate of ERK5, and activation of ERK5 has been associated with accelerated wound healing [160].
- MAP2K6 (Mitogen-Activated Protein Kinase Kinase 6) is a mitogen-activated protein (MAP) kinase member, which is required for human keratinocyte migration on dermal collagen [161].

Table 5 summarizes the significantly differentially expressed genes (DEGs) identified in AK23-treated samples compared to the mIgG control, while Table 6 presents the corresponding DEGs for PX43-treated samples compared to hIgG. Notably, no individual gene exhibited consistent upregulation or downregulation across all three time points, highlighting the temporal specificity of gene expression changes in response to treatment.

Table 4: Significantly differentially expressed genes in AK23-treated samples compared to mIgG control at 5 h, 10 h, and 24 h

	5h	10h	24h
MAGED4B	-1.09	**2.53**	-0.24
	1	**0.001**	0.94
SLC28A3	-0.03	0.26	**1.31**
	1	1	**9.00E-08**
FOSB	-0.05	0.22	**1.14**
	1	1	**0.003**
GDF15	0.51	0.21	**1.03**
	1	1	**2.24E-06**
BAMBI	0.08	-0.24	**1.11**
	1	1	**0.003**
LBHD1	-0.11	-0.23	**-1.22**
	1	1	**5.27E-19**
CPB2-AS1	0.84	-0.18	**-1.64**
	1	1	**0.00013**
FABP3	-0.05	0.25	**-1**
	1	1	**0.045**

***** Values represent log2FC (top) and padj (bottom). **Bold** indicates significance: padj < 0.05 and |log2FC| > 1.

Table 5: Significantly differentially expressed genes in PX43-treated samples compared to hIgG control at 5 h, 10 h, and 24 h

	5h	10h	24h
ARL11	**1.8**	-1.09	0.75
	0.0051	NA	0.76
CYP1A1	**1.93**	0.97	0.38
	5.23E-08	NA	0.23
CYP1B1	**2.09**	0.97	0.14
	2.48E-09	NA	0.98
CMPK2	**-1.01**	-0.17	0.12
	1.70E-04	NA	NA
KLF15	-0.02	-0.33	**1.06**

	1.00E+00	NA	**0.00012**
SMIM11	0.36	-1.15	**1.74**
	1.00E+00	NA	**1.21E-07**
PSMC1P1	-0.07	-0.2	**1.13**
	1.00E+00	NA	**0.049**
MTARC1	-0.045	0.15	**-1.18**
	1.00E+00	NA	**0.012**
NKD1	-0.047	-0.94	**-0.054**
	1.00E+00	NA	**0.0059**
PCMTD1-DT	-0.29	0.27	**-1.14**
	1.00E+00	NA	**0.033**
GABRB3	0.57	-0.23	**-1.43**
	1.00E+00	NA	**0.0018**
MEF2C	0.14	-1.13	**-2.17**
	1.00E+00	NA	**2.44E-05**
MAP2K6	-0.69	-0.61	**-1.13**
	1.00E+00	NA	**0.094**

***** Values represent log2FC (top) and padj (bottom). **Bold** indicates significance: padj < 0.05 and |log2FC| > 1.

Taken together, no consistent or robust gene expression changes were observed in response to AK23 or PX43 treatment in samples X, Y, and Z. While a limited number of significantly differentially expressed genes were detected, the statistical power remains insufficient to draw definitive conclusions. Further experimental validation is necessary to assess the biological relevance of these preliminary findings.

Similarity of Gene Expression Profiles in Samples X, Y, and Z via Correlation Analysis of TPM Values

To enable a more precise comparison of gene expression levels between treatment and control antibodies in samples X, Y, and Z (each consisting of keratinocytes from a single healthy donor), transcripts per million (TPM) values were generated for all samples. These values were $\log_2$-transformed to stabilize variance and enhance cross-sample comparability. Spearman correlation analyses were subsequently conducted to evaluate the relationships between PX43 and hIgG, AK23 and mIgG, at 5h, 10h, 24h, and 48h post-treatment.

Furthermore, the calculated correlations between AK23 and mIgG, as well as PX43 and hIgG, at 5 h, 10 h, and 24 h, revealed consistently strong positive relationships ($r = 0.94–1.00$, $p < 0.001$) between treatment and control conditions. These findings suggest that the effects of

the treatment antibodies are not easily distinguishable from those of the control antibodies at the transcriptomic level.

Consequently, the differentially expressed genes (DEGs) identified in these analyses may primarily reflect noise rather than true biological differences, which also could explain the low number of significant DEGs observed.

4.2.3 Lack of Consistently Regulated Pathways and Transcription Factors in HPEK

Beyond examining differentially expressed genes (DEGs), it is highly informative to perform pathway-level analysis, as genes typically function within coordinated modules. This approach not only facilitates a more comprehensive understanding of the underlying biological mechanisms but also enhances the interpretability of the results by highlighting key biological processes that differ between treatment and control conditions.

Significantly enriched pathways were identified using the GAGE R package in conjunction with the Hallmark gene set collection, as described in the Introduction and Methods sections. Pathways with a q-value < 0.05 are visualized either in dot plots for the PX43 versus hIgG and AK23 versus mIgG comparisons, or collectively in a heatmap encompassing all sample conditions.

In samples X, Y, and Z, the top significantly enriched pathways are presented in Figure 34, encompassing all conditions: PX43, hIgG, AK23, and mIgG at 5, 10, and 24 hours. For individual comparisons, significant pathway enrichment was observed only in the comparison of PX43 versus hIgG at 5 and 24 hours (Supplemental Figures 7 and 8), as well as in the comparison of AK23 versus mIgG at 24 hours (Supplemental Figure 9).

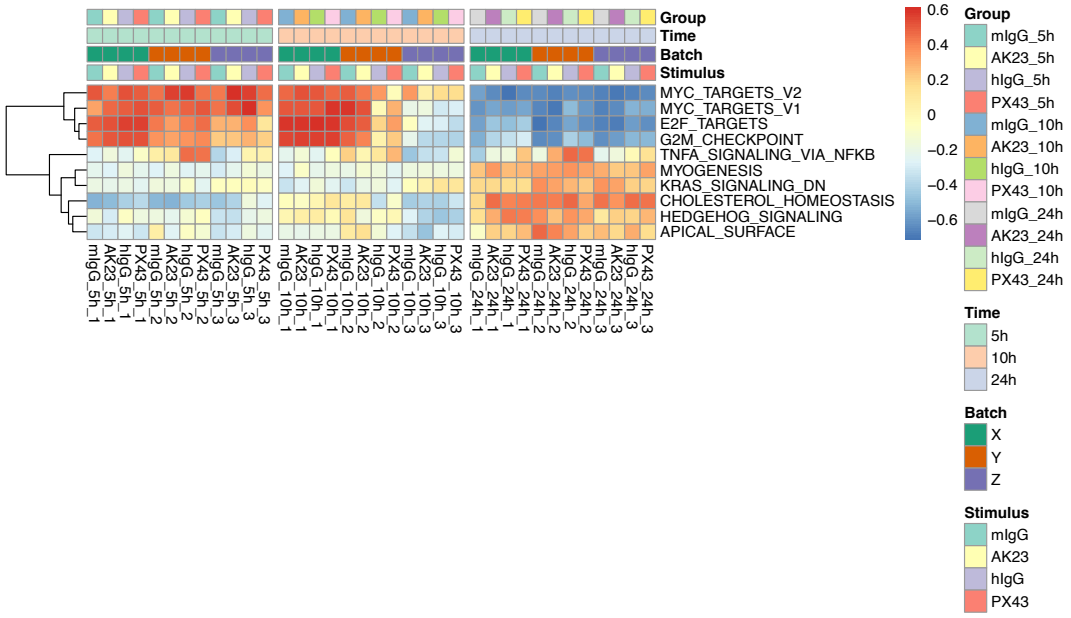


Figure 34: Top significantly enriched pathways in HPEK samples X, Y, and Z based on pathway analysis. Using the HALLMARK pathway reference database, the top significantly enriched pathways were identified across all comparisons, revealing clear time-dependent effects. Specifically, pathways such as TNFα signaling, myogenesis, KRAS signaling, cholesterol homeostasis, Hedgehog signaling, and apical surface processes were upregulated at 24 hours but downregulated at 5 and 10 hours. In contrast, MYC targets, E2F targets, and G2M checkpoint pathways were upregulated at 5 and 10 hours and downregulated at 24 hours. Moreover, the similarity in pathway enrichment profiles between AK23 and mIgG, as well as between PX43 and hIgG, suggests that the observed pathway changes are likely driven more by spatial or temporal factors rather than antibody-specific effects.

As shown in Figure 34, the similarity in pathway enrichment profiles between AK23 and mIgG, as well as between PX43 and hIgG, along with the clustering of pathways by time point across samples, suggests that the observed pathway changes are primarily driven by spatial or temporal factors rather than antibody-specific effects.

Similar patterns of top significantly enriched pathways were observed in individual samples A, B, and C, as well as in the combined analysis of samples A, B, C, X, Y, and Z.

In summary, no consistently regulated signaling pathways were identified in HPEKs across the examined time points, indicating a lack of sustained or uniform pathway activation in response to the treatments. Furthermore, the clustering of pathways by time point across samples suggests that the observed effects are not antibody specific.

In addition, transcription factor activities were inferred from the bulk RNA-seq data to identify potential hub regulators. This analysis provides a valuable approach for uncovering key regulatory elements involved in the disease process.

Top transcription factors and their predicted activity scores were identified using decoupleR. Across individual sample sets (A, B, and C or X, Y, and Z), as well as in the aggregated analysis of all samples (A–C and X–Z), no transcription factor exhibited consistent upregulation or downregulation across all three time points (5 h, 10 h, and 24 h). This lack of consistency is not unexpected, as the inference of transcription factor activity is derived from gene expression profiles for each comparison, and no uniform gene expression pattern was observed across these time points or sample groups.

To sum up, multiple analyses of the bulk RNA-seq data—including differential gene expression, pathway enrichment, and transcription factor activity—indicate that the observed changes are primarily driven by temporal dynamics rather than specific antibody responses. HPEKs may be relatively resilient to the applied treatment or stimuli, at least within the timeframe examined. This could reflect a tightly regulated homeostatic mechanism in keratinocytes that dampens transcriptional shifts in response to transient external cues, or a delayed response profile that may not manifest until later timepoints. Alternatively, it raises the possibility that the treatment exerts more subtle effects on post-transcriptional regulation, which would not be captured in bulk RNA-seq. These findings underscore the importance of considering both cell-type specific responsiveness and the kinetics of gene expression when interpreting multi-condition transcriptomic data.

4.3 HPEK scRNA-seq

The bulk RNA-seq data did not reveal antibody-specific effects, showing minimal differences between PX43 and hIgG, as well as AK23 and mIgG. Since bulk RNA-seq captures the average gene expression across a heterogeneous cell population, it is possible that transcriptional responses occurring in only a subset of cells are being masked or diluted. This raises the possibility that the subtle changes observed may reflect a heterogeneous response among human primary keratinocytes. To investigate this further, our collaborators at the University of Bern performed single-cell RNA sequencing (scRNA-seq) on AK23- and mIgG-treated samples at 10h and 24h post-treatment, aiming to resolve cell-specific responses and test this hypothesis.

Single-cell RNA sequencing (scRNA-seq) analysis of HPEK cells treated with AK23 or mIgG for 24 hours revealed multiple distinct cellular subpopulations, providing insights into treatment-

associated alterations in cell states, particularly those specifically induced by AK23 in comparison to the mIgG control.

4.3.1 Identification of 12 Subcellar Clusters in AK23 and mIgG 24h Samples in HPEK

Sample Y was subjected to single-cell RNA-seq following treatment with AK23 and mIgG at 10-hour and 24-hour time points. The target cell count distribution is presented in Figure 35. Due to the excessively high mean reads per cell observed in the mIgG 10h sample (>1 million), downstream analyses were focused on the mIgG 24h and AK23 24h samples.

In Figure 36, the Uniform Manifold Approximation and Projection (UMAP) visualization displays 12 distinct cell clusters identified using Seurat for both samples of mIgG 24h and AK23 24h. Clustering was performed with default parameters, except for the resolution, which was set to 0.7 in the FindClusters function. This resolution was selected after testing multiple values, as it provided optimal separation of the cell groups.

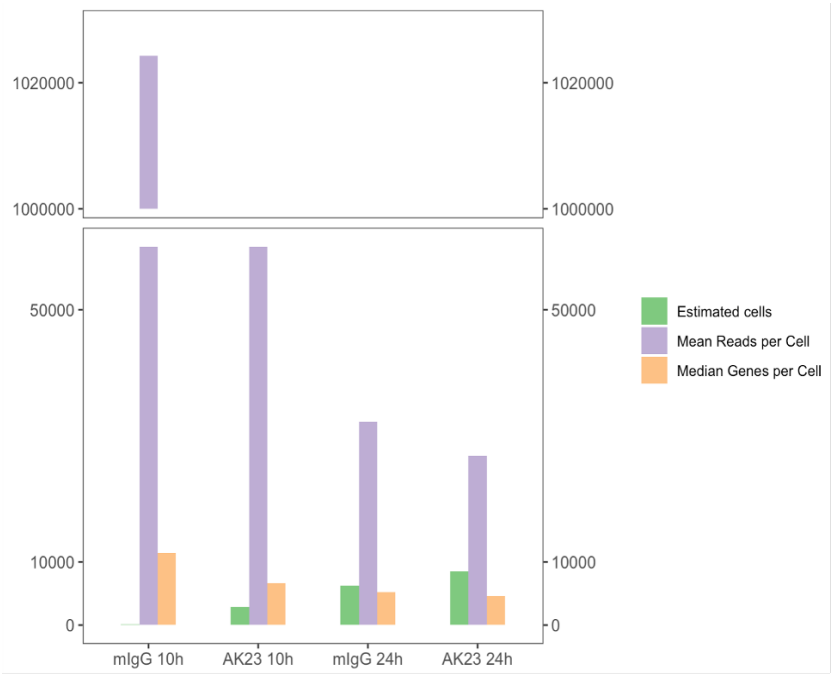


Figure 35: Target number of captures of cells for AK23 and mIgG at 10h and 24h. According to the 10x Genomics guidelines, the mean reads per cell typically ranges around 10,000. However, the mIgG 10h sample exhibits an excessively high mean reads per cell (>1 million), suggesting potential quality issues with this sample.

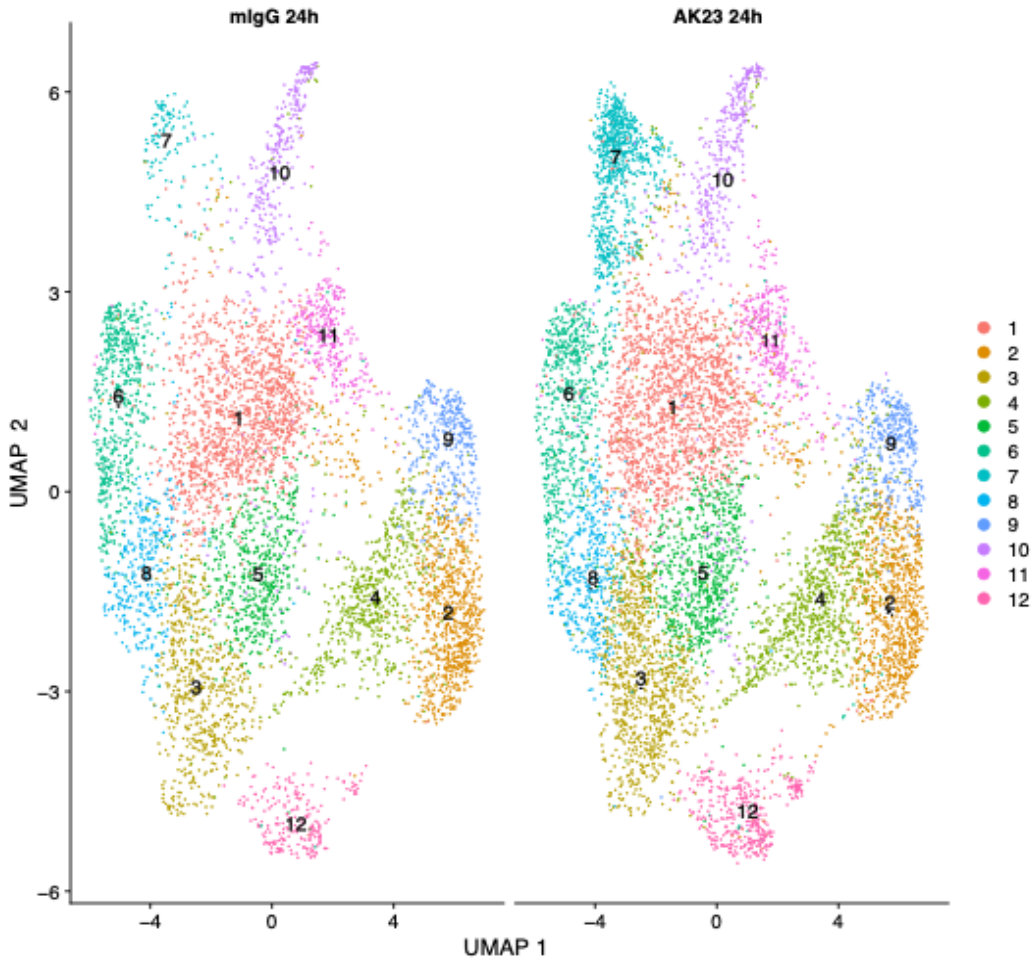


Figure 36: Uniform manifold approximation and projection (UMAP) visualization of 12 distinct cell clusters identified by Seurat. Both mIgG and AK23 samples yielded 12 distinct cell clusters, identified using Seurat. The clustering resolution was set to 0.7 in the FindClusters function, selected after testing multiple values to achieve optimal separation of cell groups.

4.3.2 AK23 Induces a Limited Transcriptomics Response Relative to mIgG Across Cell Clusters in HPEK

For the significantly differentially expressed genes (DEGs) in each cluster (AK23 vs. mIgG), applying a threshold of padj < 0.05 and |log2FC| > 1 identified only two significant DEGs of PHLDB2 and FRMD6 in Cluster 7. When the threshold was adjusted to padj < 0.05 without a log2FC cutoff, Cluster 3 additionally exhibited two significant DEGs of KRTDAP and SPRR2E. Overall, these findings suggest minimal transcriptional differences between the mIgG 24h and AK23 24h samples, limited to a small number of clusters.

Figure 37 presents a heatmap of each cell cluster's top 5 marker genes, highlighting distinct expression patterns that define each cluster. Table 7 provides a list of the top 5 marker genes for each cluster along with their associated functions, as identified using the STRING database. Additionally, GAGE analysis was performed to infer enriched pathways based on the differentially expressed genes in each cluster (AK23 vs. mIgG), with significant pathways identified at $q < 0.05$. Among the 12 cell clusters, only clusters 2, 5, and 8 showed significantly enriched pathways, all of which included the upregulated pathway TNFA Signaling via NFKB, indicating a heterogeneous response among the cell clusters, as shown in Figure 38-40.

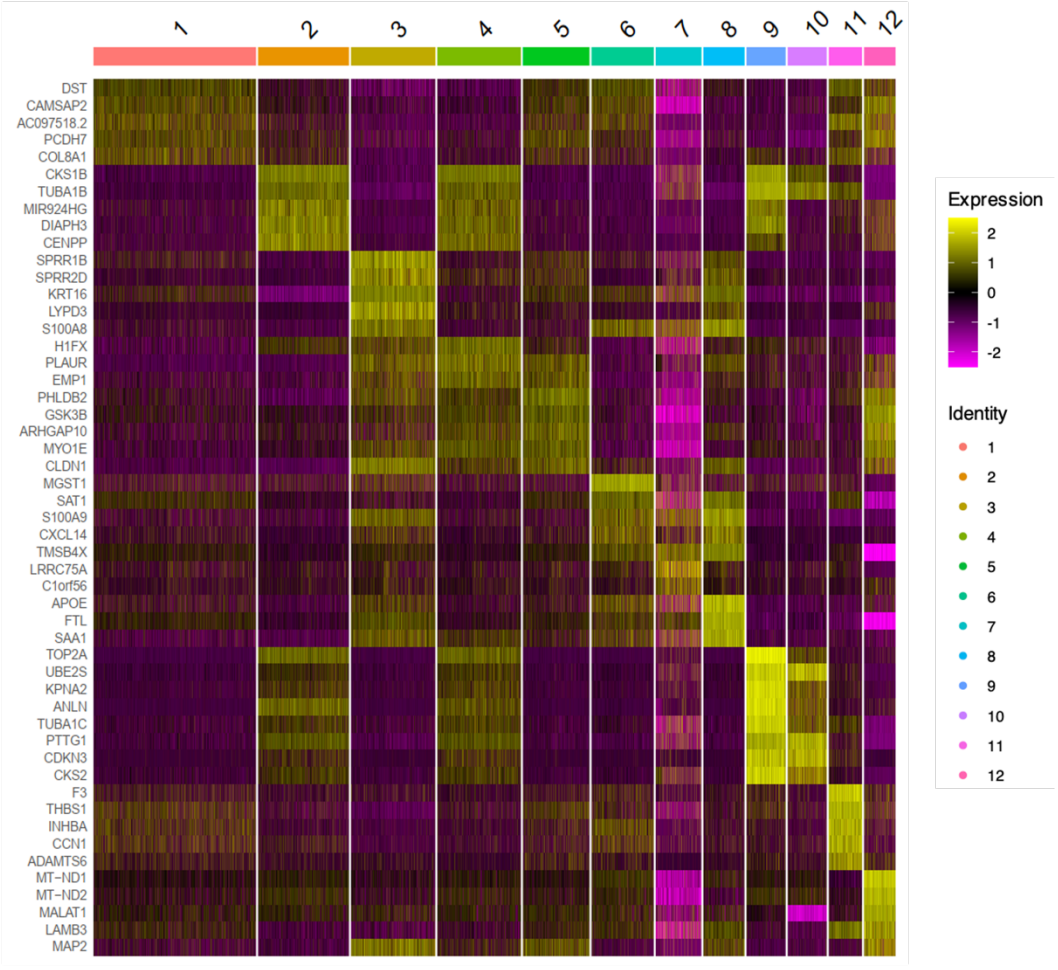


Figure 37: Heatmap of the top 5 markers for each cell cluster. Heatmap showing the top five differentially expressed marker genes for each of the 12 identified cell clusters in HPEK cells following 24-hour treatment with AK23 or mIgG. Each column represents a single cell, grouped by cluster identity (1–12), and each row corresponds to one of the top marker genes for the respective cluster. Gene expression values are scaled and color-coded from low (purple) to high (yellow), highlighting distinct transcriptional profiles among clusters. This heatmap illustrates the heterogeneity of cellular responses and reveals cluster-specific gene expression patterns potentially associated with AK23-induced effects.

Table 6: Top 5 marker genes from each cell cluster and related function annotated by Stringdb

Cluster	Gene	Stringdb annotation
1	*DST*	Microtubule end
1	*CAMSAP2*	
1	*AC097518.2*	
1	*PCDH7*	
1	*COL8A1*	
2	*CKS1B*	RHO GTPases Activate Formins
2	*TUBA1B*	
2	*MIR924HG*	
2	*DIAPH3*	
2	*CENPP*	
3	*SPRR1B*	Keratinization
3	*SPRR2D*	
3	*KRT16*	
3	*LYPD3*	
3	*S100A8*	
4	*CKS1B*	NA
4	*H1FX*	
4	*PLAUR*	
4	*EMP1*	
4	*TUBA1B*	
5	*PHLDB2*	Microtubule cytoskeleton regulation
5	*GSK3B*	
5	*ARHGAP10*	
5	*MYO1E*	
5	*CLDN1*	
6	*MGST1*	Chronic inflammatory response
6	*SAT1*	
6	*S100A8*	
6	*S100A9*	
6	*CXCL14*	
7	*TMSB4X*	Chronic inflammatory response
7	*LRRC75A*	
7	*S100A9*	
7	*C1orf56*	
7	*S100A8*	
8	*APOE*	Chronic inflammatory response
8	*FTL*	
8	*SAA1*	
8	*S100A9*	
8	*S100A8*	
9	*TOP2A*	Cell division
9	*UBE2S*	
9	*KPNA2*	

9	*ANLN*	
9	*TUBA1C*	
10	*PTTG1*	Cell division
10	*UBE2S*	
10	*CDKN3*	
10	*TUBA1B*	
10	*CKS2*	
11	*F3*	Positive regulation of transmembrane receptor protein serine/threonine kinase
11	*THBS1*	
11	*INHBA*	
11	*CCN1*	
11	*ADAMTS6*	
12	*MT-ND1*	Oxidative phosphorylation
12	*MT-ND2*	
12	*MALAT1*	
12	*LAMB3*	
12	*MAP2*	

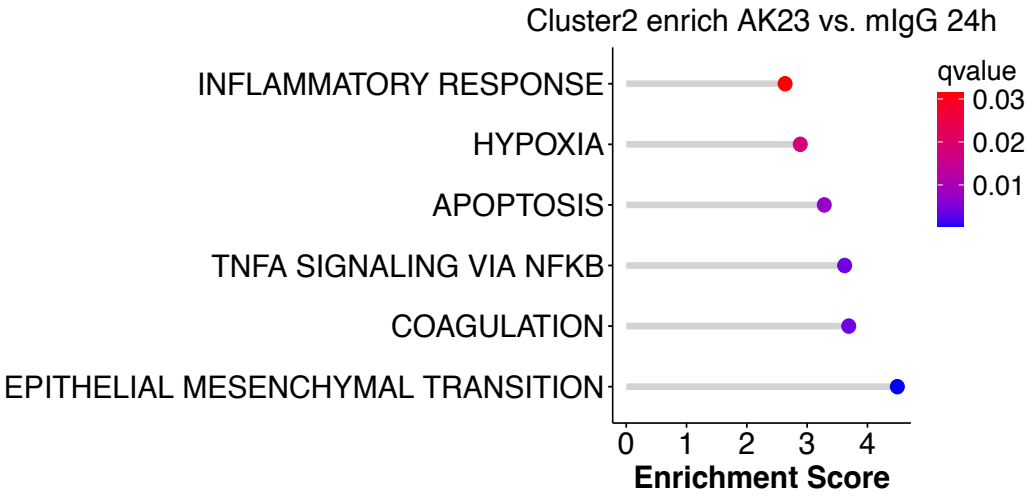


Figure 38: Significantly enriched pathways of Cluster2 AK23 vs. mIgG 24h. Differentially expressed genes were analyzed using GAGE to identify enriched pathways, with a significance threshold of $q < 0.05$.

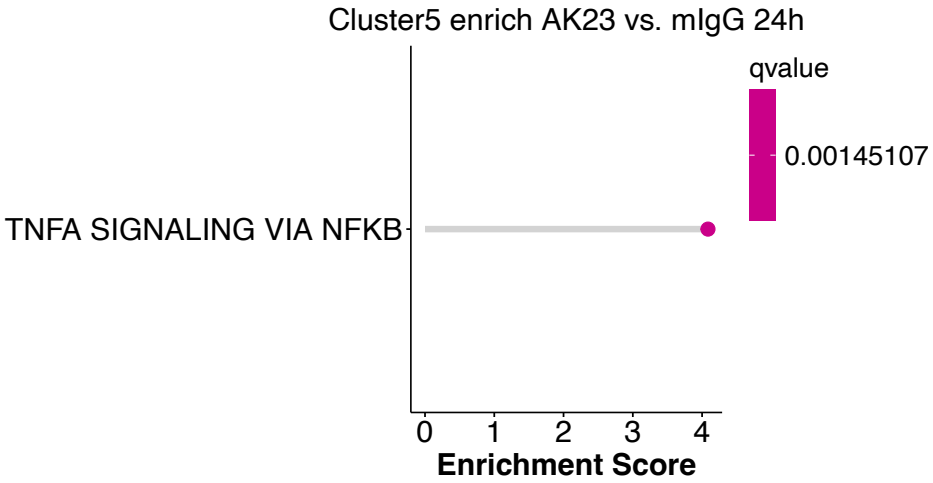


Figure 39: Significantly enriched pathways of Cluster5 AK23 vs. mIgG 24h. Differentially expressed genes were analyzed using GAGE to identify enriched pathways, with a significance threshold of $q < 0.05$.

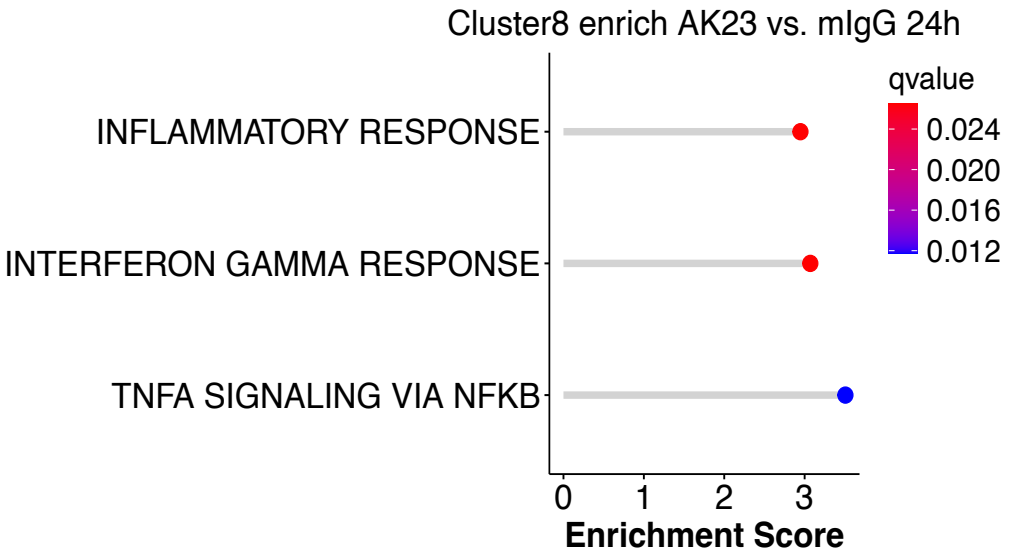


Figure 40: Significantly enriched pathways of Cluster8 AK23 vs. mIgG 24h. Differentially expressed genes were analyzed using GAGE to identify enriched pathways, with a significance threshold of $q < 0.05$.

4.3.3 SingleR Confirms Majority of Sequenced Cells Are Keratinocyte Cells in HPEK

Additionally, Figure 41 illustrates the cell type predictions generated by SingleR, with the majority of cells identified as epithelial cells. This observation aligns with expectations, as keratinocytes are a subtype of epithelial cells.

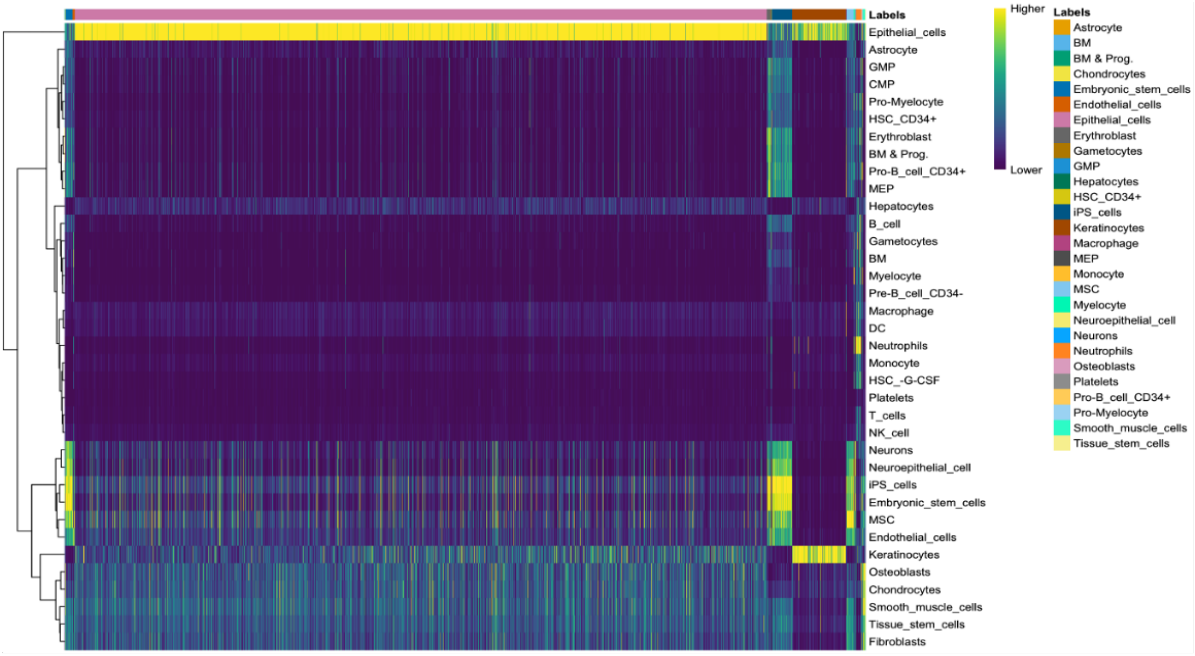


Figure 41: Cell type predictions performed using SingleR. Most cell types are predicted to be epithelial cells, with yellow indicating the highest probability of the corresponding cell type classification.

4.3.4 DA-seq Confirms Subtle Transcriptomics Differences Between AK23 and mIgG 24h Samples in HPEK

Differential abundant analysis shows the minor difference of the cells of mIgG and AK23, only a small portion of cells (3.23%) were identified to be more abundant in the mIgG 24h sample, as shown in Figure 42. While DA-seq identifies regions of significant differences in cell abundance between conditions rather than comparing the total cell counts, this can initially appear contradictory to the UMAP visualization, where the AK23 24h sample shows a larger number of cells in cluster 7 compared to mIgG 24h sample. This discrepancy arises because DA-seq focuses on local neighborhood-based differences in abundance, which may not directly align with global cluster-level patterns observed in UMAP.

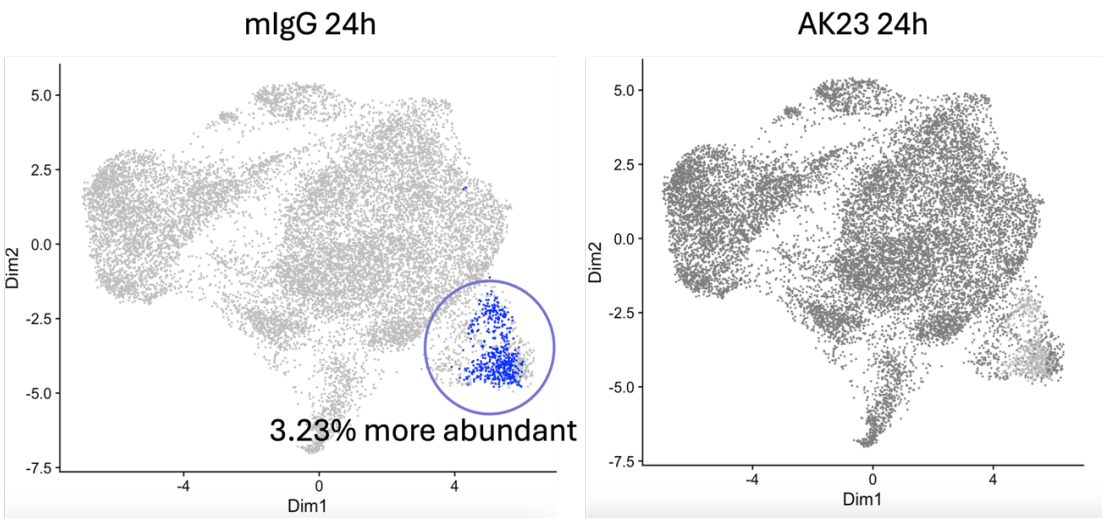


Figure 42: Differentially abundant cell subpopulations by DA-seq. DA-seq analysis identified a specific cell subpopulation that was 3.23% more abundant in the mIgG 24-hour sample relative to the AK23 24-hour sample.

Overall, the scRNA-seq analysis revealed differences in only a few cell clusters in the AK23 24h sample compared to the mIgG 24h sample. This finding may explain the limited number of significant differences observed in the bulk RNA-seq analysis, as the response is likely restricted to a few subpopulations of keratinocyte cells.

The scRNA-seq results further support the findings from bulk RNA-seq, confirming that antibody-induced transcriptomic changes are relatively modest. This consistency across platforms suggests that the observed effects are not artifacts of sequencing modality but rather reflect a genuine biological phenomenon. In particular, the subtle shifts in gene expression at the single-cell level may point to heterogeneous or transient responses among individual cells, rather than a coordinated, robust transcriptional program. Such patterns could indicate that the antibodies primarily modulate signaling or cellular behaviour through mechanisms other than large-scale transcriptional reprogramming—such as epigenetic remodeling, protein-level regulation, or post-transcriptional control. However, to draw more definitive conclusions, additional biological replicates will be necessary to improve the robustness and reproducibility of the single-cell data, and to better resolve these nuanced cellular dynamics.

4.4 HPEK Shotgun Proteomics

Although no antibody-specific responses were detected at the transcriptomic level, it remains possible that consistent responses may be present at the protein level. This discrepancy could be attributed to post-transcriptional regulation, alternative splicing, and other processes influencing the translation of RNA into protein.

4.4.1 Donor Effects Observed of Proteins in HPEK

The initial step in analyzing proteomic data involves an exploratory assessment of sample similarity using Principal Component Analysis (PCA). PCA plots for the shotgun proteomics data are presented in Figure 43.

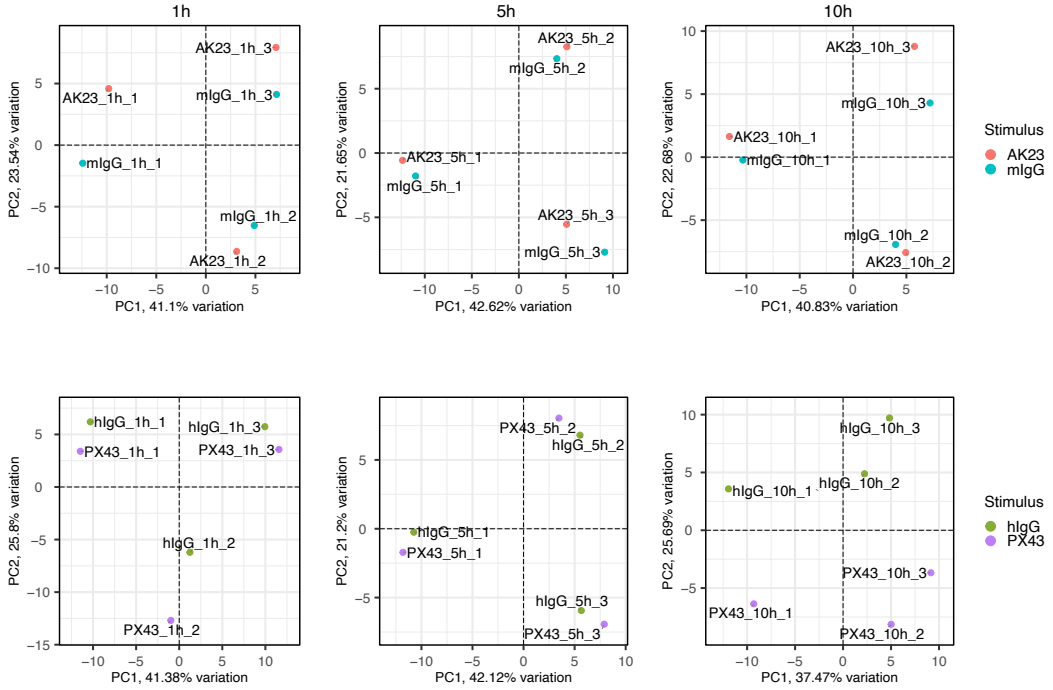


Figure 43: Principal component analysis (PCA) plots for X, Y and Z shotgun proteomics. Samples at 1h, 5h and 10h, including AK23 vs. mIgG and PX43 vs. hIgG. AK23 vs. mIgG and PX43 vs. hIgG have separation mainly at the first principal component (PC1), still, samples at the same time points tend to cluster closer.

Based on the results, although there is a separation between AK23 vs. mIgG and PX43 vs. hIgG along both Principal Component 1 (PC1) and Principal Component 2 (PC2), samples from the same donor tend to cluster more closely together, irrespective of the treatment antibody. It has also been observed that samples from the same time point tend to cluster together,

further supporting the notion that temporal factors exert a stronger influence on sample clustering than antibody-induced variations.

4.4.2 No Significant Protein Alterations in HPEK

Differentially Abundant Proteins (DAPs) were identified using the limma (Linear Models for Microarray Data) package, as described in the Introduction and Methods sections. This approach employs linear modeling combined with empirical Bayes moderation to improve statistical power, particularly in datasets with relatively small sample sizes. Log_2-transformed protein intensity values, generated by the sequencing facility at the University of Bern, were used as input. Linear models were fitted to the data for each protein across the experimental groups. However, no statistically significant DAPs were detected in the comparisons of AK23 vs. mIgG and PX43 vs. hIgG, suggesting that antibody treatment did not result in substantial changes in protein abundance under the conditions tested.

This outcome is consistent with the observation of only minor differences at the transcriptome level, suggesting that the proteomic changes may be subtle or below the detection threshold of the current experimental approach. These findings highlight the potential alignment between transcriptomic and proteomic profiles, where limited transcriptional variation translates to minimal changes in protein abundance.

4.5 HPEK Phosphoproteomics

Although no differentially abundant proteins were identified in the comparisons of PX43 vs. hIgG and AK23 vs. mIgG, there is still the potential for consistent changes at the level of protein phosphorylation. To explore this, phosphoproteomic analysis was performed and analyzed using a methodology analogous to that applied to the shotgun proteomics data, utilizing the limma package for statistical modeling.

4.5.1 Donor Effects Observed of Phosphoproteins in HPEK

Principal Component Analysis (PCA) plots for the initial exploration of the phosphoproteomics data are presented in Figure 44. A distinct separation between the AK23 and mIgG groups, as well as between the PX43 and hIgG groups, is primarily observed along Principal Component

1 (PC1). This component accounts for a substantial proportion of the total variance, ranging from 65% to 90% across the different time points, suggesting that differential treatment effects are the primary source of variation in the data. The high variance explained by PC1 highlights a significant divergence in the global proteomic and phosphoproteomic profiles between the treatment groups. However, it is also noteworthy that samples from the same donor tend to cluster closely together, indicating a degree of individual variability within each treatment condition.

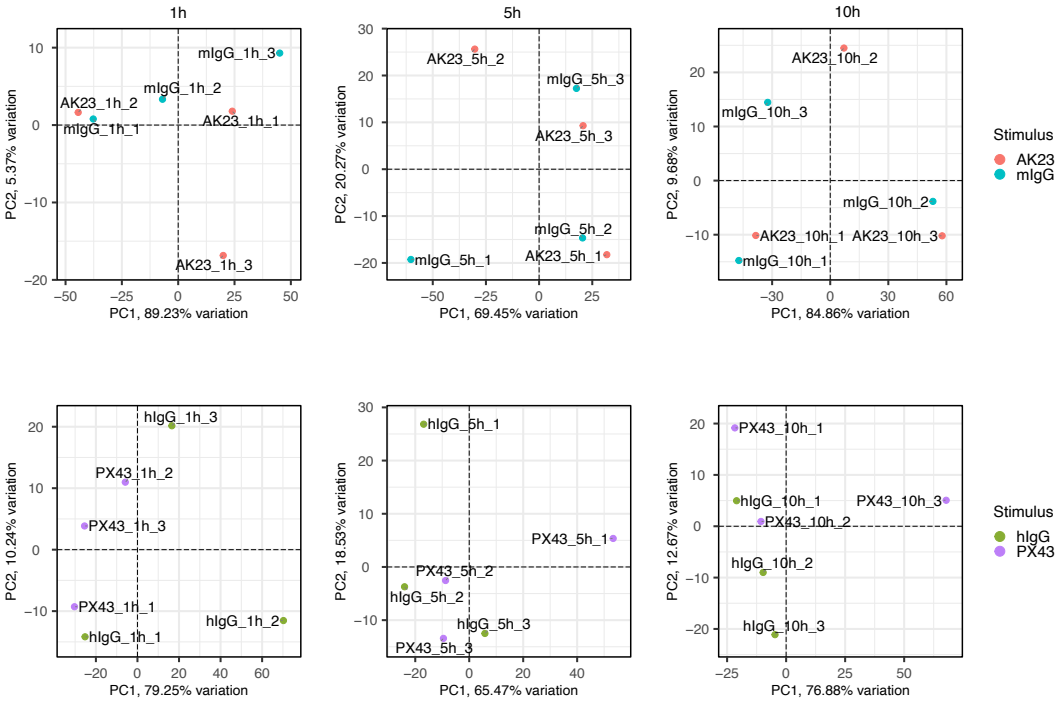


Figure 44: Principal component analysis (PCA) plots for X, Y and Z phosphoproteomics samples at 1h, 5h and 10h, including AK23 vs. mIgG and PX43 vs. hIgG. Separation between AK23 and mIgG, as well as PX43 and hIgG, is mainly observed along Principal Component 1 (PC1).

4.5.2 No Significant Phosphoproteins Alterations in HPEK

Upon further analysis of the Differentially Abundant Phosphorylated Proteins (DAPPs) using the limma method, analogous to the approach employed in shotgun proteomics data analysis, no DAPPs were identified in the comparisons between AK23 vs. mIgG and PX43 vs. hIgG. Although a more distinct separation was observed along Principal Component 1 (PC1) in the PCA plots, these specific comparisons did not reveal significant differences in phosphorylation

levels. This suggests that the treatments, under the conditions examined, may not induce substantial or detectable changes in global phosphorylation patterns.

This result aligns with the minimal differences observed at the transcriptomic level, these findings underscore the potential correlation between transcriptomic and phosphoproteomic profiles, where limited transcriptional variation leads to minimal alterations in protein phosphorylation.

These results align with the minimal differences observed at the transcriptomic level and are further supported by both phosphoproteomic and shotgun proteomics analyses. Across all three modalities, the data consistently reveal modest changes in response to treatment, suggesting a tightly regulated system in which limited transcriptional variation leads to correspondingly subtle shifts in protein abundance and phosphorylation. This concordance between the transcriptome and proteome highlights a potential mechanistic link, where weak or transient upstream signals may not be sufficient to trigger broad transcriptional or post-translational responses. Instead, the cells may maintain a poised state, with only minimal activation across molecular layers—reflecting a conservative response that prioritizes stability and homeostasis over dramatic molecular reprogramming. Such findings emphasize the importance of multi-omics integration in capturing the full scope of cellular behaviour.

4.6 HSOC Bulk RNA-seq

The human skin organ culture (HSOC) model closely recapitulates the in vivo architecture and cellular complexity of human skin, making it a valuable ex vivo system for studying disease mechanisms. In this study, we utilized bulk RNA sequencing (RNA-seq) to assess transcriptomic changes in the HSOC model following treatment with PX43, hIgG, AK23, and mIgG at 5, 10, and 24 hours, as described in the experimental design section. This approach aimed to determine whether these treatments induce consistent transcriptional perturbations over time.

4.6.1 Donor Effects Observed in HSOC

As part of the initial exploratory analysis using principal component analysis (PCA), the most pronounced differentiation between AK23 and mIgG was observed along the first principal

component (PC1), which accounted for 27% to 56% of the total variance. In contrast, the separation between PX43 and hIgG was more distinct along PC1, which explained 22% to 28% of the variance, as shown in Figure 45. Within individual donors, AK23 and mIgG tended to cluster closely, with a slightly more defined clustering observed for PX43 and hIgG samples.

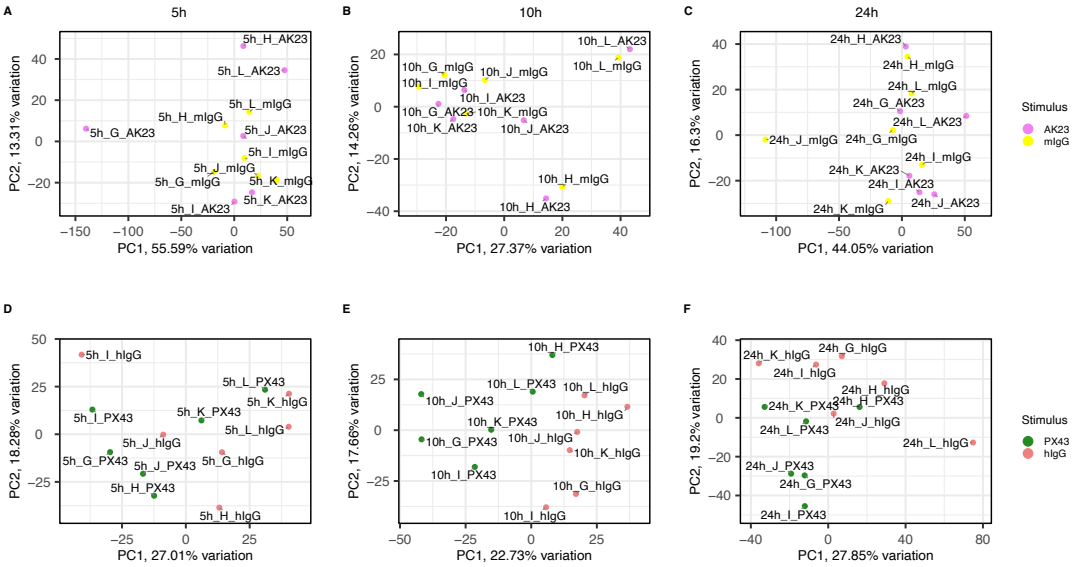


Figure 45: Principal component analysis (PCA) for human skin organ culture samples at 5h, 10h and 24h. The first three plots illustrate the comparisons between AK23 and mIgG, while the last three plots pertain to PX43 versus hIgG. Letters G, H, I, J, K, and L represent biological replicates for the respective samples (AK23, mIgG, PX43, and hIgG).

4.6.2 PX43-Specific Differential Gene Expression Correlates with Split Formation Absence in AK23

The significantly differentially expressed genes (DEGs) are shown in the Figure 46, volcano plots were generated to identify significant differentially expressed genes (DEGs) under the defined thresholds of padj < 0.05 and $|\log_2$ fold change$| > 1$ comparing AK23 with mIgG and PX43 with hIgG across three time points: 5, 10, and 24 hours. Given the substantial number of DEGs identified in the PX43 versus hIgG comparison, only the top DEGs are labeled in the volcano plot to minimize overlap of gene names and enhance the clarity of the display.

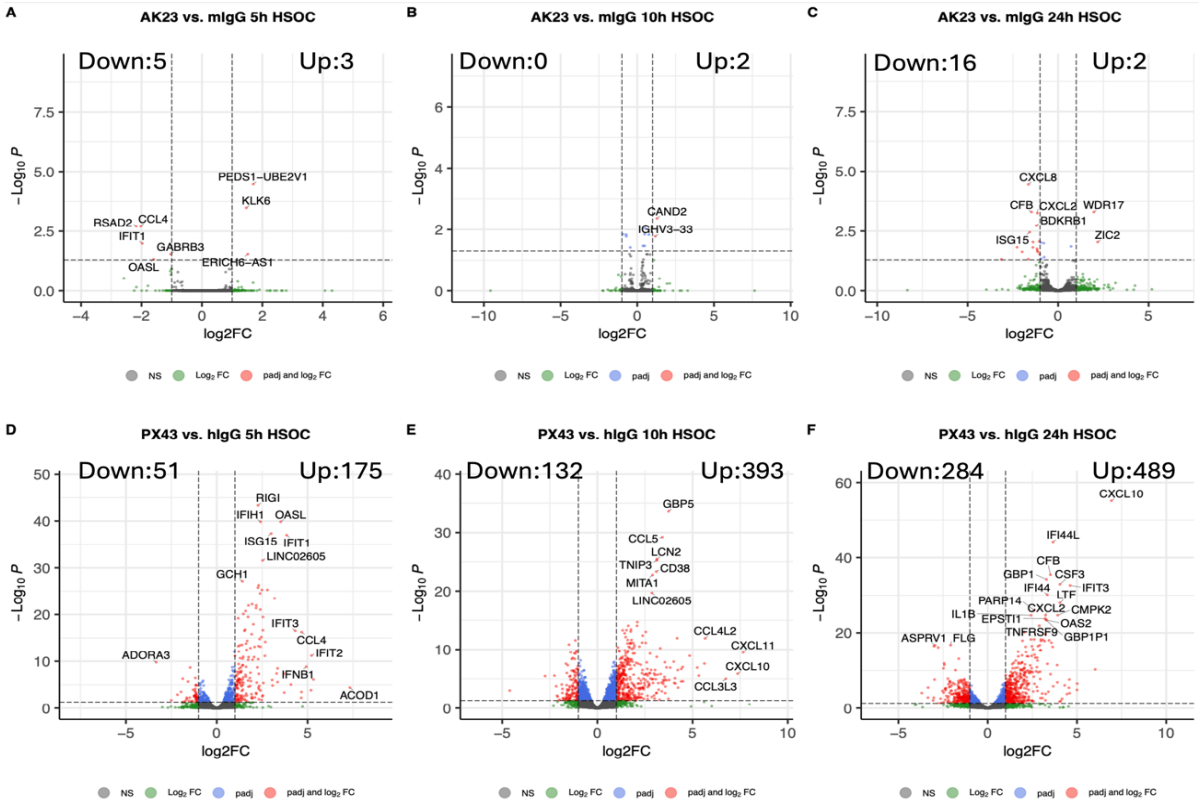


Figure 46: Volcano plots for significant differentially expressed genes (DEGs) in human skin organ culture (HOSC) of AK23 vs. mIgG and PX43 vs. hIgG at time points 5h, 10h and 24h. The first three plots represent the comparison between AK23 and mIgG, revealing a limited number of significant differentially expressed genes (DEGs). In contrast, the last three plots correspond to the comparison between PX43 and hIgG, identifying over 200 DEGs at each time point (5h, 10h, and 24h).

Few significant differentially expressed genes (DEGs) were identified in comparing AK23 to mIgG at 5h, 10h, and 24h. In contrast, many significant DEGs were observed in comparing PX43 to hIgG at the same time points. This pattern aligns closely with the split formation data depicted in Figure 47, generated by Veronika Hartmann from Professor Jennifer Hundt's lab. Specifically, no split formation was observed in AK23 compared to mIgG at 5h, 10h, and 24h, corresponding to the minimal number of significant DEGs. In PX43-treated human skin, split formation increased over time, from 5h to 24h, consistent with a progressive increase in the number of significant DEGs identified in the PX43 versus hIgG comparison.

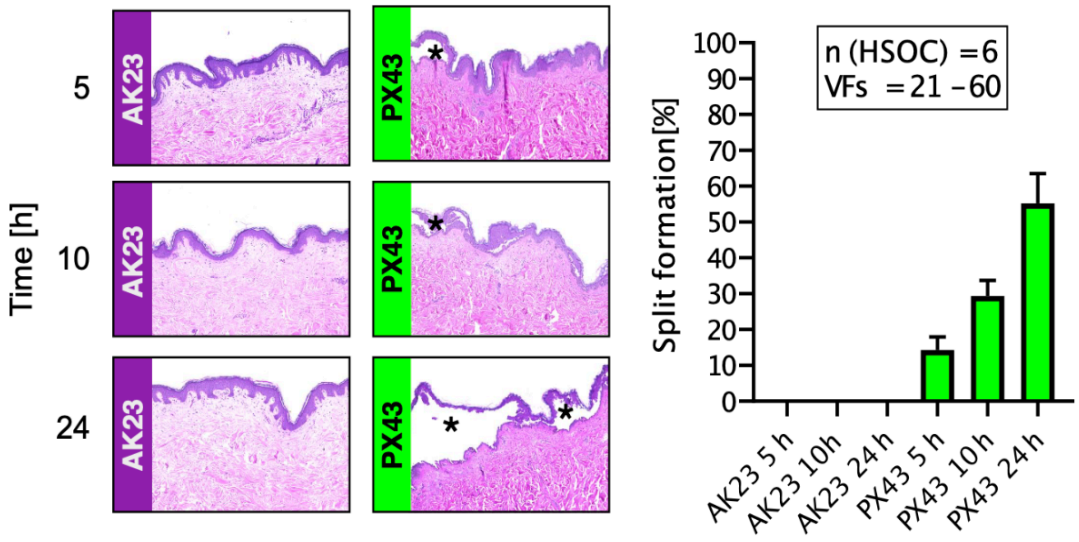


Figure 47: Split formation in human skin following scFv PX43 and AK23 treatment. Hematoxylin and eosin (H&E) staining of human skin revealed no evidence of split formation in samples treated with AK23, which targets DSG3. In contrast, split formation was observed to increase over time in samples treated with PX43, which targets both DSG3 and DSG1. VFs (visual fields). The HE staining and bar plots are generated by Veronika Hartmann from Professor Jennifer Hundt's lab.

Transcriptomic Responses Associated With Split Formation Show A Positive Correlation With Keratinocyte Detachment And Acute Skin Wounding

Besides, the transcriptome perturbation induced by PX43, compared to hIgG in HSOC, exhibits a moderate positive correlation with keratinocyte cell detachment, as reported in the publicly available data from Crispin et al. (2012) [162]. As shown in Figure 48, the correlation coefficient increased over time, from 0.31 at 5 hours to 0.45 at 10 hours, and further to 0.53 at 24 hours. This progressive increase in correlation suggests that PX43 treatment may contribute to the disruption of keratinocyte cell adhesion, potentially indicating a time-dependent effect on keratinocyte responses.

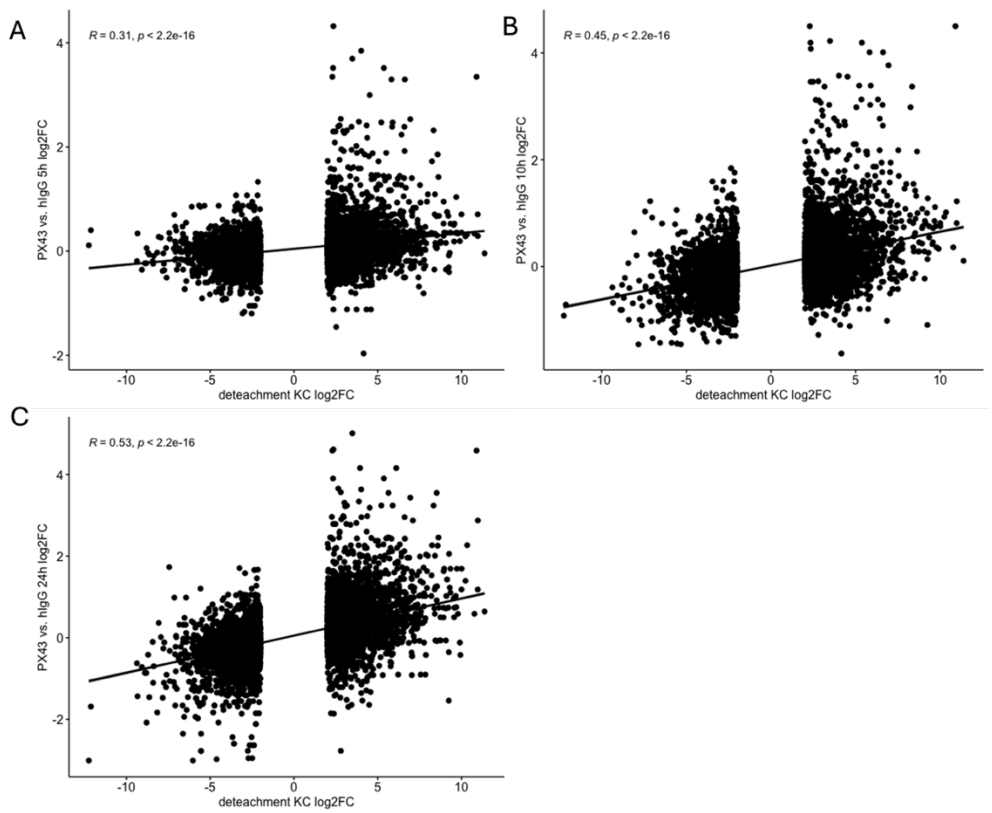


Figure 48: PX43 transcriptome response linked to cell detachment. Correlation analysis between the PX43 vs. hIgG transcriptomic data in HSOC at 5h, 10h, and 24h and the keratinocyte cell detachment RNA-seq data from Crispin et al. (2012) [162] revealed a moderate positive correlation.

Furthermore, a study by Nuutila et al. (2012) [163] on skin wound healing revealed a moderate positive correlation of 0.34 between differentially expressed genes (DEGs) at day 3 post-wounding and the transcriptomic changes induced by PX43 compared to hIgG at 24 hours in the HSOC model, as shown in Figure 49. In their study, biopsies were taken at acute wounding (immediately after removal of the split-thickness skin graft), day 3, and day 7, with gene expression profiles compared against intact skin as a normal control. The authors identified matrix metalloproteinase 1 (MMP1) and keratin 2 (KRT2) as the most significantly altered genes on day 3 relative to intact skin. Based on these findings, incorporating additional time points, both earlier (e.g., 15 minutes, 0.5 hours, 1 hour) and later (e.g., 72 hours), in our model may provide deeper insights and a more comprehensive understanding of PV antibody effects in the HPEK and HSOC models.

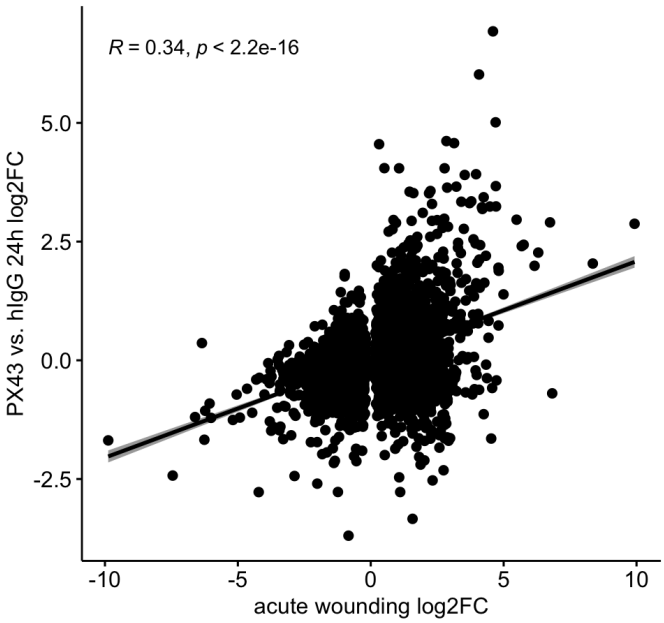


Figure 49: PX43 transcriptome response linked to skin wound. Correlation analysis between the PX43 vs. hIgG transcriptomic data in HSOC at 24h and the skin wound healing RNA-seq data from Nuutila et al. (2012) [163] revealed a moderate positive correlation.

Minimal Transcriptomic Changes Induced by AK23 in HSOC

For AK23 compared to mIgG at 5h in the human skin organ culture (HSOC) model, there are 3 significantly upregulated genes (*PEDS1-UBE2V1*, *KLK6*, *ERICH6-AS1*) and 5 significant downregulated genes (*CCL4*, *GABRB3*, *OASL*, *IFIT1*, *RSAD2*). AK23 compared to mIgG at 10h, there are 2 significant upregulated genes (*CAND2*, *IGHV3-33*), AK23 compared to mIgG at 24h has 2 significant upregulated genes (*WDR17*, *ZIC2*) and 16 significant downregulated genes (*BDKRB1*, *BIVM-ERCC5*, *CCL3*, *CFB*, *COL22A1*, *CSF3*, *CXCL2*, *CXCL3*, *CXCL8*, *G0S2*, *IRF1*, *ISG15*, *NINJ1*, *OASL*, *SAA2*, *TMEM265*).

- *PEDS1-UBE2V1* (PEDS1-UBE2V1 Readthrough) regulates cell autophagy [164].
- *KLK6* (Kallikrein Related Peptidase 6) plays an important role in the proliferation of keratinocyte cells [165].
- ERICH6-AS1 (ERICH6 antisense RNA 1) is located near the ERICH6 gene and has been reported to be related to necroptosis [166].
- CCL4 (C-C motif chemokine Ligand 4) plays an important role in the immune response [167].

- GABRB3 (Gamma-Aminobutyric Acid Type A Receptor Subunit Beta3) upregulated in HaCaT keratinocytes treated with lipopolysaccharide (LPS) and cyclosporine A (CSA) [168].
- OASL (2'-5'-Oligoadenylate Synthetase Like) contributes to epidermal keratinocyte proliferation [169].
- IFIT1 (Interferon induced protein with Tetratricopeptide repeats 1) is an important gene in the innate immune system [170].
- RSAD2 (Radical S-adenosyl methionine domain containing 2) is an antiviral gene overexpressed in the lesional psoriatic skin [171].
- CAND2 (Cullin Associated and Neddylation Dissociated 2) plays critical roles in protein degradation via ubiquitination pathways [172].
- IGHV3-33 (Immunoglobulin Heavy Variable 3-33) has been reported to be correlated with the adaptive immune response [173].
- WDR17 (WD repeat domain 17) regulates cell proliferation, cell cycle progression and apoptosis [174].
- ZIC2 (Zic Family Member 2) expression was higher in most cancer tissues than in adjacent normal tissue [175].
- BDKRB1 (Bradykinin Receptor B1) is beneficial for the wound healing [176].
- BIVM-ERCC5 (BIVM-ERCC5 readthrough) is a DNA repair endonuclease that plays a crucial role in nucleotide excision repair (NER) [177].
- CCL3 (C-C motif chemokine Ligand 3) promotes cutaneous wound healing by recruiting mouse macrophages [178].
- CFB (Complement Factor B) plays a crucial role in the immune response by participating in the activation of the complement system [179].
- COL22A1 (Collagen type XXII Alpha 1 chain) is an important member of the extracellular matrix [180].
- CSF3 (Colony Stimulating Factor 3) belongs to the chemokines that can be induced by IL-17A [181].

- CXCL2 (C-X-C motif chemokine Ligand 2) acts as a chemokine that mediates the recruitment and activation of neutrophils during inflammatory responses [182].
- CXCL3 (C-X-C motif chemokine Ligand 3) acts as a chemokine that attracts immune cells to sites of inflammation [183].
- CXCL8 (C-X-C motif chemokine Ligand 8) acts as a chemokine and recruits neutrophils to sites of injury [184].
- G0S2 (G0/G1 switch 2) encodes the protein that plays an important role in proliferation and apoptosis [185].
- IRF1 (Interferon Regulatory Factor 1) plays a significant role in halting cell proliferation and contributes to the regulation of immune responses [186].
- ISG15 (ISG15 ubiquitin like modifier) promotes keratinocyte proliferation in psoriasis by regulating the HIF-1α signaling pathway and influencing cell cycle progression [187].
- NINJ1 (Ninjurin 1) encodes a membrane protein that mediates plasma membrane rupture during specific types of cell death, such as pyroptosis and plays a role in inflammation, vascular development, and cellular adhesion [188].
- SAA2 (Serum Amyloid A2) is an acute-phase gene with a transcriptional advantage over SAA1 under cytokine stimulation, contributing prominently to inflammation, but lacks glucocorticoid responsiveness due to an inactive glucocorticoid response element [189].
- TMEM265 (Transmembrane Protein 265) is a gene associated with tumor-infiltrating immune cells in nasopharyngeal carcinoma, playing a role in the immune response and influencing patient prognosis [190].

The top genes identified in this list are closely linked to critical biological processes, particularly in immune response regulation, cell proliferation, and tissue repair. Many of these genes, such as *CCL4*, *RSAD2*, and *CXCL8*, are involved in mediating immune cell recruitment, inflammation, and antiviral defense. Others, like *G0S2*, *WDR17*, and *ZIC2*, are implicated in regulating cell cycle progression, apoptosis, and cancer-related processes. Additionally, genes

like *CFB*, *ISG15*, and *SAA2* play key roles in inflammatory responses, tissue regeneration, and immune modulation.

Given the limited number of differentially expressed genes in the AK23 versus mIgG comparison and the absence of consistent transcriptomic alterations across time points, these findings suggest that AK23 elicits a modest and potentially context-dependent transcriptional response in the HSOC model. Further validation will be necessary to confirm the biological relevance of these observations.

Robust Transcriptomic Alterations Induced by PX43 in HSOC

Numerous significant differentially expressed genes (DEGs) were identified when comparing Px43 to hIgG across all three time points (5h, 10h, and 24h). To highlight key findings, the top 5 upregulated and top 5 downregulated genes at each time point will be presented. Top 5 upregulated genes (*ACOD1*, *CCL3L3*, *IFIT2*, *CCL4L2*, *IFNB1*) and top 5 downregulated genes (*ADORA3*, *FCGR1A*, *ARHGEF38*, *FZD2*, *ADGRG3*) at 5h, top 5 upregulated genes (*CXCL11*, *CXCL10*, *CCL3L3*, *CCL4L2*, *CCL4*) and top 5 downregulated genes (*ADORA3*, *PPDPFL*, *APOB*, *CHAD*, *CCDC187*) at 10h and top 5 upregulated genes (*CXCL10*, *CXCL11*, *RSAD2*, *CCL3L3*, *IFIT3*) and top 5 downregulated genes (*CHAD*, *CFAP73*, *VSIG4*, *TMIGD3*, *DSP-AS1*) at 24h. Notably, among the top five upregulated genes, CCL3L3 was consistently upregulated at 5, 10, and 24 hours post-treatment in PX43-treated HSOC samples. Given its established role in immune cell recruitment and the modulation of inflammatory responses, CCL3L3 may contribute to the immune-related mechanisms associated with PX43 treatment, particularly in samples that exhibit split formation, as independently validated.

- *ACOD1* (Aconitate Decarboxylase 1) encodes a key enzyme in itaconate biosynthesis that modulates immune responses, oxidative stress, and antiviral defense through metabolic reprogramming [191].
- CCL3L3 (C-C motif Chemokine Ligand 3 Like 3) is involved in immune signaling, contributing to the recruitment of immune cells and the modulation of inflammatory responses [192].

- IFIT2 (Interferon Induced Protein with Tetratricopeptide Repeats 2) is an interferon-stimulated gene that plays a critical role in antiviral defense by inhibiting viral replication and modulating immune responses to viral infections [193].
- CCL4L2 (C-C Motif Chemokine Ligand 4 Like 2) is a chemokine involved in modulating immune cell recruitment and inflammation, with its expression associated with milder forms of psoriasis [194].
- IFNB1 (Interferon Beta 1) encodes interferon-beta, a type I interferon that plays a key role in the immune response by activating signaling pathways that modulate inflammation and immune cell function, particularly in response to viral infections and other stimuli [195].
- ADORA3 (Adenosine A3 Receptor) encodes the adenosine A3 receptor, which mediates various cellular responses, including modulation of inflammation, immune cell activation, and regulation of signaling pathways such as those involved in cancer progression [196].
- FCGR1A (Fc Gamma Receptor Ia) is a receptor on immune cells that plays a crucial role in recognizing and responding to immune complexes, contributing to the body's defense mechanisms [197].
- ARHGEF38 (Rho Guanine Nucleotide Exchange Factor 38) plays a key role in regulating cell signaling pathways involved in cell migration, adhesion, and cytoskeletal dynamics [198].
- FZD2 (Frizzled Class Receptor 2) regulates the proliferation and apoptosis of keratinocytes [199].
- ADGRG3 (Adhesion G Protein-Coupled Receptor G3) plays a critical role in modulating immune responses, keratinocyte proliferation, and differentiation [200].
- CXCL11 (C-X-C Motif Chemokine Ligand 11) encodes a chemokine that binds to the CXCR3 receptor, playing a crucial role in immune responses by recruiting activated T cells, particularly type 1 T cells, to sites of inflammation, and contributing to the amplification of inflammatory processes in various dermatoses [201].

- CXCL10 (C-X-C Motif Chemokine Ligand 10) encodes a chemokine that binds to the CXCR3 receptor, playing a pivotal role in immune regulation by recruiting type 1 T helper cells to sites of inflammation [202].
- CCL3L3 (C-C Motif Chemokine Ligand 3 Like 3) encodes a chemokine involved in immune signaling, playing a significant role in recruiting and activating immune cells, particularly during viral co-infections [203].
- CCL4 (C-C motif chemokine ligand 4) has a strong expression during the inflammatory phase of mouse wound healing [204].
- PPDPFL (Pancreatic Progenitor Cell Differentiation and Proliferation Factor Like) encodes a protein potentially involved in modulating immune responses and tissue remodeling, with lower expression linked to stronger glucocorticoid responsiveness in conditions like chronic rhinosinusitis with nasal polyps [205].
- APOB (Apolipoprotein B) encodes apolipoprotein B, a key component of low-density lipoproteins (LDL) that is critical for lipid transport and metabolism, and its expression in keratinocytes is linked to epidermal differentiation and lipid homeostasis in the skin [206].
- CHAD (Chondroadherin) encodes a leucine-rich repeat protein that plays a key role in cartilage by binding collagen type II, influencing collagen fibril organization, and mediating cell-matrix interactions critical for chondrocyte signaling and extracellular matrix integrity [207].
- CCDC187 (Coiled-Coil Domain Containing 187) encoding proteins that are involved in intercellular transmembrane signal transduction and genetic signal transcription [208].
- RSAD2 (Radical S-adenosyl methionine Domain containing 2) encodes viperin, an interferon-stimulated protein that plays a critical role in antiviral defense by disrupting viral replication, modulating immune signaling, and influencing lipid metabolism [209].

- IFIT3 (Interferon Induced Protein with Tetratricopeptide Repeats 3) encodes an interferon-induced protein that plays a key role in antiviral defense, immune regulation, and inflammation, with elevated expression in psoriatic lesions [210].
- CFAP73 (Cilia and Flagella Associated Protein 73) encodes a cilia-associated protein critical for the structural integrity and motility of motile cilia, and its dysfunction affects cilia assembly, microtubule architecture, and cilia-independent cellular processes [211].
- VSIG4 (V-set and Immunoglobulin domain containing 4) encodes a transmembrane receptor that plays a crucial role in immune homeostasis by regulating macrophage phagocytosis and suppressing complement activation and T-cell responses [212].
- TMIGD3 (Transmembrane and Immunoglobulin Domain containing 3) encodes a protein that suppresses osteosarcoma (OS) cell proliferation, tumor formation, and metastasis by inhibiting NF-κB activity [213].
- DSP-AS1 (DSP Antisense RNA 1) is a long non-coding RNA (lncRNA) that modulates the expression of the DSP gene, which encodes desmoplakin, a protein integral to desmosomes that is essential for maintaining cell-cell adhesion in both cardiac and epithelial tissues [214].

The top genes identified are closely associated with critical cellular processes, immune responses, and inflammatory pathways. Many of these genes, such as *CCL3L3*, *IFIT2*, and *RSAD2*, are involved in modulating immune cell recruitment, antiviral defense, and immune regulation. Others, like *DSP-AS1* and *TMIGD3*, play significant roles in maintaining cell-cell adhesion, tissue remodeling, and regulating disease progression. Additionally, genes like *APOB* and *CHAD* are linked to metabolic processes and extracellular matrix integrity, highlighting the diverse biological functions that contribute to disease pathogenesis, particularly in inflammatory and immune-mediated conditions.

4.6.3 Split Formation Correlates with Divergent Inflammatory and Apoptotic Pathway Regulation: Sustained Upregulation in PX43, Downregulation in AK23

Pathway enrichment analysis facilitates the interpretation of bulk RNA-seq results from the HSOC model by providing insights beyond individual gene-level changes, as genes typically function in coordinated modules or biological pathways.

The significantly enriched pathways in the human skin organ culture (HSOC) were identified using the gage package, with a threshold of $q < 0.05$, based on the analysis of differentially expressed genes and their corresponding log2 fold changes, and are presented in the dot plot or heatmap that include all the time points.

Compared to mIgG, AK23 treatment does not result in consistent enrichment or suppression of specific pathways across time points in the HSOC model. As illustrated in Figure 50, the top differentially enriched pathways appear to be primarily associated with temporal changes rather than stimulus-specific effects. Within each time point, both AK23 and mIgG biological replicates display similar pathway enrichment profiles, suggesting that AK23 does not elicit a distinct or robust transcriptomic signature at the pathway level in this context.

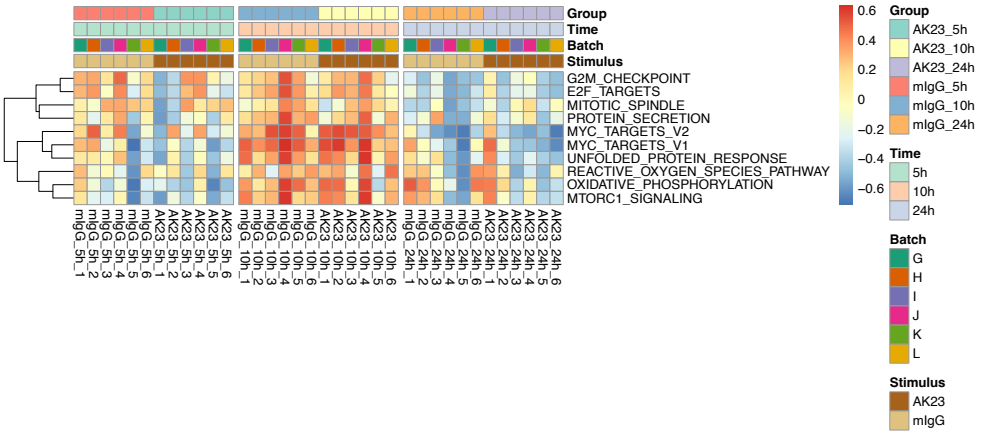


Figure 50: Pathway enrichment profiles of HSOC samples treated with AK23 and mIgG across time points. Heatmap showing the enrichment scores of hallmark pathways in HSOC samples treated with AK23 or mIgG at 5, 10, and 24 hours. Each column represents an individual biological replicate, grouped by treatment and time. The rows correspond to selected hallmark gene sets from MSigDB. While distinct temporal patterns are observed in pathway activity, no consistent stimulus-specific pathway changes are evident. AK23- and mIgG-treated samples display comparable enrichment profiles within each time point, suggesting a limited transcriptional impact of AK23 at the pathway level.

Compared to hIgG, PX43 consistently induces consistent significant upregulation of pathways associated with the inflammatory response, including IL6_JAK_STAT3, TNFα signaling via NF-

κB and interferon alpha/gamma responses, across all three time points (5h, 10h, and 24h), as shown in Figure 51, individual comparison for 5h, 10h and 24h can be find in the supplemental Figures 10-12.

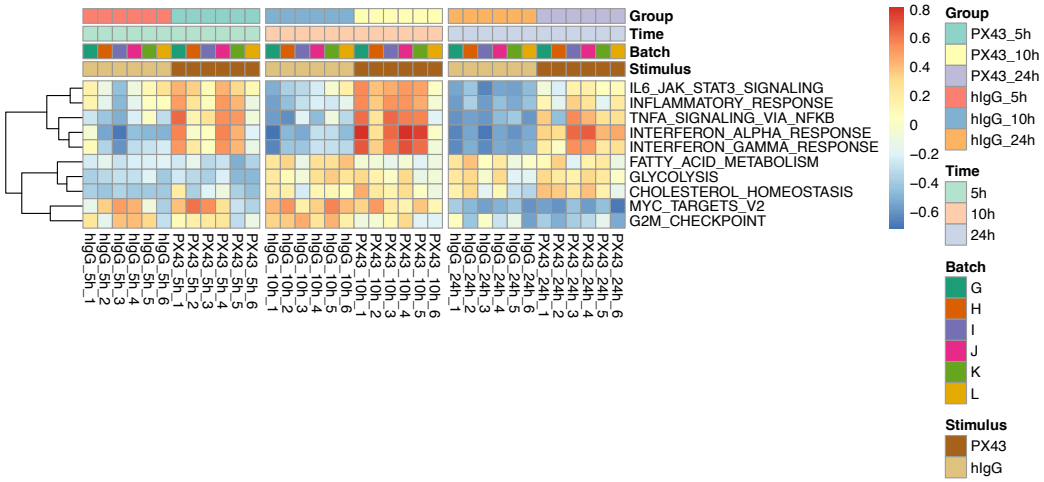


Figure 51: Pathway enrichment profiles of HSOC samples treated with PX43 and hIgG across time points. Heatmap displaying enrichment scores for selected hallmark pathways in HSOC samples treated with PX43 or hIgG at 5, 10, and 24 hours. Columns represent individual biological replicates, grouped by treatment and time, while rows correspond to hallmark gene sets from MSigDB. Compared to hIgG controls, PX43-treated samples show stronger and more consistent upregulation of immune- and inflammation-related pathways, including IL6/JAK/STAT3 signaling, TNFα signaling via NF-κB, and interferon responses, particularly at later time points. These results suggest that PX43 elicits a robust immune-modulatory transcriptomic response in the HSOC model.

Besides, pathway analysis of apoptosis in the HSOC bulk RNA-seq data revealed a significant enrichment of apoptotic pathway in the PX43-treated group compared to hIgG control at all time points (5 h, 10 h, and 24 h), with a q-value < 0.05. In contrast, the AK23-treated group showed no significant enrichment at 5 h and exhibited a decrease in apoptotic pathway at 10 h and 24 h relative to mIgG control, as shown in Table 9.

Table 7: Enrichment of the apoptosis pathway in HSOC bulk RNA-seq data

	5h	10h	24h
PX43 vs. hIgG	**3.64**	**3.35**	**4.31**
	0.001	**0.003**	**0.00006**
AK23 vs. mIgG	0.89	-1.6	**-3.34**
	0.52	0.35	**0.002**

* Values represent Enrichment Score (top) and padj (bottom). **Bold** indicates significance: padj < 0.05.

These findings confirm that PX43 treatment, which induces split formation, is also associated with the activation of apoptosis. In contrast, AK23 treatment, which does not lead to split

formation, shows no significant evidence of apoptotic activity. This suggests a potential correlation between the occurrence of split formation and the induction of apoptosis in the HSOC system.

4.6.4 PX43-Specific Upregulation of Inflammatory Transcription Factors, AK23 Exhibits No Consistent Regulation

Given the large number of significantly differentially expressed genes in PX43 relative to hIgG within the HSOC, it would be valuable to identify the top enriched transcription factors, as this may reveal novel key regulators underlying the observed transcriptional changes.

Transcription factors with significantly altered activity were inferred using the decoupleR framework. Top-ranking TFs were selected and visualized based on a q value threshold of < 0.05.

In the comparison between AK23 and hIgG, as illustrated in the supplemental Figures 13-15, no transcription factors (TFs) were commonly upregulated across all three time points (5h, 10h, and 24h). Similarly, no TFs were consistently downregulated across these time points.

In comparing PX43 and hIgG (Figures 52-54), the transcription factors *IRF3*, *NFKB*, *NFKB1*, and *REL* were commonly upregulated across all three time points (5h, 10h, and 24h). These upregulated TFs align with the pathway enrichment analysis results, identifying the TNFα signaling via the NFκB pathway as commonly upregulated.

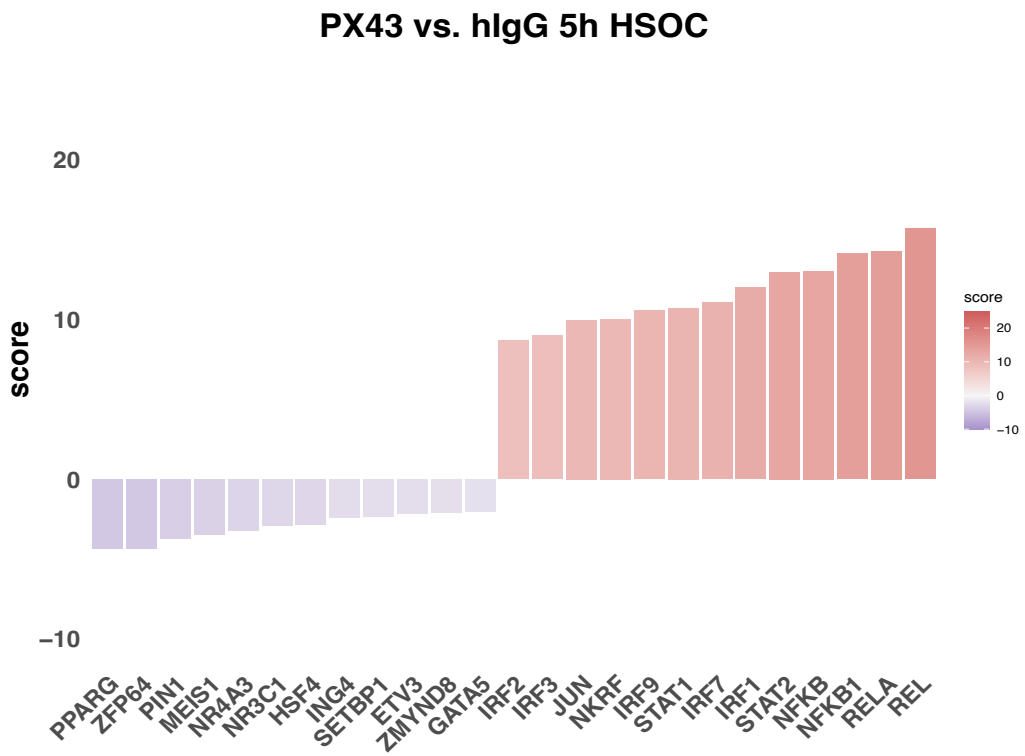


Figure 52: Key differentially expressed transcription factors (TFs) in human skin organ culture (HSOC) at 5h for PX43 vs. hIgG. Top transcription factors were inferred using decoupleR (q-value < 0.05), based on differentially expressed genes and their corresponding log2 fold-change values. Predicted transcription factor activity scores are color-coded, with purple indicating negative activity scores and red indicating positive activity scores.

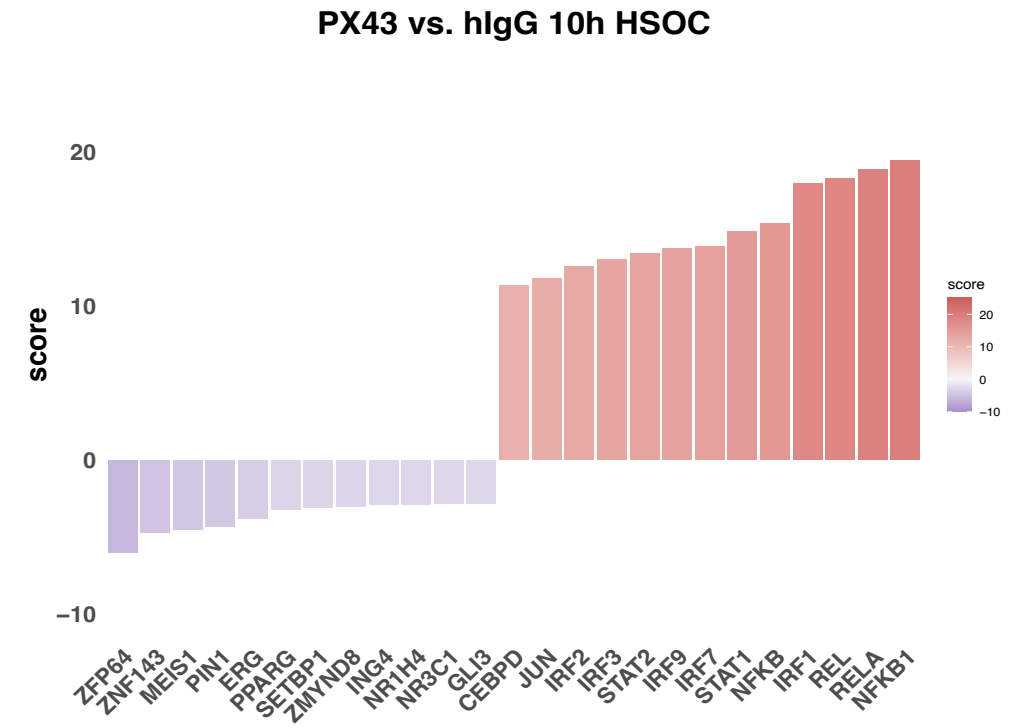


Figure 53: Key differentially expressed transcription factors (TFs) in human skin organ culture (HSOC) at 10h for PX43 vs. hIgG. Top transcription factors were inferred using decoupleR (q-value < 0.05), based on differentially expressed genes and their corresponding log2 fold-change values. Predicted transcription factor activity scores are color-coded, with purple indicating negative activity scores and red indicating positive activity scores.

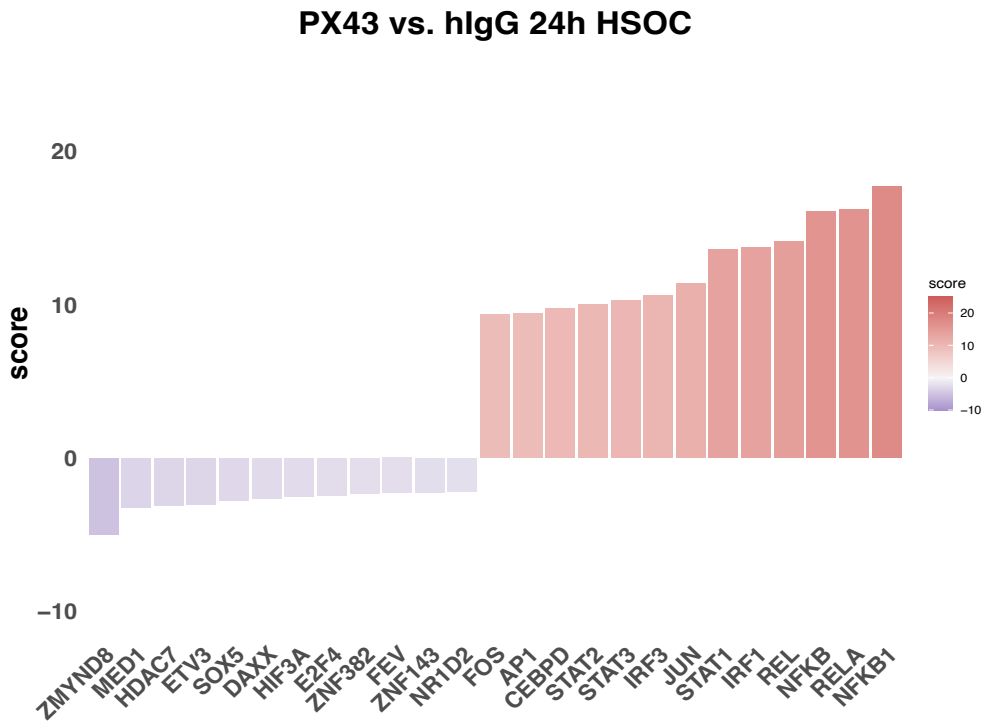


Figure 54: Key differentially expressed transcription factors (TFs) in human skin organ culture (HSOC) at 24h for PX43 vs. hIgG. Top transcription factors were inferred using decoupleR (q-value < 0.05), based on differentially expressed genes and their corresponding log2 fold-change values. Predicted transcription factor activity scores are color-coded, with purple indicating negative activity scores and red indicating positive activity scores.

4.7 HSOC Shotgun Proteomics

4.7.1 HSOC Shogun Proteomics Samples Showed Distinct PC1-Driven Separation in PCA Plot

An initial exploratory analysis of the shotgun proteomics data from the human skin organ culture (HSOC) model was performed using principal component analysis (PCA), as shown in Figure 55.

The PCA plot illustrates the clear separation of samples by treatment groups (PX43, hIgG) and their distinction from the control group (T0). This separation highlights the differential protein expression patterns attributable to the treatments in comparison to the control (T0).

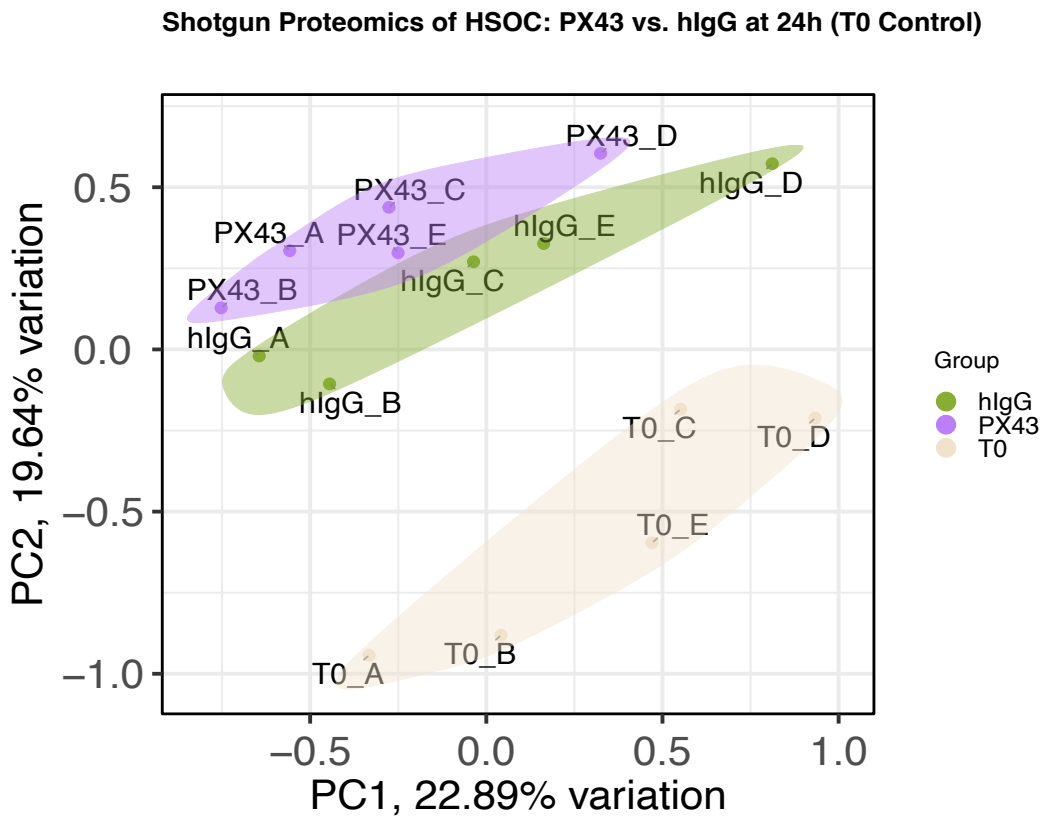


Figure 55: Shotgun proteomics analysis of HSOC model: PX43 vs. hIgG at 24h with T0 as control. Samples cluster distinctly by treatment group, with PX43 (purple), hIgG (green), and untreated T0 controls (tan) forming separate ellipses. The first two principal components explain 22.89% and 19.64% of the variance, respectively. The separation between PX43 and hIgG treatment groups is primarily driven by variation along the first principal component (PC1).

4.7.2 Proteomics Data Correlate Well with Bulk RNA-seq in HSOC

The significantly differentially abundant proteins (DAPs) identified in the human skin organ culture (HSOC) model using shotgun proteomics are summarized in Table 8, based on the thresholds of adjusted p-value (padj < 0.05) and absolute log2 fold change (|log2FC| > 1). At 24h, 4 proteins were significantly upregulated in PX43 compared to hIgG. In PX43 compared to the control (T0), 22 proteins were upregulated and 3 were downregulated. Similarly, in hIgG compared to the control (T0), 12 proteins were upregulated and 1 was downregulated.

Table 8: Significant differentially abundant proteins in HSOC shotgun proteomics

Comparison	Upregulated	Downregulated
PX43 vs. hIgG	S100A9, S100A8, ADAMTS4, SERPINB4	None
PX43 vs. T0	S100A9, S100A8, SERPINB4, SOD2, NAMPT, KRT6A, STEAP4, KRT16, KRT75, JUNB, SLC2A1, CDH3, CENPF, IGFBP3, ADAMTS4, CYP51A1, SERPINB1, SPRR2E, CHI3L2, HMGCS1, PTX3, KRT6C	LGMN, SERPINF2, IL37
hIgG vs. T0	S100A9, S100A8, KRT6A, KRT75, KRT16, CENPF, TIMP1, JUNB, SERPINB4, STEAP4, PTX3, KRT6C	LGMN

Figure 56 presents a volcano plot illustrating the differential abundant proteins between PX43 and hIgG at 24h.

- S100A9 (S100 calcium binding protein A9) regulates psoriatic skin and joint disease by modulating inflammatory pathways and neutrophil infiltration [215].
- S100A8 (S100 calcium binding protein A8) modulates inflammatory responses in atopic dermatitis and reduces skin barrier protein expression [216].
- ADAMTS4 (ADAM Metallopeptidase with Thrombospondin Type 1 Motif 4) belongs to extracellular proteases that mediate cellular interactions and cell signaling via the modulation of adhesion and the cleavage of cell surface protein ectodomains and extracellular matrix molecules [217].
- SERPINB4 (Serpin Family B Member 4) promotes keratinocyte inflammation in psoriasis by activating the p38 MAPK signaling pathway [218].

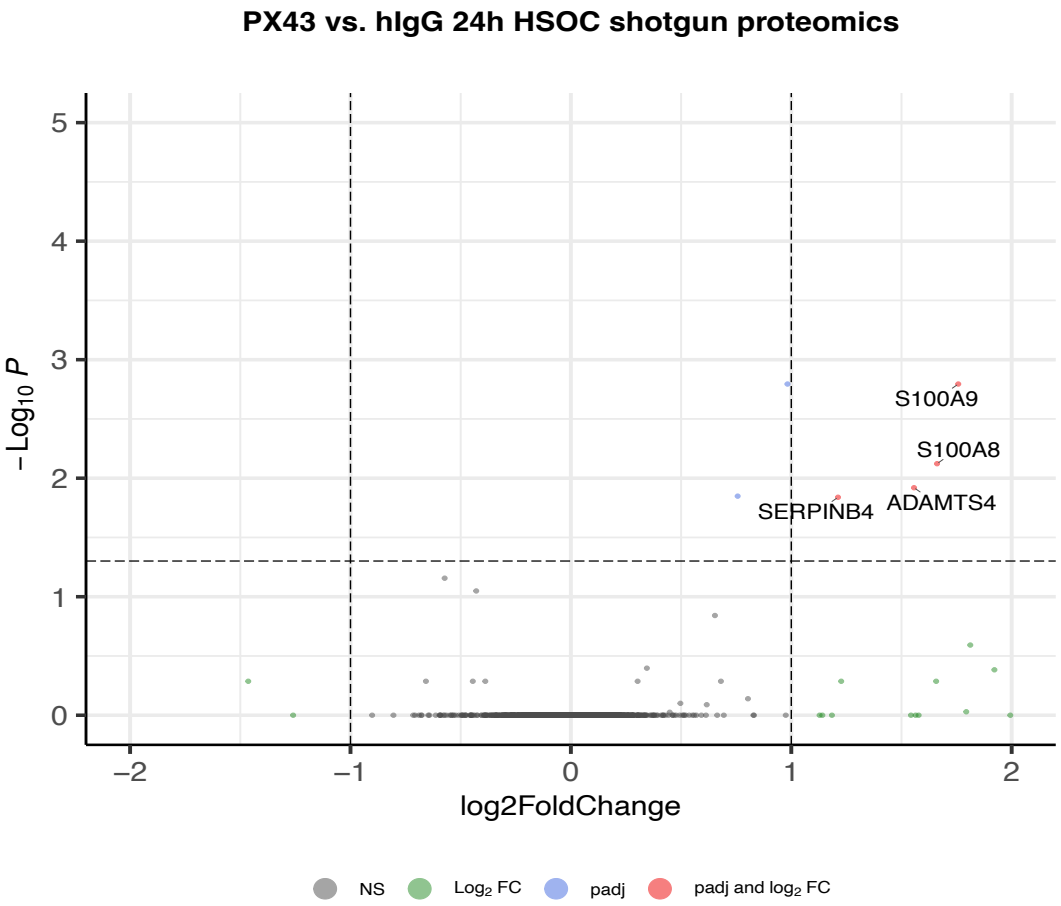


Figure 56: Volcano plot of shotgun proteomics in HSOC: PX43 vs. hIgG at 24h. Differentially abundant proteins (DAPs) identified between PX43 and hIgG in HSOC, using significance thresholds of adjusted p-value (padj) < 0.05 and |log2 fold-change| > 1, are visualized in the volcano plot. It is important to note that the number of detected proteins is currently fewer than 4,000, which is substantially lower than the ~20,000 protein-coding genes in the human genome [219].

Although a substantial number of differentially expressed genes (DEGs) are observed in the comparison of PX43 versus hIgG at 24 hours in the HSOC model, relatively few proteins exhibit significant differential abundance under the stringent threshold of adjusted $p < 0.05$ and $|\log_2 FC| > 1$. This discrepancy is not unexpected and can be attributed to various biological and technical factors. When the threshold is relaxed to adjusted $p < 0.05$ and $|\log_2 FC| > 0.58$, 15 proteins meet the criteria for significant differential abundance. It is also important to consider that the total number of proteins detected in the current dataset is fewer than 4,000, in contrast to the approximately 20,000 protein-coding genes in the human genome [219].

Not so many significantly differentially abundant proteins may be explained by the fact that gene expression at the RNA level does not always directly correspond to protein levels due to various regulatory mechanisms, such as mRNA stability, translational efficiency, and protein degradation. Proteins have varying half-lives, and their steady-state levels may not

immediately reflect recent transcriptional changes captured by RNA-seq. While PX43 and hIgG may induce rapid transcriptional responses, protein-level changes might take longer to manifest or stabilize, especially if the measurement time point does not capture the dynamic processes of protein turnover. Additionally, the timing of the measurements may have missed critical windows where protein-level changes are more pronounced.

To assess the relationship between shotgun proteomics and bulk RNA-seq data in HSOC, a correlation analysis was performed on the log2 fold-change (log2FC) values of differentially abundant proteins (DAPs) and their corresponding log2FC values in the bulk RNA-seq dataset. This approach specifically focused on proteins that were also detected in the bulk RNA-seq data. The correlation between the abundance of 3743 differentially abundant proteins (PX43 vs. hIgG 24h) and their corresponding transcription levels in HSOC revealed a positive correlation of 0.36 (Figure 57). This value aligns with previously reported correlations (~0.4) from human skin bulk RNA-seq and proteomic datasets [220]. The observed moderate correlation suggests a reasonable degree of consistency between protein and mRNA abundance in the HSOC model, supporting the validity of the proteomic data.

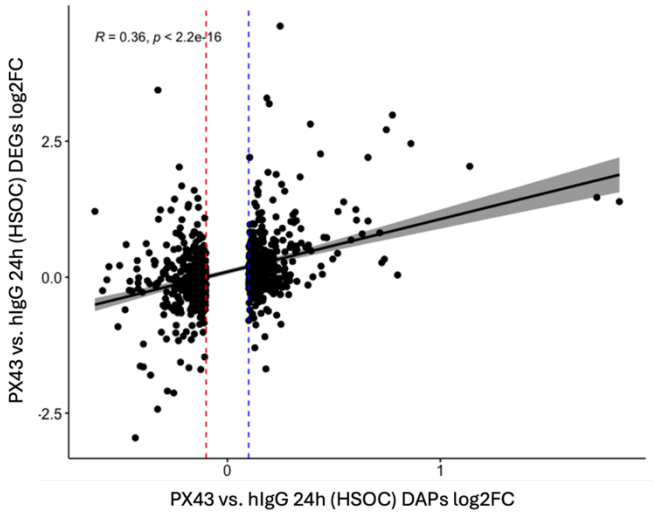


Figure 57: Significantly enriched pathways in HSOC shotgun proteomics data: PX43 vs. hIgG at 24h. Correlation analysis between the log2 fold-change (log2FC) values of 3,743 differentially abundant proteins (DAPs) and their corresponding log2FC values from bulk RNA-seq data in HSOC revealed a moderate positive correlation (r = 0.36). This level of correlation is consistent with previously reported values for human skin proteomics and transcriptomics, supporting the reliability and consistency of the proteomics data in relation to the RNA-seq results.

Enrichment analysis utilized the $\log_2$ fold change values of all 3,743 differentially abundant proteins identified in the comparison between PX43 and hIgG treated samples at the 24 hour time point revealed a significant enrichment of upregulated pathways, notably TNFα signaling, interferon gamma response, and oxidative phosphorylation, suggesting that PX43 treatment may activate key inflammatory and metabolic pathways as shown in Figure 58.

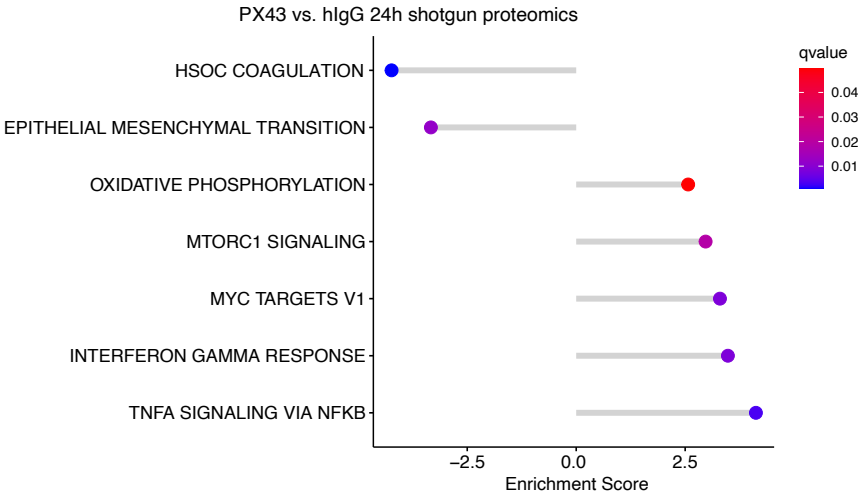


Figure 58: Significantly enriched pathways in HSOC shotgun proteomics data: PX43 vs. hIgG at 24h. Differentially expressed genes were analyzed using GAGE to identify enriched pathways, with a significance threshold of $q < 0.05$.

In summary, the proteomics data show strong concordance with the bulk RNA-seq results in HSOC, supporting the observation that changes at both the transcriptomic and proteomic levels are consistent with the phenotypic manifestation of split formation. Upon PX43 treatment of human skin organ culture, split formation is observed, which is reflected at both the transcriptome and proteome levels.

5. Assessment of the Impact of Endotoxin Contamination

5.1 Endotoxin: The Origin and Impact

Lipopolysaccharide (LPS), now commonly referred to as endotoxin is a collective term for structural components of the outer membrane of Gram-negative bacteria, such as Escherichia coli and Salmonella [221,222]. These molecules share a common architectural framework. Lipopolysaccharides (LPS) are large, complex molecules composed of three distinct regions: an outer core polysaccharide known as the O-antigen, an inner core oligosaccharide, and Lipid A; the primary component responsible for toxicity. In modern terminology, 'endotoxin' is often used interchangeably with LPS. However, some endotoxins—originally defined as toxins retained within bacterial cells and released upon cell lysis—are not related to LPS. An example is the delta endotoxin proteins produced by Bacillus thuringiensis [223].

Figure 59 provides a comprehensive depiction of endotoxin structures.

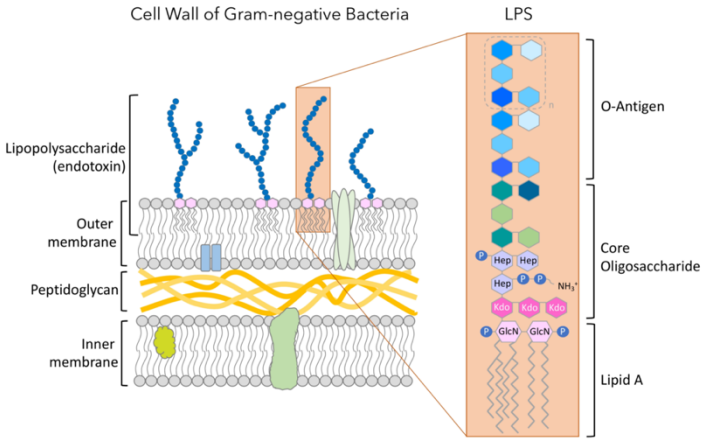


Figure 59: A schematic representation of the structure of endotoxins. The three structural components of lipopolysaccharide (LPS): the variable O-antigen, the core oligosaccharide, and the highly conserved Lipid A. Lipid A is embedded in the outer membrane of Gram-negative bacteria and is the primary mediator of endotoxic activity, while the O-antigen contributes to serotype diversity. The picture is extracted from online blogs by AB Biosciences (2020) [224].

LPS binds to lipopolysaccharide-binding protein (LBP) via its lipid A component and, in conjunction with other proteins, facilitates the transport of foreign substances to sites of metabolism and degradation, such as monocytes and macrophages. Lung macrophages and blood monocytes express the cell surface protein CD14, which serves as a receptor for the LBP-LPS complex. Additionally, with the involvement of CD14 and the accessory protein MD2, the LBP-LPS binding signal is transmitted to toll-like receptor 4 (TLR4) [225,226]. Following

endotoxin internalization, a complex signal transduction cascade is initiated. A schematic representation of the cellular endotoxin signaling pathway is provided in Figure 60.

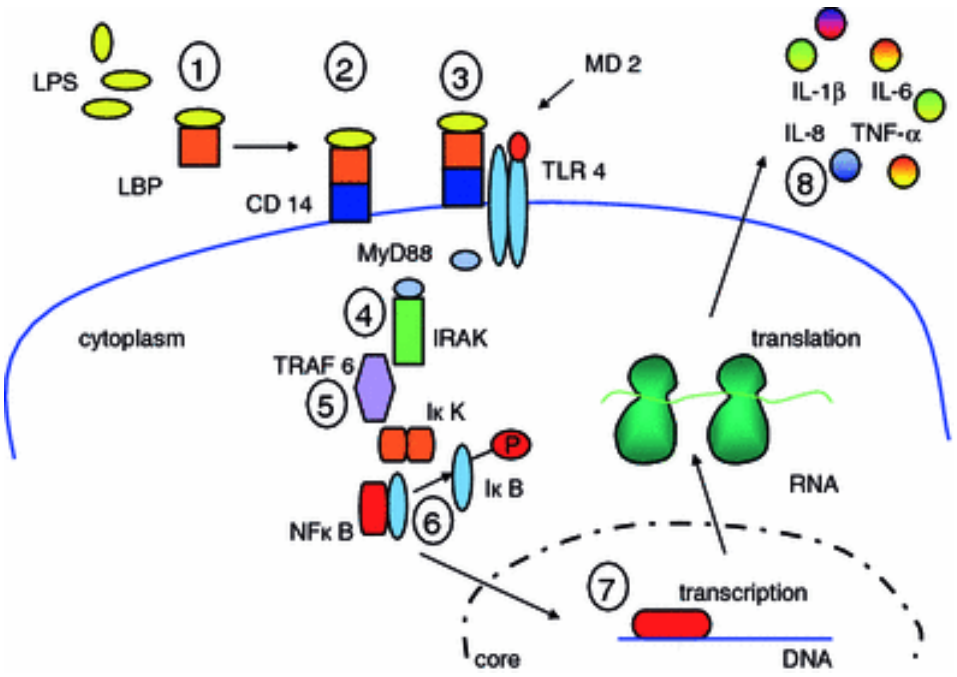


Figure 60: A schematic view on the cellular endotoxin signal transduction pathway. LPS-binding protein (LBP), present in plasma, binds to LPS (1). This LBP-LPS complex subsequently interacts with the cell surface receptor CD14 (2). The aggregation of the LBP-LPS-CD14 complex with MD2 and the transmembrane toll-like receptor 4 (TLR4) (3) initiates intracellular signal transduction (4, 5). This cascade ultimately leads to the activation of the transcription factor NF-κB (6), which drives the transcription and translation of various genes, including those encoding proinflammatory mediators such as IL-1β (7, 8). The figure is extracted from the paper Liebers et al. (2008) [227].

Endotoxin identification also plays a role in the pharmaceutical and medical industry for product quality and safety. Parenteral and injectable drugs, biologicals (e.g., insulin) and medical implants must be sterile. The sterilisation process can however release endotoxins in case Gram-negative bacteria are present and killed. Endotoxins are heat stable and persist even after bacterial death. Their inactivation is neither possible with boiling nor with autoclaving.

The Limulus Amebocyte Lysate (LAL) assay is the standard method for detecting endotoxins, relying on blood from the horseshoe crab (Limulus polyphemus) [228]. Even trace amounts of LPS can trigger coagulation of the Limulus lysate due to a highly sensitive enzymatic amplification cascade. However, concerns over the declining horseshoe crab population and potential assay interferences have driven efforts to develop alternative detection methods. Among the most promising alternatives are ELISA (Enzyme Linked Immunosorbent Assay)

based assays utilizing a recombinant form of Factor C (rFC), a key protein in the LAL assay [229].

The rFC is capable of detecting endotoxin in a solid phase [230]. ELISA-based assay (Figure 61), lipid A (the bioactive pathogenic moiety of endotoxin) immobilized on microtitre plate captures rFC, which binds anti-rFC antibody to interact with a secondary antibody conjugated with horseradish peroxidase. The captured secondary antibody then hydrolyses a chromogenic substrate [ABTS: 2,2'-azino-di-(3-ethylbenzthiazoline sulfonate)] yielding a coloured product, which corresponds to the presence of endotoxin.

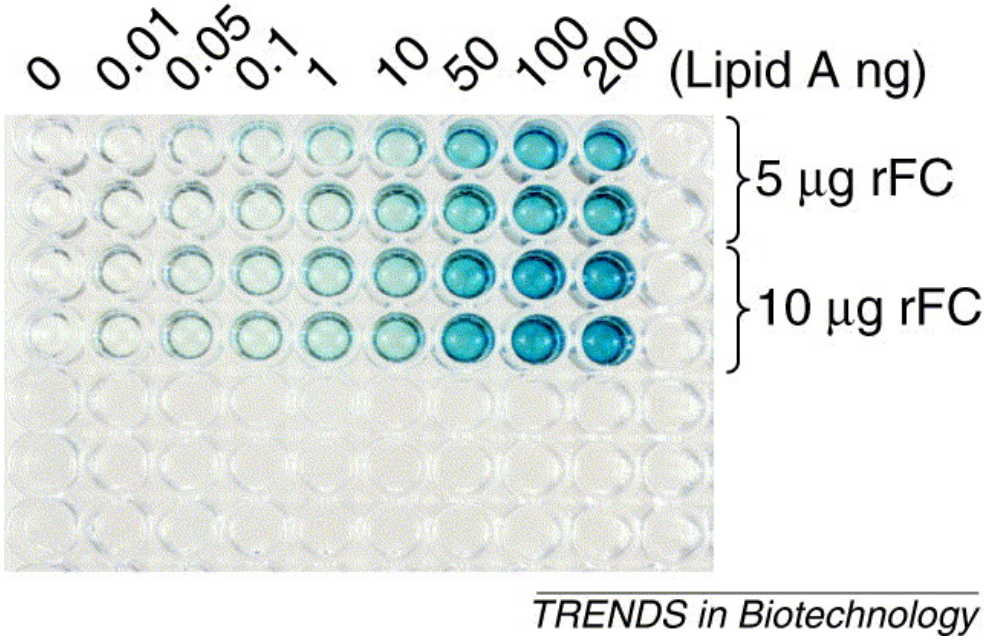


Figure 61: ELISA-based detection of lipid A immobilized on a microtitre plate. Increasing intensity of colour development (at OD405 nm) of the enzymatically hydrolysed product is seen from 0.01 to 200 ng lipid A. The figure is extracted from the online google patents US6645724B1 by Ding et al. (2003) [230].

Due to the potential adverse effects of endotoxins, the U.S. Food and Drug Administration (FDA) has established stringent limits on endotoxin levels in medical devices and therapeutic drug products. For medical devices, endotoxin limits are determined based on the device's contact with the body. Specifically, using the recommended extraction volumes, the allowable endotoxin level is 0.5 EU/mL or 20 EU/device for products that directly or indirectly contact the cardiovascular and lymphatic systems. For devices that come into contact with cerebrospinal fluid (CSF), the limit is 0.06 EU/mL or 2.15 EU/device.

For therapeutic drug products, the FDA recommends that manufacturers conduct endotoxin screening at a dilution level just above the threshold that neutralizes interference. For instance, if a product has a maximum valid dilution (MVD) of 1:100 and demonstrates inhibition at 1:10, but not at 1:20, it is advisable to begin screening at 1:30. If endotoxins are detected at this level, a full quantification should be performed to accurately determine the endotoxin concentration.

5.2 Endotoxin Detected in the Antibodies Except hIgG

At later time points, our collaborators from the University of Bern evaluated the antibodies using the standard Limulus Amebocyte Lysate (LAL) assay. The results indicated that only hIgG was free from endotoxin, while the other antibodies, including mIgG, AK23, and scFv PX43, exhibited varying levels of endotoxin contamination.

Table 9: Endotoxin levels in the antibodies tested by amebocyte lysate assay

Date (DD/MM/YYYY)	Product	Batch #	Limit (EU/mg)	Result (EU/mg)
27/07/2023	PX43 [8mg/mL]	7	0.16 (for 20g mouse)0.06 (for 1g of skin)	>0.625
27/09/2023	PX43 [8mg/mL]	7	0.16	>156
23/10/2023	PX43 [8mg/mL]	7	0.16	>1562.5 <15625
24/10/2023	PX43 [8mg/mL]	7	0.16	<3125
27/07/2023	hIgG, IVIg	B791838	0.16 (for 20g mouse)0.06 (for 1g of skin)	<0.02
27/07/2023	2G4[1.05mg/mL]	10/2022	0.4 (for 20g mouse)	>4.16 <41.6
27/09/2023	2G4[0.95mg/mL]	11/2022	0.4	<4
21/06/2023	mIgG [1mg/mL in PBS,NaCOOH 3mM]	220303-156	0.4 (for 20g mouse)	>31.25
23/10/2023	mIgG	220303-156	0.4	>395.56 <3955.6
24/10/2023	mIgG	220303-156	0.4	<1978
17/05/2023	AK23[3.4mg/mL]	XVI	0.4 (for 20g mouse)	<31.25
21/06/2023	AK23 0%FBS	I	0.4 (for 20g mouse)	> 3.125 <31.25
03/05/2023	ddH2O	3052023	0.125	<0.06EU/mL
27/09/2023	ddH2O	27092023	0.125	<0.06EU/mL
23/10/2023	ddH2O	23102023	0.125	<0.06EU/mL

5.3 Approaches for Identifying Endotoxin Related Genes

To identify endotoxin-related genes, employed an approximate approach by selecting differentially expressed genes (DEGs) with an p-value < 0.05 using DESeq2. This analysis compared PX43 to hIgG at 5 hours, 10 hours, and 24 hours using X, Y, Z samples, which each only including one pure donor to avoid confounding factors. A unified list of DEGs was generated, considering that PX43 exhibits high endotoxin levels, whereas hIgG is endotoxin-free.

Differential expression analysis using DESeq2 identified 716, 1,209, and 1,212 differentially expressed genes (DEGs) for PX43 compared to hIgG at 5 hours, 10 hours, and 24 hours, respectively, under a significance threshold of p-value < 0.05. In total, 2726 unique DEGs were detected across all time points, as illustrated in the Venn diagram (Figure 62). Notably, 30 genes were consistently differentially expressed at all three time points (5h, 10h, and 24h), while pathway enrichment analysis of these 30 genes using the STRING database did not reveal any significantly enriched pathways. This may suggest the absence of a consistent perturbation induced by the endotoxin.

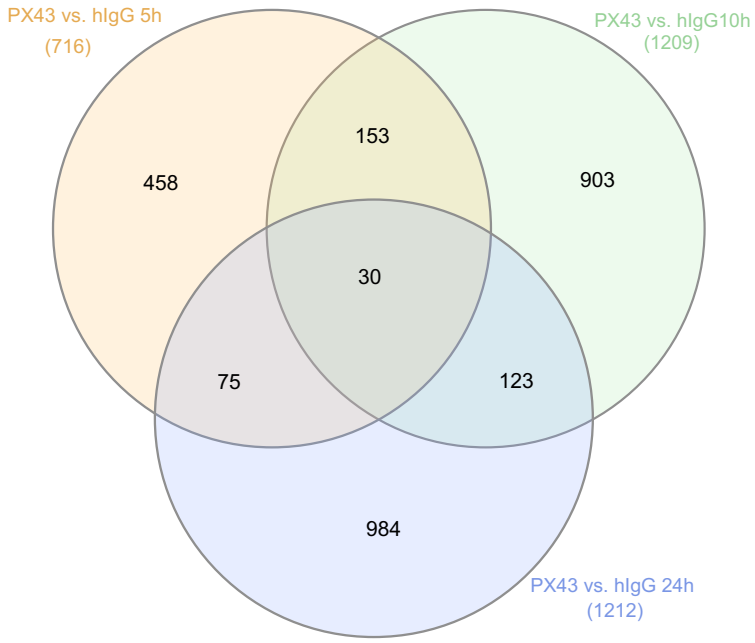


Figure 62: Identification of endotoxin-related genes using an approximate method. Differentially expressed genes (DEGs) were identified by comparing the high-endotoxin PX43 condition to the endotoxin-free hIgG condition at 5h, 10h, and 24h (p-value < 0.05). Endotoxin-related genes were then determined based on this differential expression analysis.

5.4 The Omission of Endotoxin-Related Genes Does Not Qualitatively Alter the Findings

To evaluate the impact of endotoxins, the 2726 endotoxin-related genes identified using the approximate method were excluded from the differential expression analysis of PX43 compared to hIgG at 5h, 10h, and 24h in samples X, Y, and Z. Following the removal of these genes, differentially expressed genes (DEGs) were reassessed, and the top enriched pathways were analyzed.

After removing the 2,726 endotoxin-related genes, no significantly differentially expressed genes (DEGs) were identified for PX43 compared to hIgG in samples X, Y, and Z at 5h, 10h, and 24h under the criteria of adjusted p-value < 0.05 and $|\log_2 FC| > 1$. This result is consistent with the findings prior to endotoxin gene removal, where fewer than 10 significant DEGs were observed across the same conditions.

Spearman correlation analysis of log2-transformed TPM values for PX43 and hIgG in samples X, Y, and Z at 5h, 10h, and 24h—both before and after removal of 2,726 endotoxin-related genes—revealed consistently high positive correlations ($r = 0.95–1.00$, $p < 0.001$). These results suggest that endotoxin-related genes have minimal impact on the overall transcriptomic profiles in these datasets.

Additionally, pathway enrichment analysis identified significantly enriched pathways for PX43 compared to hIgG at 5h, whereas no significant pathways were detected at 10h and 24h. Moreover, the pathway enrichment results at 5h remained consistent after the removal of endotoxin-related genes, with both analyses indicating a reduction in inflammation-related pathways, as shown in Figure 63.

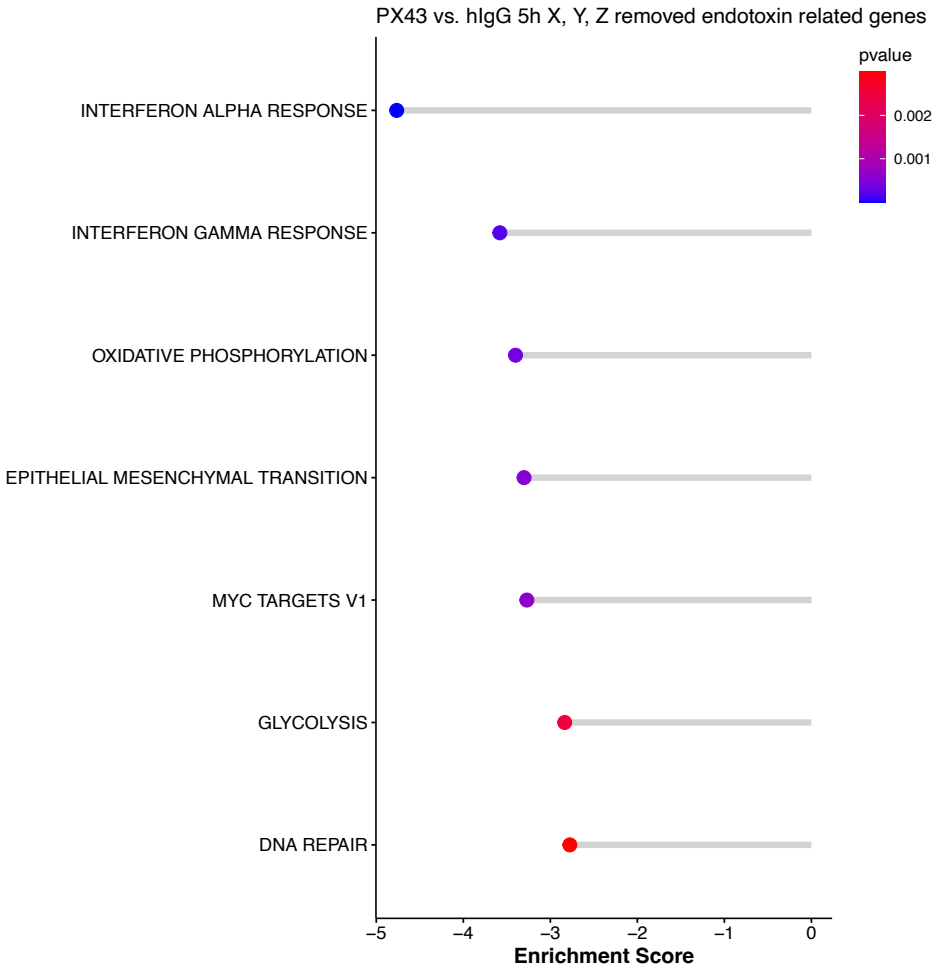


Figure 63: Enriched pathways of PX43 vs. hIgG at 5h after removing endotoxin related genes. Differentially expressed genes were analyzed using GAGE to identify enriched pathways, with a significance threshold of $q < 0.05$.

The strong positive correlation observed between the log2-transformed TPM values of PX43 and hIgG at 5h, 10h, and 24h across the X, Y, and Z samples, along with the consistent results for the enriched pathways before and after the exclusion of endotoxin-related genes, suggests that endotoxin does not qualitatively impact the findings.

6. Discussion

6.1 Summary

The present study advances our understanding of pemphigus vulgaris (PV) by dissecting the localized molecular and structural consequences of pathogenic antibodies using physiologically relevant models. As introduced in Section 1.2, two prevailing hypotheses dominate the field: one posits that PV is primarily driven by desmosome remodeling [34], while the other suggests a broader immune dysregulation mediated by autoantibody signaling [35]. Our findings provide support for the former, highlighting the primacy of adhesion disruption over widespread transcriptional or inflammatory shifts.

In the human primary epidermal keratinocyte (HPEK) model, we observed only minimal differences between the disease-related antibodies (scFv PX43 and AK23) and the control antibodies (hIgG and mIgG) at the RNA and protein levels across the tested time points (5h, 10h and 24h). These findings suggest that, at the global transcriptomic and proteomic levels, the response to the disease antibodies is subtle. Further analysis through single-cell RNA sequencing (scRNA-Seq) revealed that only a small subset of keratinocyte subpopulations exhibited detectable changes, further validated the findings in the bulk RNA-seq data.

In the HSOC model, treatment with PX43 induced a progressive and significant increase in epidermal splitting over time (5 h to 24 h), which was accompanied by a parallel rise in the number of differentially expressed genes (DEGs), consistently exceeding 150 at each time point. In contrast, AK23 treatment did not result in skin split formation or significant transcriptional changes, with fewer than 10 DEGs detected. This clear divergence in molecular and phenotypic outcomes underscores fundamental differences in the pathogenic potential of these antibodies. These findings suggest that PV pathogenesis is influenced not merely by the presence of autoantibodies, but also by factors such as epitope specificity, binding affinity, and the resulting conformational effects on desmosomal protein complexes.

HSOC proteomic analysis at the 24-hour time point revealed only four proteins with significantly altered abundance between PX43 and control (hIgG), under a stringent significance threshold (adjusted $p < 0.05$ and $|\log_2 FC| > 1$). This relatively low number of

differentially abundant proteins is not unexpected and may reflect both biological complexity and technical limitations. Upon applying a more relaxed threshold (adjusted $p < 0.05$ and $|\log_2 FC| > 0.58$), the number of significantly altered proteins increased to 15, highlighting the sensitivity of statistical outcomes to threshold selection.

It is also important to note that HSOC proteomic dataset included fewer than 4,000 reliably quantified proteins, compared to the approximately 20,000 protein-coding genes in the human genome, which may limit comprehensive proteome coverage. Nonetheless, a moderate positive correlation ($r = 0.36$) was observed between the abundance of the 3,743 quantified proteins and their corresponding transcript levels in the HSOC model following PX43 treatment at 24 h. This correlation is consistent with previously reported transcriptome-proteome concordance in human skin ($r \sim 0.4$) [220] , supporting the robustness and biological relevance of the HSOC results.

Collectively, the results suggest that the disease antibodies associated with Pemphigus Vulgaris (PV), specifically PX43 and AK23, do not induce a widespread, long-lasting transcriptomic or proteomic response at the cellular level in either the human epidermal keratinocyte (HEPK) or human skin organ culture (HSOC) models. These findings provide strong evidence supporting the hypothesis that PV primarily functions as a disease of desmosome remodeling [34] rather than a broad systemic alteration of keratinocyte gene expression or protein synthesis. The data indicate that the PV antibodies bind specifically to desmosomal components, such as DSG3 and DSG1, thereby disrupting desmosomal integrity and disturbing cellular homeostasis. This disruption leads to the separation of keratinocytes, the formation of skin splits, and the subsequent initiation of inflammatory processes. These findings highlight the pivotal role of desmosome dysfunction in the pathogenesis of PV and underscore the localized nature of the disease's impact on epidermal cell adhesion and skin integrity.

Future studies should aim to capture early and spatially restricted signaling events, such as calcium flux, kinase activation, or desmosome internalization, which may precede transcriptional or proteomic responses. Understanding these early events may offer novel therapeutic windows focused on preserving desmosomal integrity or interrupting the conformational cascade initiated by specific pathogenic antibodies.

6.2 Open Questions

6.2.1 What Is the Optimal Number of Biological Replicates?

Biological replicates use different biological samples of the condition to measure the biological variation between samples. For differential expression analysis, the more biological replicates, the better the estimates of biological variation and the more precise our estimates of the mean expression levels. This leads to more accurate modeling of our data and identification of more differentially expressed genes. Liu et al. (2014)[231] reported that biological replicates are of greater importance than sequencing depth, which is the total number of reads sequenced per sample. The figure 64 shows the relationship between sequencing depth and number of replicates on the number of differentially expressed genes identified.

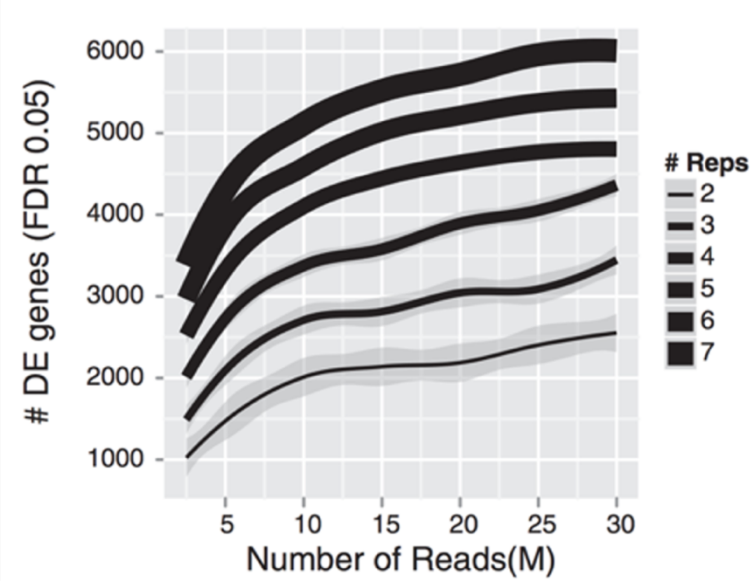


Figure 64: Effect of read depth and replicates on differential gene expression detection. Increase in biological replication significantly increases the number of DE genes identified. Numbers of sequencing reads have a diminishing return after 10 M reads. Line thickness indicates depth of replication, with 2 replicates the darkest and 7 replicates the lightest. The lines are smoothed averages for each replication level, with the shaded regions corresponding to the 95% confidence intervals. The figure is extracted from the paper by Liu et al. (2014) [231].

Other researchers also recommend using at least six biological replicates in the RNA-seq experiments and rising to at least 12 when it is important to identify the significantly differentially expressed genes for all fold changes [232]. Their results show the fraction of all genes edgeR calls as significantly differentially expressed gene increases as a function of n_r (number of replicates) and impact of sampling effects on this fraction shrinks as nr increase (Figure 65 A). The true positive rate (TPR) performance changes as a function of both replicate

number and fold-change threshold (Figure 65 B,C). However, edgeR (exact) successfully controls its false discovery rate (FDR) for all combinations of both nr and T and the primary effect of increasing the number of replicates or imposing a fold-change threshold is to increase the sensitivity of the tool, converting false negatives to true positives (Figure 65 D).

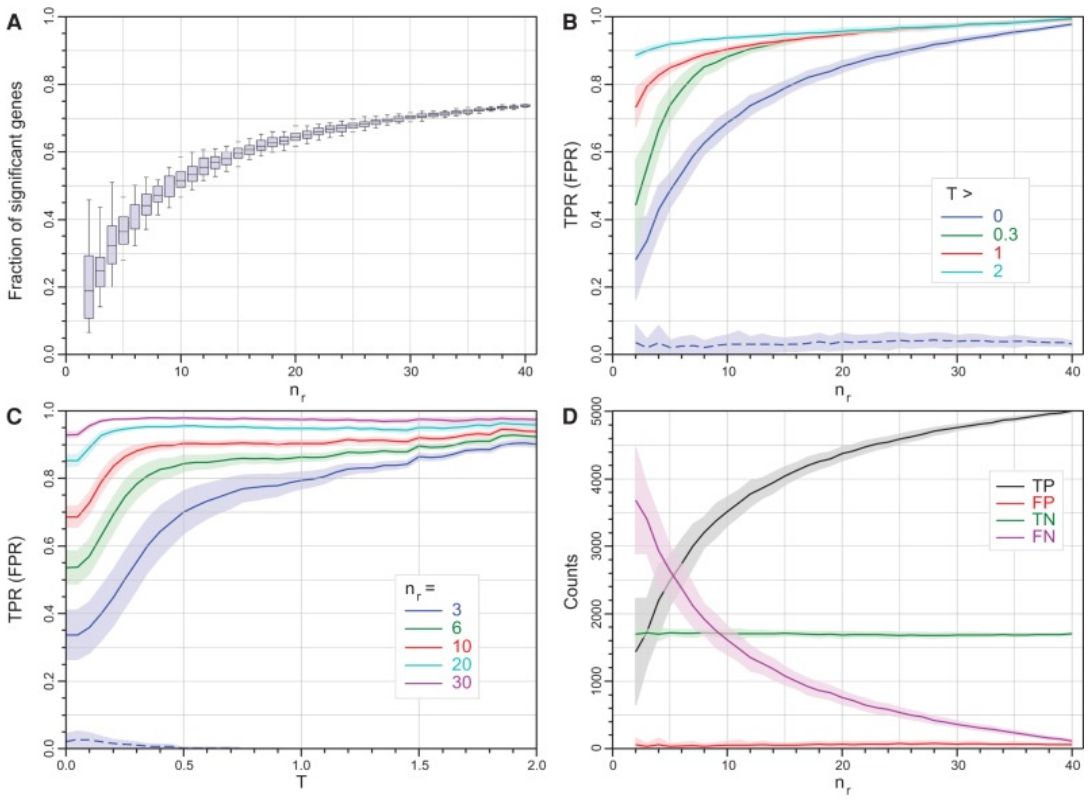


Figure 65: Enhanced performance in differential expression analysis with increased biological replicates. Statistical properties of edgeR (exact) as a function of $|\log_2(FC)|$ threshold, T, and the number of replicates, nr. Individual data points are not shown for clarity; however, the points comprising the lines are each an average over 100 bootstrap iterations, with the shaded regions showing the 1 SD limits. (A) The fraction of all (7126) genes called as SDE as a function of the number of replicates (boxplots show the median, quartiles and 95% limits across replicate selections within a bootstrap run). (B) Mean true positive rate (TPR) as a function of nr for four thresholds T∈{0,0.3,1,2} (solid curves, the mean false positive rate [FPR] for T = 0 is shown as the dashed blue curve, for comparison). Data calculated for every $\Delta n_r = 1$. (C) Mean TPR as a function of T for nr∈{3,6,10,20,30} (solid curves, again the mean FPR for nr = 3 is shown as the dashed blue curve, for comparison). Data calculated every ΔT = 0.1. (D) The number of genes called as true/false positive/negative (TP, FP, TN, and FN) as a function of nr. The FPR remains extremely low with increasing nr demonstrating that edgeR is excellent at controlling its false discovery rate. Data calculated for every $\Delta n_r = 1$. The figure is extracted from the paper Schurch et al. (2016) [232].

In the HSOC model, bulk RNA-seq was performed with six biological replicates, providing sufficient robustness for the identification of significantly differentially expressed genes. Nevertheless, increasing the number of biological replicates is generally recommended when feasible, as it enhances statistical power and confidence in the results.

6.2.2 What are the major changes of Keratinocyte cells on Pemphigus Vulgaris Susceptibility?

Keratinocytes, as the terminally differentiated cells of the epidermis, maintain a natural balance between proliferation and programmed cell death. Over the course of approximately two weeks, they undergo a well-orchestrated differentiation process known as cornification [233]. During this time, keratinocytes in the suprabasal layers sequentially express characteristic differentiation markers as they migrate toward the stratum corneum, ultimately contributing to the formation and maintenance of the skin barrier.

Exposure to ultraviolet (UV) radiation [234] or cytotoxic agents [235]can induce apoptosis in keratinocytes. Moreover, apoptosis has been implicated in the pathogenesis of pemphigus, suggesting its involvement in disease-related epidermal damage [236].

Figure 66 illustrates the key features of keratinocyte cornification and apoptosis.

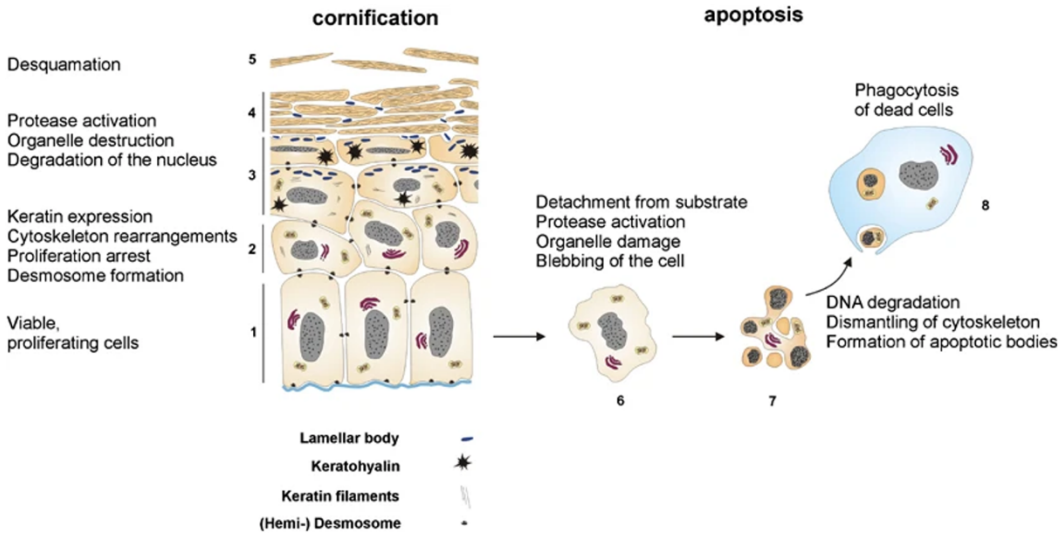


Figure 66: Morphological and biochemical changes during keratinocyte cornification and apoptosis. (1) The basal layer of the epidermis consists of undifferentiated, mitotically active keratinocytes attached to the basement membrane. (2) As keratinocytes migrate into the spinous layer, they detach from the basal lamina and begin terminal differentiation, marked by cell flattening, expression of keratins 1 and 10, and coordinated cytoskeletal reorganization across the epithelial sheet. Cell surface projections terminate in desmosomes, promoting intercellular adhesion. (3) In the granular layer, keratohyalin granules become evident; keratin filaments are crosslinked, DNA is degraded, organelles are dismantled, and the plasma membrane is replaced by a cornified envelope with ceramide deposition from lamellar bodies. (4) The outermost cornified layer comprises enucleated, dead keratinocytes (corneocytes), forming the skin's protective barrier. (5) Unlike apoptotic cells, corneocytes are not removed by phagocytic Langerhans cells but are instead shed into the environment through desquamation. (6) In apoptosis, activation of the caspase cascade triggers membrane blebbing and initiates controlled cellular dismantling. (7) Chromatin and cytoplasm condense, internucleosomal DNA fragmentation occurs, the cytoskeleton is disassembled, and the plasma membrane invaginates to form apoptotic bodies while maintaining membrane integrity to prevent cytoplasmic leakage. Mitochondrial damage also leads to the release of pro-apoptotic factors such as cytochrome c into the cytosol. (8) Finally, apoptotic bodies are recognized, engulfed, and degraded by phagocytes via lysosomal digestion. Figure is extract from paper by Lippens et al. (2005) [237].

Understanding the fate of keratinocytes in the context of Pemphigus Vulgaris (PV) is a key question, particularly within the 3D human skin organ culture (HSOC) model, which closely mimics the in vivo environment. Determining whether keratinocytes primarily undergo detachment, apoptosis, or differentiation under pathogenic conditions can provide critical insights into disease progression and therapeutic targeting.

Transcriptomic response to PX43 associate with split formation has a positive correlation with keratinocyte cell detachment and acute skin wounding, as shown in the above section 4.6.2. Additionally, the transcriptomic response to PX43 has been shown to negatively correlate with keratinocyte differentiation, as reported by Hartmann et al. (2025) [238], this recently published online biorxiv is our collaborate research summary for this PV project, myself also as a co-first author, led by Professors Hauke Busch, Jennifer Hundt, Eliane Müller and Christoph Hammers. These findings further support the notion that, in Pemphigus Vulgaris, keratinocytes predominantly undergo apoptosis and cellular detachment rather than progressing through normal differentiation pathways.

It would also be valuable to use large-scale single-cell RNA sequencing (scRNA-seq) to further investigate keratinocyte cell trajectories under PV conditions. While bulk transcriptomic analyses provide insight into overall pathway activation, they cannot resolve the heterogeneity of individual cell states within the epidermis. By applying scRNA-seq to HSOC models treated with PV IgG or PX43, researchers could map the dynamic transitions of keratinocytes—from healthy states toward detachment, apoptosis, or alternative stress responses. Trajectory inference and pseudotime analysis could help clarify whether certain keratinocyte populations are more resistant to split formation or capable of initiating repair programs, offering new angles for therapeutic targeting and precision intervention.

6.2.3 Can Inhibitors That Reduce Split Formation Alter Transcriptome Perturbations Qualitatively?

Recent PV research increasingly focuses on inhibitors that stabilize desmosomes by targeting signaling pathways involved in acantholysis, as these mechanisms go beyond direct desmoglein inhibition and have shown relevance in both in vitro and in vivo models as summarized by Schmitt et al. (2021) [39].

In a 3D human skin organ culture model, Burmester et al. screened a panel of kinase inhibitors and identified five candidates targeting MEK1, TrkA, PI3Kα, p38 MAPK, and VEGFR2, that were able to attenuate, though not completely prevent, PV-induced split formation as shown in Figure 67 [239]. Notably, inhibitors were pre-incubated for 2 hours prior to the addition of PV IgG. These findings support the concept that upstream signaling cascades can modulate desmosomal stability.

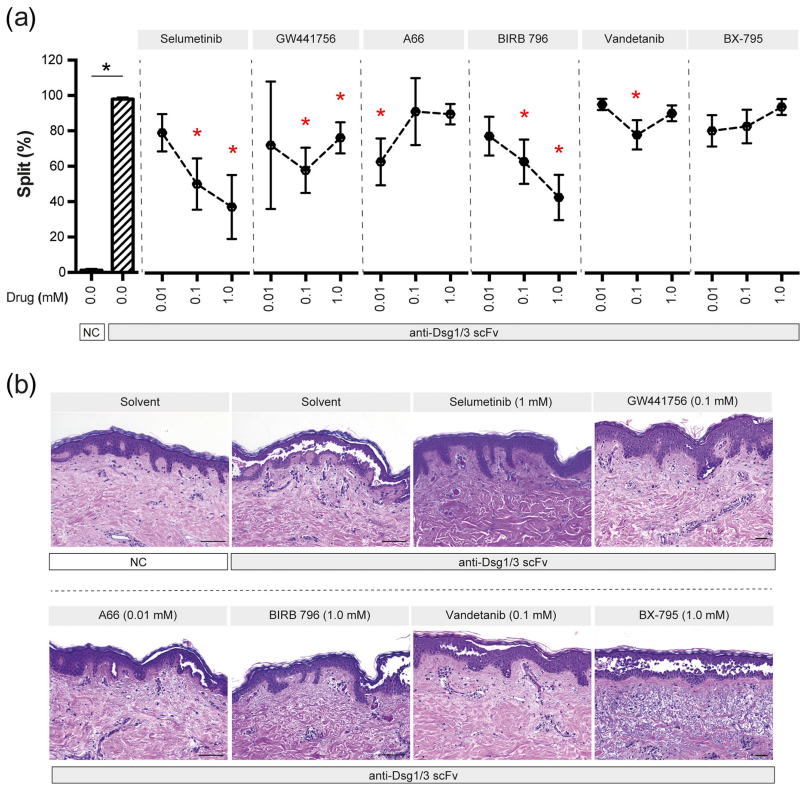


Figure 67: MEK1, TrkA, PI3Kα, p38 MAPK, and VEGFR2 inhibitors reduce PV IgG-induced blistering in the HSOC model. (a) The anti-Dsg1/3 scFv was injected into human skin samples, in the absence or presence of the indicated inhibitors. Data are shown as individual values, with medians ± SD from 8–19 sections per group from three organ cultures. *P<0.05, significantly different from anti-Dsg1/3 scFv alone; Kruskal–Wallis H test followed by Dunn's post hoc test. (b) Representative H&E-stained images from the experiments are shown in panel (a). For better contrast, white correction was applied using GIMP (www.gimp.org). Scale bars correspond to 100 µm. The figure is extracted from the paper by Burmester et al. (2020) [239].

Comparison of the 24-hour transcriptomic response to PX43 with RNA-seq data from inhibitor-pretreated HSOC samples subsequently exposed to PV IgG revealed a shared activation of key upstream pathways—including TGFα/NF-κB, TGF-β, and JAK-STAT—across all conditions, regardless of inhibition. This suggests these pathways may be critical bottlenecks. Targeting these with specific inhibitors such as NF-κB inhibitors (e.g., BAY 11-7082), TGF-β receptor blockers (e.g., SB431542), or JAK inhibitors (e.g., Tofacitinib) could be

prioritized in follow-up experiments to test whether these late-stage signaling nodes play a decisive role in split formation or epithelial repair.

In contrast, pathways such as EGFR, MAPK, and PI3K, which were upregulated following PX43 stimulation, were selectively suppressed by Vandetanib (a VEGFR2, EGFR, and RET inhibitor) and Selumetinib (a MEK1 inhibitor), as illustrated in Figure 68.

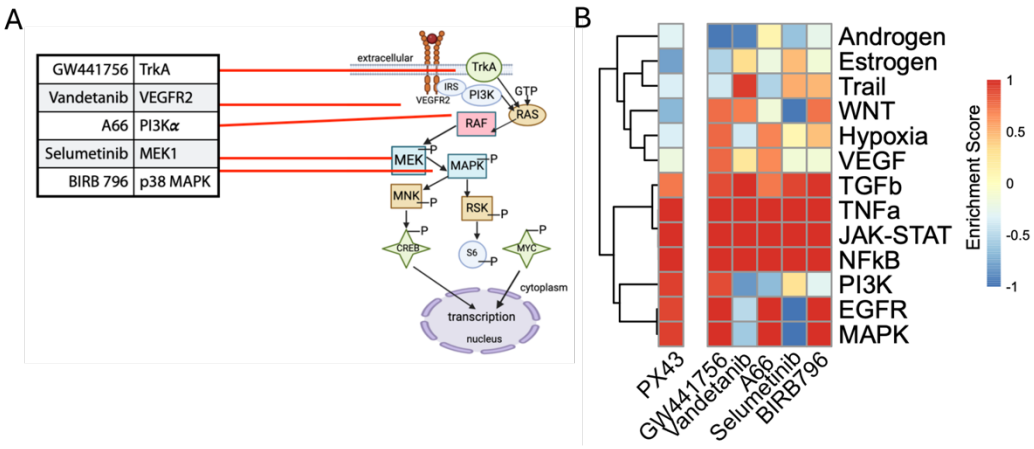


Figure 68: Inhibitors of PV-Related kinases modulate PX43-Induced pathways. Inhibitors targeting MEK1 (Selumetinib), TrkA (GW441756), PI3Kα (A66), p38 MAPK (BIRB 796), and VEGFR2 (Vandetanib) modulate key signaling pathways also activated by PX43 at 24 hours.(A) Schematic illustration of the MAPK signaling cascade and the specific upstream targets of the tested inhibitors.(B) Heatmap showing pathway enrichment scores from transcriptomic analysis comparing PX43 stimulation with inhibitor-pretreated HSOC samples following PV IgG exposure. Common activation of TGFa/NF-κB, TGFβ, and JAK-STAT signaling was observed across all conditions, while pathways such as EGFR, MAPK, and PI3K were upregulated by PX43 but suppressed by Vandetanib and Selumetinib treatment. The Figure is extracted from Hartmann et al. (2025) [238] online biorxiv as previously introduced.

Thus, while the inhibitors modulated upstream signaling pathways and reduced split formation, pathways related to cell detachment and wound response remained predominant. The inhibitors did not substantially alter the overall cellular response, suggesting that the transcriptomic profile is primarily dictated by the presence or absence of intraepidermal splits.

Future studies could also pursue other directions, such as 1) Does-Response and Time studies: Since the inhibitors were only pre-incubated for 2 hours before PV IgG addition, it remains unclear whether longer pre-treatment or post-treatment regimens might enhance efficacy. Controlled time-course studies with varied pre- and post-incubation intervals could help determine optimal therapeutic windows. 2) Combination Therapy Studies: Given that no single inhibitor fully prevented blistering, combining agents that target distinct but convergent pathways—such as pairing a MEK1 inhibitor (Selumetinib) with a VEGFR2/EGFR inhibitor (Vandetanib)—could provide synergistic effects. These combinations should be

tested in HSOC models with quantitative scoring of split severity and transcriptomic profiling to assess additive or compensatory pathway activation. 3) Delayed Intervention Studies: Most current protocols apply inhibitors prior to PV IgG exposure, which models a prophylactic setting. However, patients typically receive treatment after lesion formation begins. To better reflect this clinical reality, future experiments should investigate whether the same inhibitors can attenuate or reverse acantholysis when administered after PV IgG exposure—e.g., 2, 4, or even 6 hours post-treatment. This would help determine the potential for these agents as therapeutic, rather than purely preventive, interventions.

6.2.4 What Therapeutic Inspirations Can Be Drawn from Our Models for Pemphigus Vulgaris?

In general, our model of HPEK and HSOC mostly relied on keratinocyte cells as the sole material for sequencing, which may not fully capture the complexity of the PV disease process in vivo. Interactions with other skin cell types, such as immune cells, fibroblasts, and endothelial cells, may contribute to the overall disease pathology and were not considered in this model.

Fang et al. (2020) [240] provided a comprehensive review of the role of T cells in the immunopathogenesis of Pemphigus Vulgaris (PV). As illustrated in Figure 69, distinct T cell subsets, which are key regulators of autoimmunity, exhibit various dysregulations that contribute to disease progression and skin inflammation. T helper (Th)1 cells mediate pro-inflammatory immune responses primarily through interferon-gamma (IFN-γ), while Th2-derived cytokines, such as interleukin-4 (IL-4), promote B cell proliferation, antibody production, and immunoglobulin class switching. Th17 cells drive inflammatory responses and exacerbate tissue damage in the skin, whereas regulatory T cells (Tregs) play a critical role in suppressing autoreactive $CD4^+$ T cell activation and maintaining immune homeostasis. Additionally, T follicular helper (Tfh) cells engage in crosstalk with B cells, thereby facilitating autoantibody production, a hallmark of PV pathogenesis.

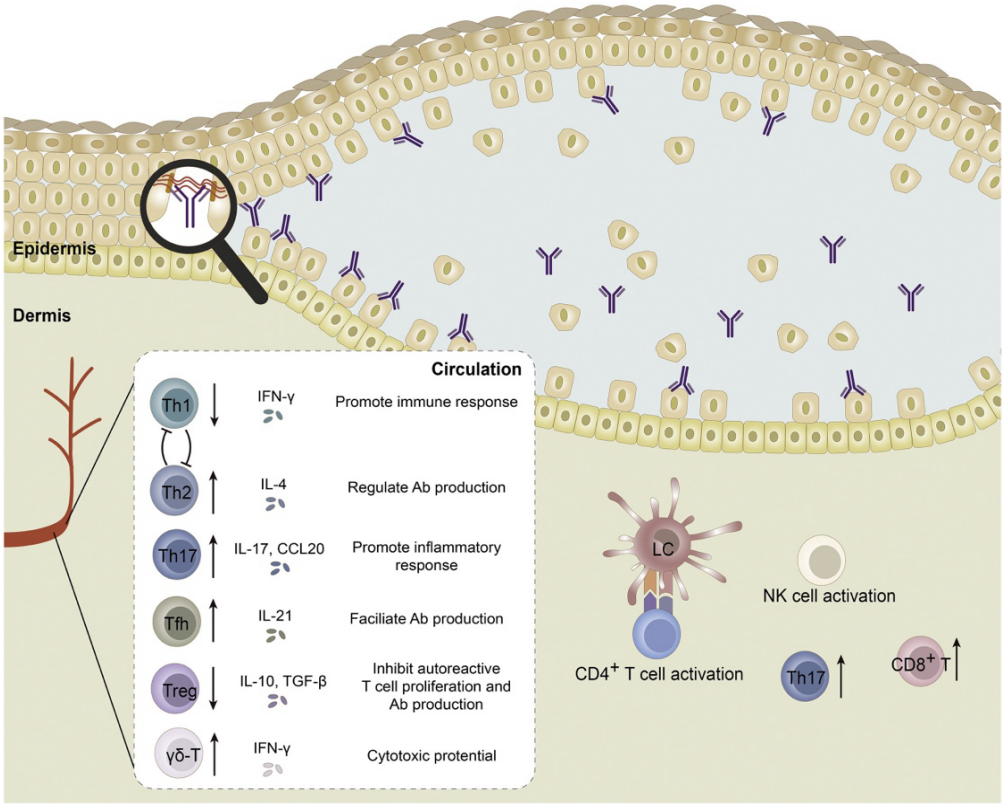


Figure 69: T cells in the pathogenesis of pemphigus vulgaris. PV is characterized by immune complex deposits that usually consist of IgG bound to desmosomal transmembrane proteins of keratinocytes. Autoantibodies binding to autoantigens can directly compromise desmosomal function and cause acantholysis. In the skin, Langerhans cells act as antigen presenting cells, take up and present Dsg peptides to CD4+ T cells and induce their activation. In the circulation, the ratios and functions of T cell subsets are both abnormal. Th1 cells promote immune response and Th2 cells regulate pathogenic Dsg-specific IgG antibody production. Th17 cells secrete IL-17, which promote an inflammatory response, while Treg cells inhibit Dsg3-autoreactive T cell proliferation and antibody production. Tfh cells, which crosstalk with B cells, facilitate autoantibody production. The Figure is extracted from the paper by Fang et al. (2020) [240].

Egami et al. (2023) [241] reported that autoreactive B cells play a pivotal role in pemphigus, with the number of DSG-specific B cells correlating with serum autoantibody titers. Their findings indicate that patients with anti-DSG3 IgG exhibit a higher frequency of DSG-specific B cells compared to healthy controls, as illustrated in Figure 70.

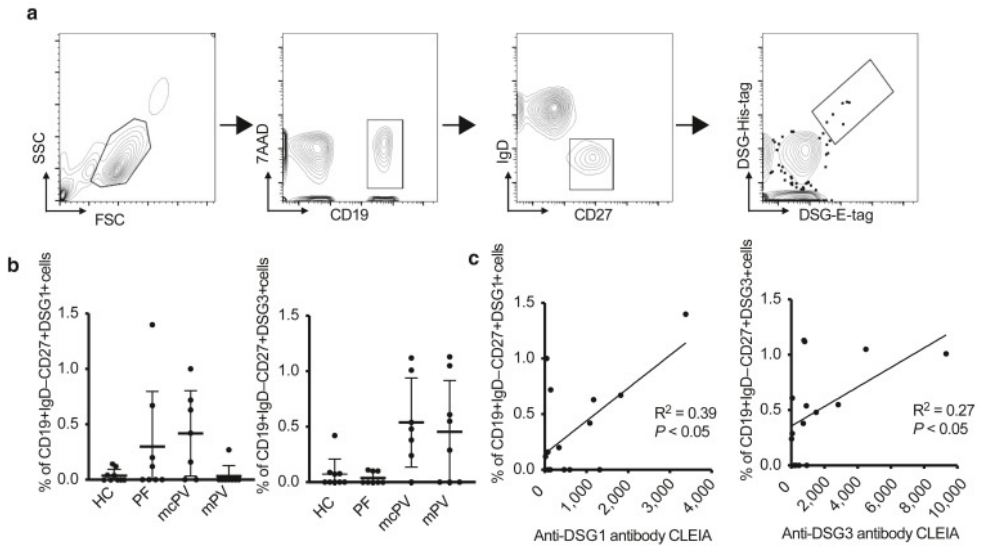


Figure 70: Detection of DSG-specific B cells in patients with pemphigus. (a) Scheme of detection. Memory B cells were categorized into CD19+IgD−CD27+ cells. After incubating with recombinant DSG proteins, DSG-specific B cells (E-tag+His-tag+) were detected in each patient population (7AAD−CD19+IgD−CD27+E-tag+His-tag+). (b) Proportion of DSG-specific memory B cells (CD19+IgD−CD27+DSG+ cells) in total memory B cells (CD19+IgD−CD27+ cells) in each patient population. plot: mean ± SD. (c) Scatter diagram of the proportion of DSG-specific memory B cells and the patient's serum DSG-specific autoantibody titer. 7AAD, 7-aminoactinomycin D; CLEIA, chemiluminescence enzyme immunoassay; DSG, desmoglein; E-tag, epitope tag; FSC, forward scatter; HC, healthy control; His-tag, histidine tag; mcPV, mucocutaneous pemphigus vulgaris; mPV, mucous pemphigus vulgaris; PF, pemphigus foliaceus; SSC, side scatter. The Figure is extracted from the paper by Egami et al. (2023) [241].

Fetter et al. (2020) [242] summarized that skin has the potential to act as a niche for localized antibody production, and tertiary lymphoid structure (TLS) may act as a site of restimulation of memory lymphocytes or priming of precursors that boost differentiation and expansion of effector cells and maintain self-reactive inflammatory responses, and may arise in the skin in response to chronic inflammation in cutaneous autoimmune disease (CAD) as shown in the Figure 71.

Drugs that inhibit the initiation of tertiary lymphoid structures (TLS) include compounds targeting the lymphotoxin-β pathway, as well as pro-inflammatory cytokines such as IL-17 and IL-22, which have been identified as critical mediators of TLS development in animal models [243,244].

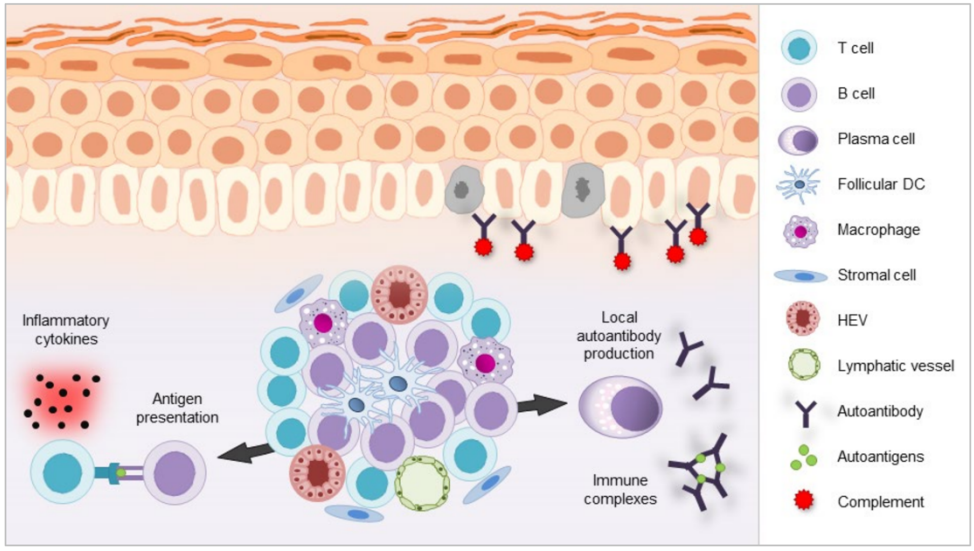


Figure 71: Schematic illustration of a potential tertiary lymphoid structure (TLS) in skin autoimmunity. TLS represent accumulations of lymphoid and stromal cells arising at ectopic sites as a reflection of chronic inflammation. As depicted, they may consist of B- and T-cell clusters and a network of stromal cells, follicular dendritic cells (DC) and macrophages as well as vascular structures such as high endothelial venules (HEV) and lymphatic vessels, possibly forming GC-like structures. TLS may create a microenvironment that enables localized autoantigen-directed immune responses such as T cell activation, which might also be driven by B cells with antigen-presenting features. Proinflammatory cytokines are subsequently released from activated T cells and autoantibody production by PCs as well as immune complex formation are initiated. TLS formation has been described in pemphigus and lupus erythematosus panniculitis, yet data concerning the appearance and specific role of TLS in autoimmune skin diseases are limited. The Figure is extracted from the paper by Fetter et al. (2020) [242].

Pemphigus vulgaris pathogenesis involves multiple interconnected processes, including desmosome binding by autoantibodies, autoantibody-mediated endocytosis, impaired desmosome regeneration, disruption of cellular homeostasis, immune cell activation, weakened keratinocyte adhesion, and inflammation. Given the complexity of these mechanisms, a therapeutic approach targeting multiple pathways may be more effective. For instance, Strandmoe et al. (2024) [245] reported a case in which plasma cell-targeted therapy, consisting of four cycles of bortezomib (1.3 mg/m²), dexamethasone (40 mg), and cyclophosphamide (500 mg/m²) administered every three weeks, led to a rapid clinical response and near-complete remission, as evidenced by a reduction in the Pemphigus Disease Area Index (PDAI) score from 73 to 3. Therefore, further investigation into additional therapeutic targets and combination strategies is warranted to optimize treatment outcomes.

6.3 Outlook

Building on the current findings, future studies should aim to address the limitations identified in this study to further elucidate the molecular mechanisms underlying Pemphigus Vulgaris (PV). Expanding the sample size is a key step to enhance the statistical power and generalizability of the results, as larger cohorts are more likely to account for individual variability and improve reproducibility. Furthermore, the inclusion of additional time points, particularly at earlier or later stages of disease progression, would provide a more comprehensive understanding of the temporal dynamics involved in PV pathophysiology.

A more diverse experimental model could also yield further insights into PV by incorporating interactions between keratinocytes and other cell types within the skin, such as immune cells, fibroblasts, and endothelial cells. This would offer a more holistic view of the disease process, as PV is likely driven by complex cellular interactions that extend beyond keratinocyte dysfunction alone.

Moreover, investigating a broader range of disease-related antibodies would be valuable, as PV patients present with diverse autoantibodies targeting various epitopes. Including multiple antibodies in future studies would allow for the examination of antibody-specific effects and their contribution to the pathogenesis of different forms of PV.

Finally, while RNA-seq and shotgun proteomics are powerful tools, they have inherent limitations in capturing the full scope of molecular changes. Future studies could complement these approaches with additional techniques such as single-cell proteomics, mass spectrometry-based phosphoproteomics, and spatial transcriptomics to gain a more comprehensive view of protein expression, post-translational modifications, and their dynamic changes across different regions of the skin.

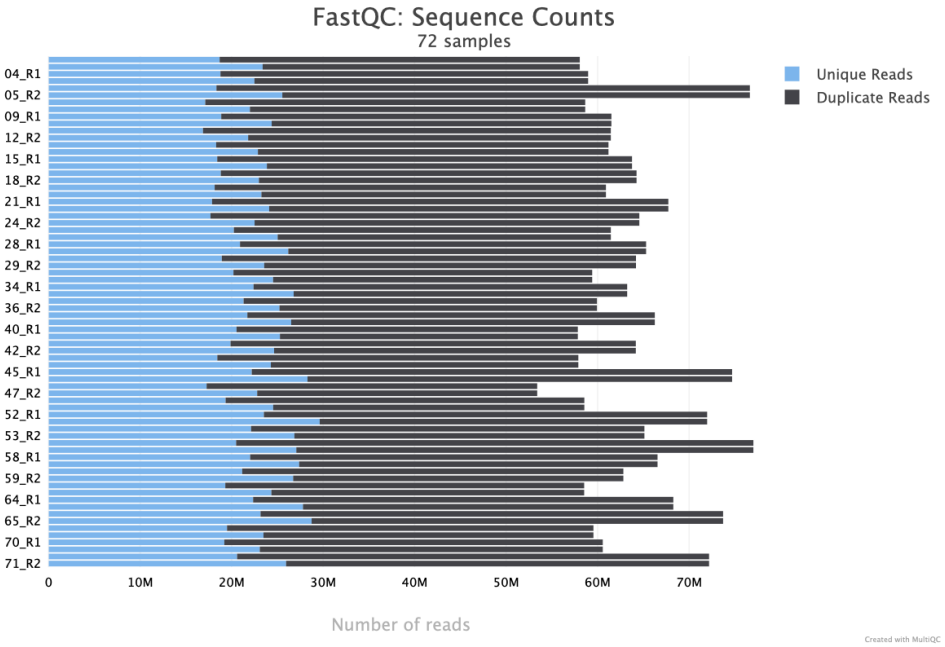
Pemphigus vulgaris pathogenesis is multifaceted, highlighting the need for further exploration of therapeutic targets and combination approaches to enhance treatment efficacy.

By addressing these limitations, future research will not only deepen our understanding of PV but also facilitate the development of targeted therapeutic strategies that address the root causes of this devastating autoimmune skin disorder.

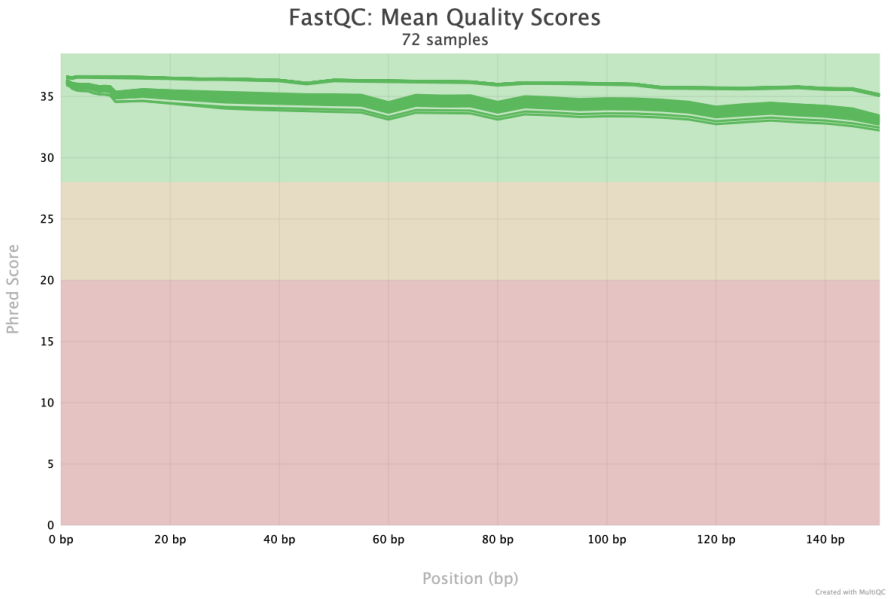
7. Appendix

7.1 Supplementary Figures

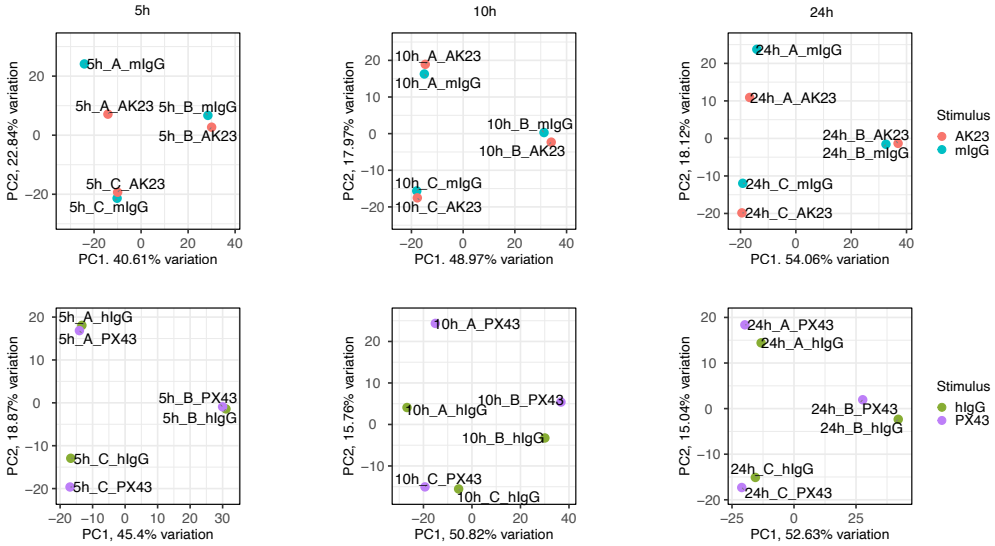
This section presents the supplementary figures referenced in the main chapters.



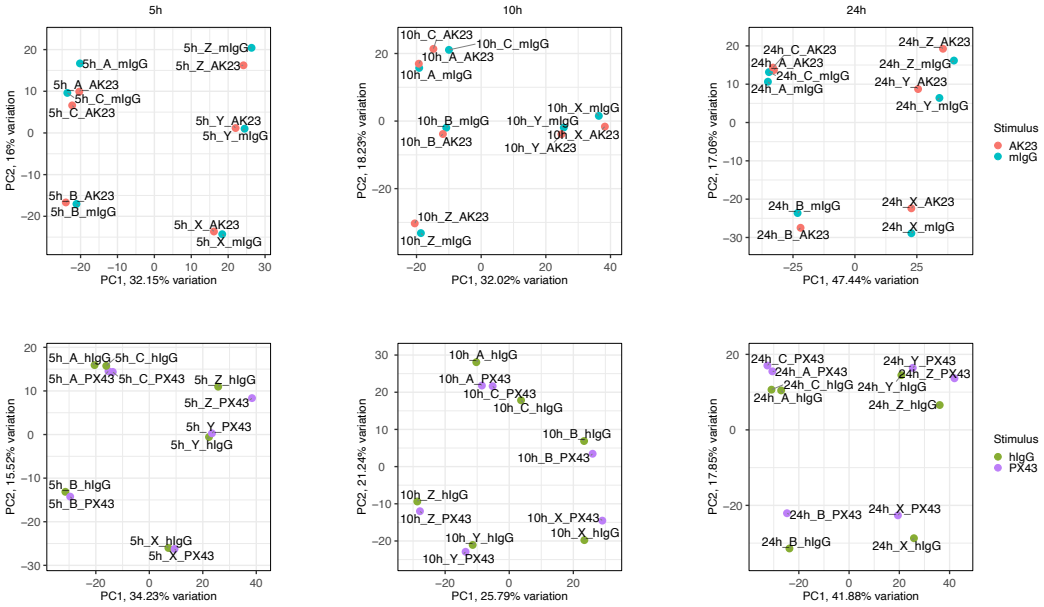
Supplemental Figure 1: Total read counts from bulk RNA-seq for HPEK samples A, B, and C. The quality control results aggregated by MultiQC from FastQC outputs indicate that all samples achieved sequencing depths exceeding 50 million reads, with the majority surpassing 60 million reads. These read counts meet or exceed the recommended thresholds for bulk RNA-seq experiments as outlined by ENCODE guidelines [132], suggesting sufficient sequencing depth for downstream analyses.



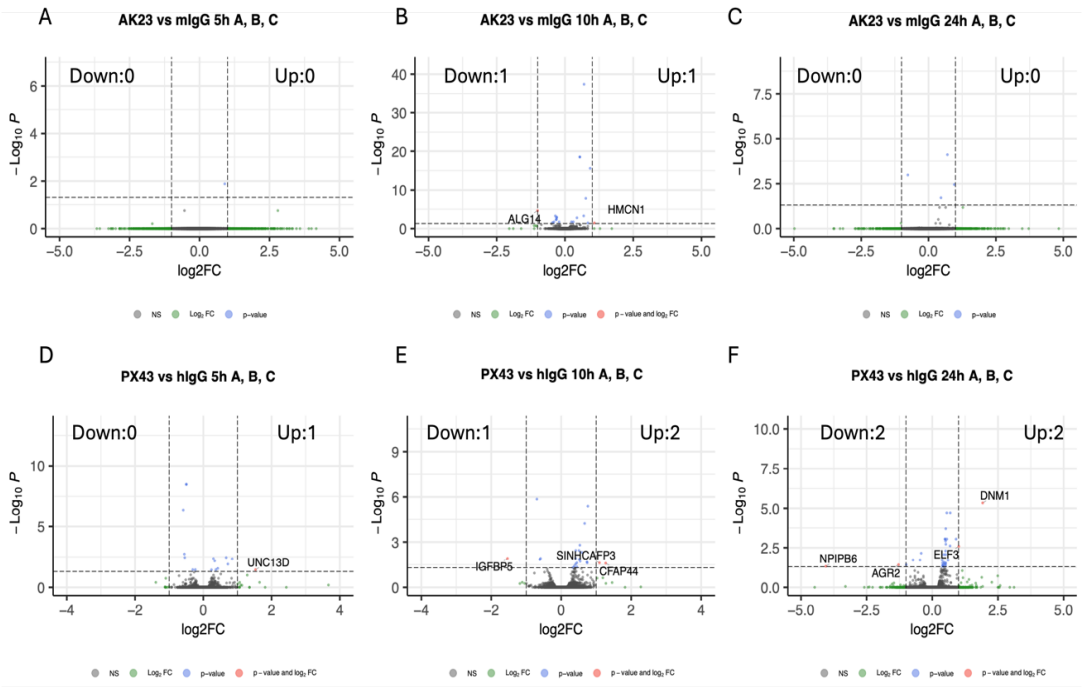
Supplemental Figure 2: Mean base quality scores across reads from bulk RNA-seq of HPEK samples A, B, and C. All samples exhibited mean quality scores above 33.8, as reported by FastQC and summarized by MultiQC. This corresponds to a base call accuracy exceeding 99.9%, indicating that the raw sequencing reads are of exceptionally high quality.



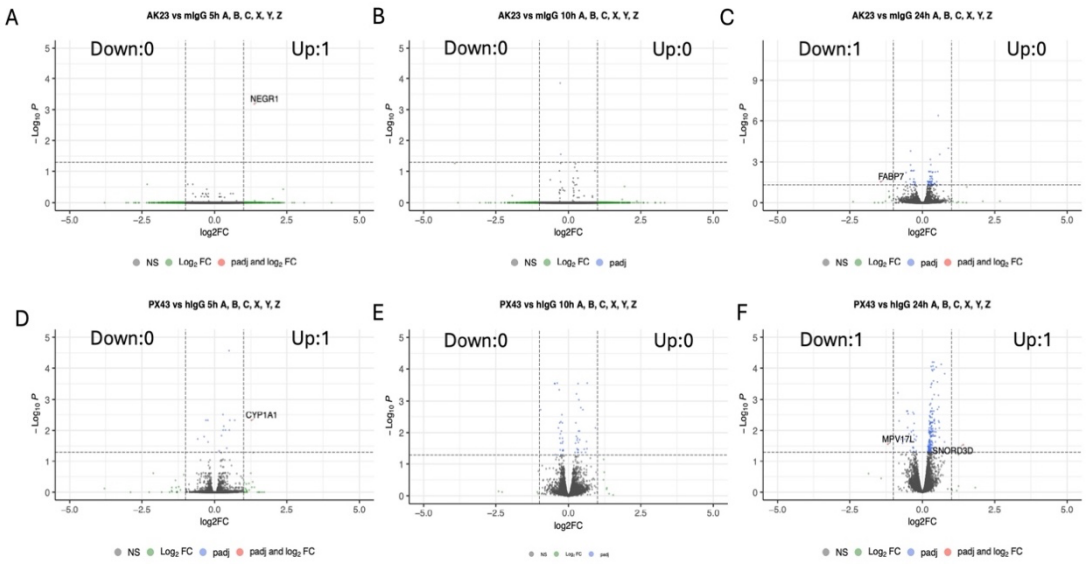
Supplemental Figure 3: Principal Component Analysis (PCA) plots comparing samples A, B, and C from experiment 1 in HPEK. The upper three plots show comparisons between AK23 and its control (mIgG), while the lower three plots display comparisons between PX43 and its control (hIgG). Although PC1 reflects the main transcriptional response to AK23 or PX43 stimulation, treated samples remain in close proximity to controls, suggesting only moderate global expression changes.



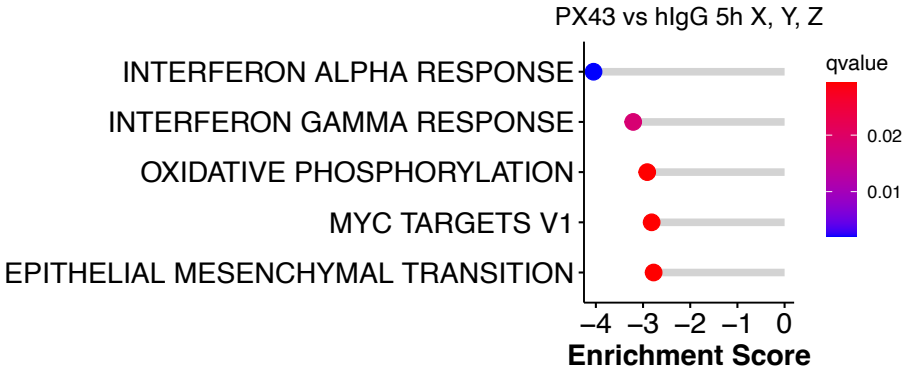
Supplemental Figure 4: Principal Component Analysis (PCA) plots comparing samples A, B, C, X, Y, and Z from experiment 1 and experiment 2 in HPEK. The upper three plots show comparisons between AK23 and its control (mIgG), while the lower three plots display comparisons between PX43 and its control (hIgG). Separation between samples occurs mainly along the first principal component (PC1), accounting for 25-48% of the total variance.



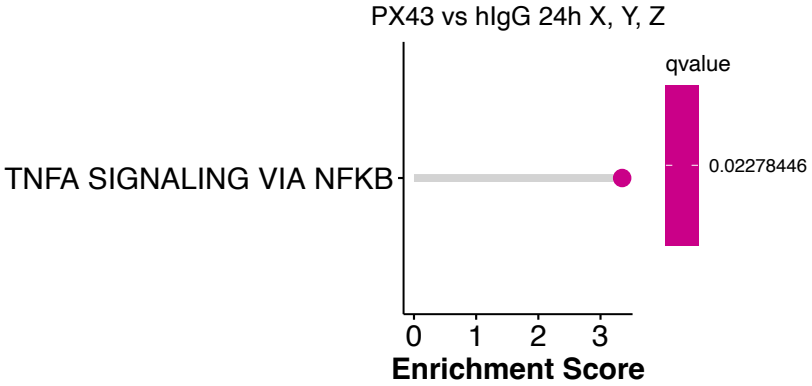
Supplemental Figure 5: Volcano plots of differentially expressed genes in samples A, B, and C comparing PV antibodies with control antibodies in HPEK. Each subplot displays the log2 fold-change (log2FC) on the x-axis and the negative log10 of the adjusted p-value (-log10 P) on the y-axis, with upregulated (Up) and downregulated (Down) genes indicated. Significant genes are labeled in each plot. AK23 vs. mIgG at 10h (Up: HMCN1, Down: ALG14); PX43 vs. hIgG at 5h (Up: UNC13D), 10h (Up: CFAP44, SINHCAFP3, Down: IGFBP5), 24h (Up: DNM1, ELF3, Down: AGR2, NPIPB6). Overall, only a few DEGs were identified.



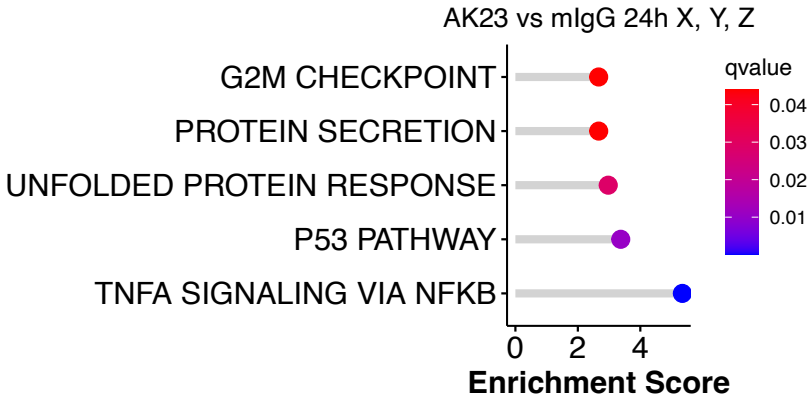
Supplemental Figure 6: Volcano plots of differentially expressed genes in samples A, B, C, X, Y, and Z comparing PV antibodies with control antibodies in HPEK. Each subplot displays the log2 fold-change (log2FC) on the x-axis and the negative log10 of the adjusted p-value (-log10 P) on the y-axis, with upregulated (Up) and downregulated (Down) genes indicated. Significant genes are labeled in each plot. AK23 vs. mIgG at 5h (Up: NEGR1), 24h (Down: FABP7); PX43 vs. hIgG at 5h (Up: CYP1A1), 24h (Up: SNORD3D, Down: MPV17L) Overall, only a few DEGs were identified.



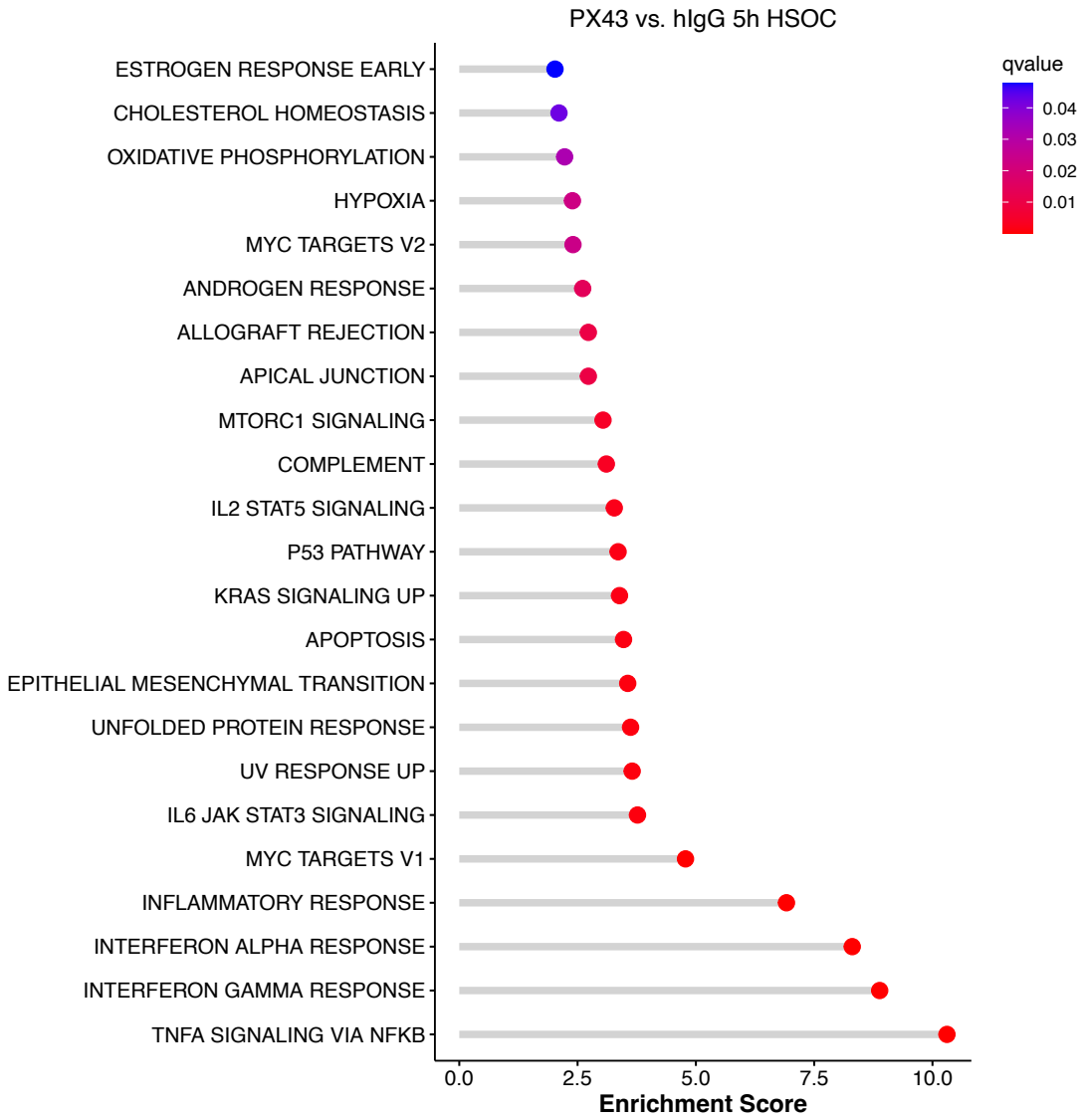
Supplemental Figure 7: Comparative pathway analysis of PX43 and hIgG treatments at 5 hours in samples X, Y, and Z. Downregulation of these pathways suggests a reduction in inflammatory response, compromised energy metabolism, and decreased epithelial-mesenchymal transition activity.



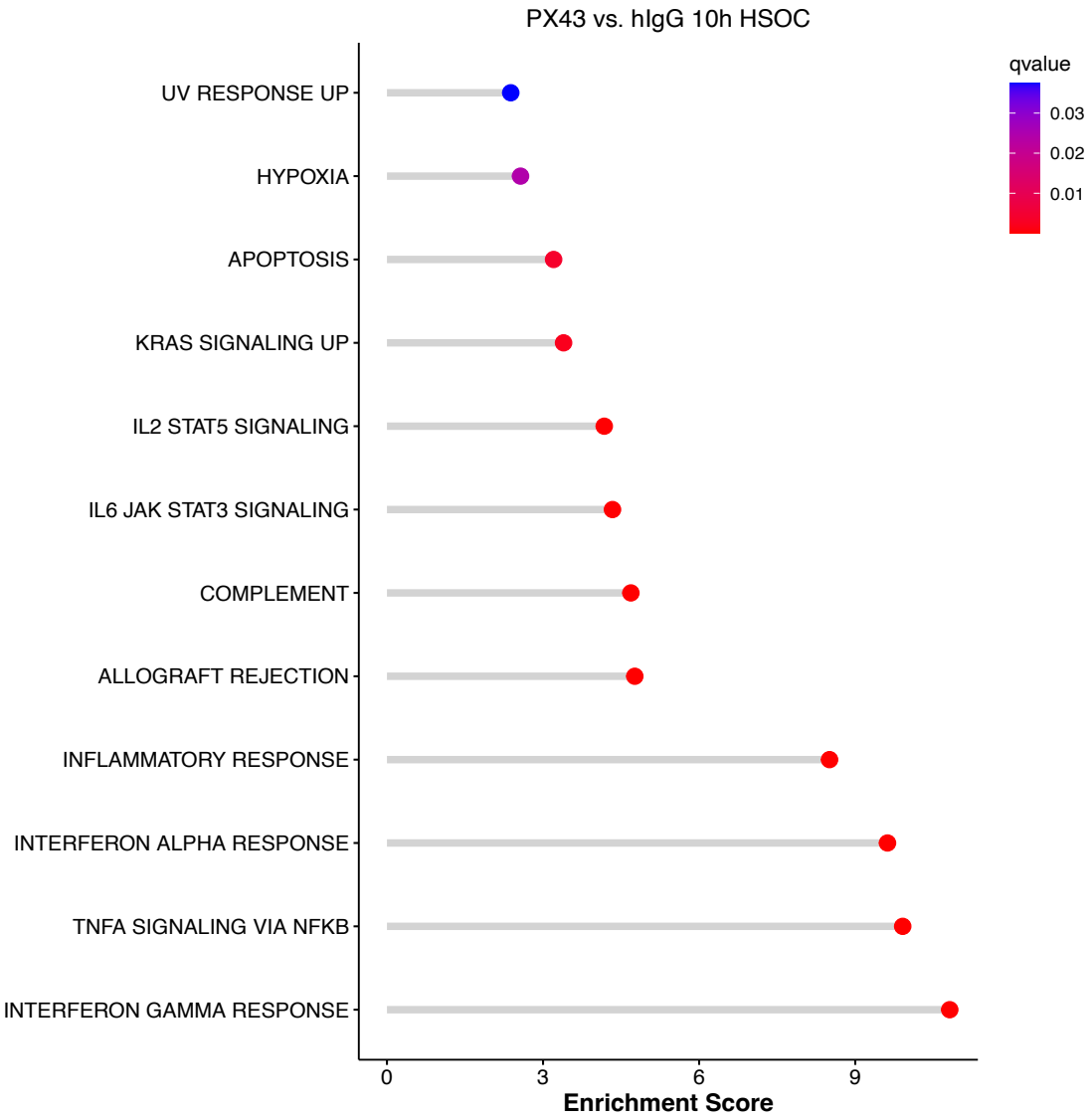
Supplemental Figure 8: Comparative pathway analysis of PX43 and hIgG treatments at 24 hours in samples X, Y, and Z. Upregulation of the TNFA pathway suggests an increase in inflammatory response.



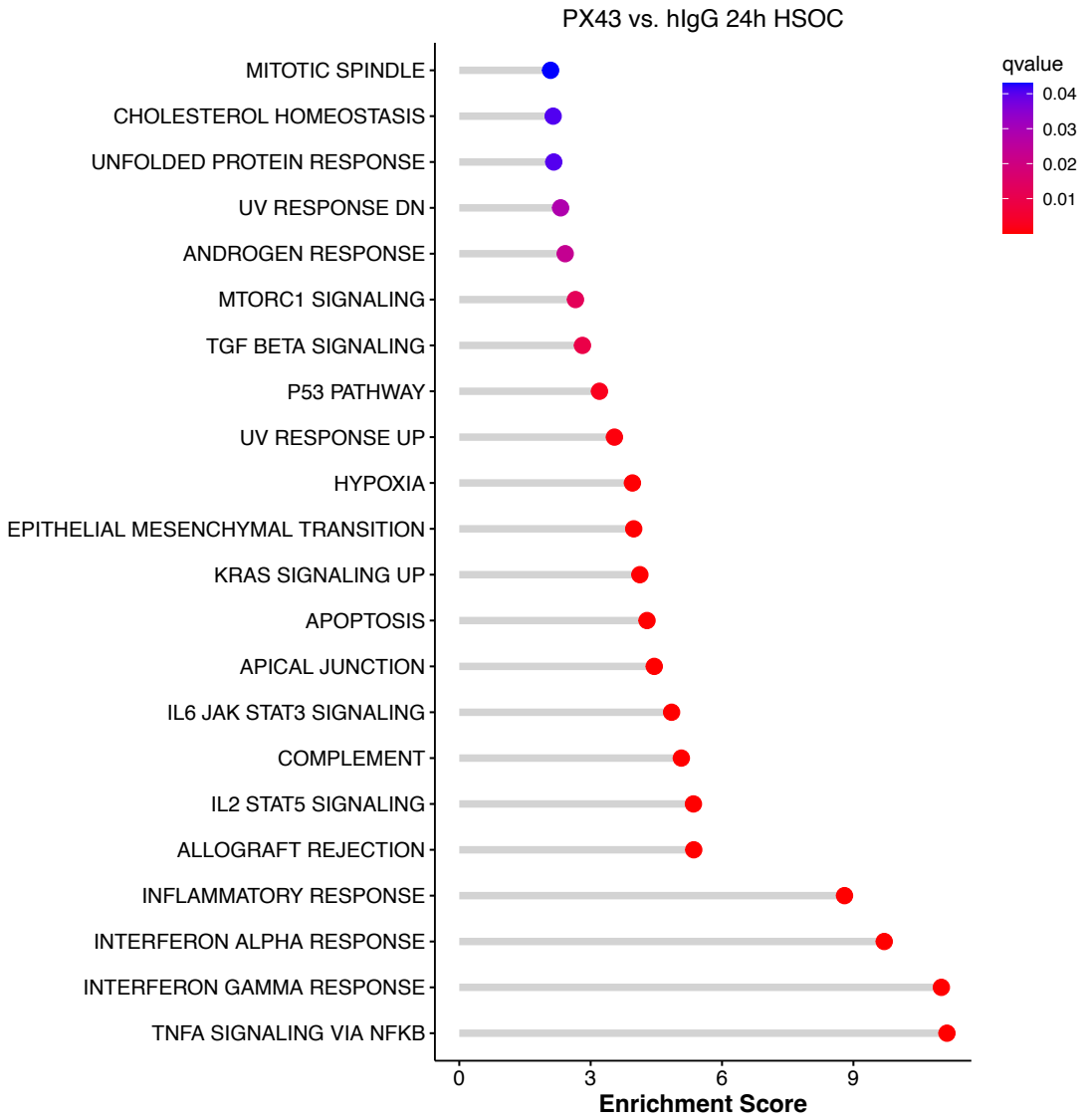
Supplemental Figure 9: Comparative pathway analysis of AK23 and mIgG treatments at 24 hours in samples X, Y, and Z. Upregulation of the TNFα pathway suggests an increase in inflammatory response, enhanced DNA damage response and apoptosis, elevated protein folding stress management, heightened protein export, and reinforced cell cycle control.



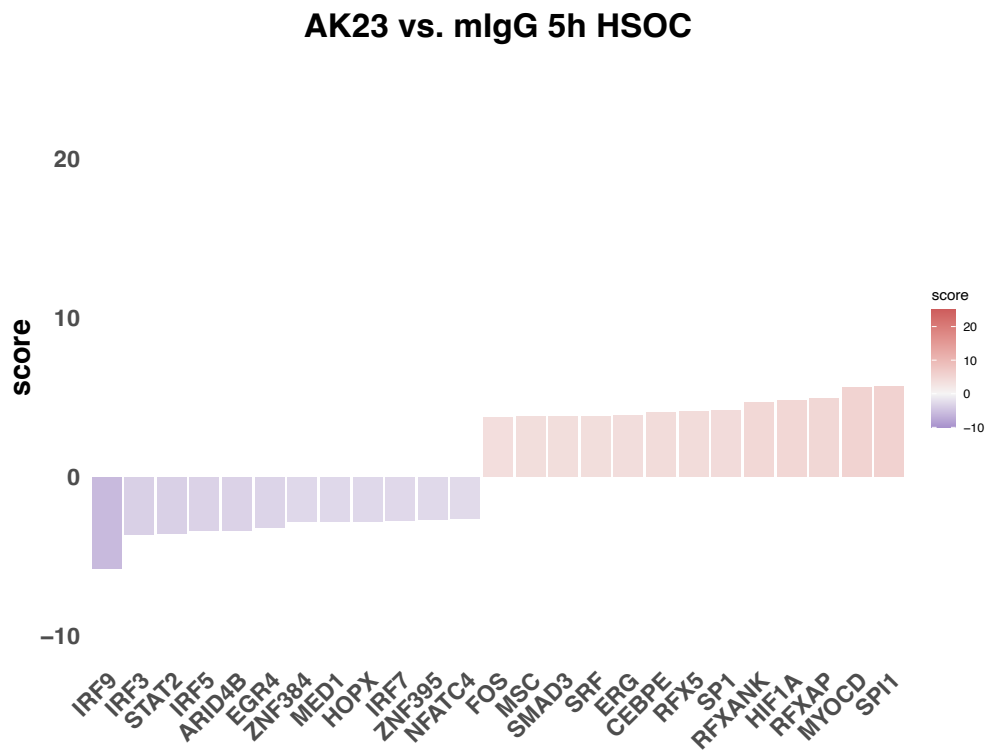
Supplemental Figure 10: Significantly enriched pathways of PX43 vs. hIgG 5h in human skin organ culture (HSOC) model. Pathways are from the HALLMRK collections, identified by the gage package, and the significance threshold is $q < 0.05$.



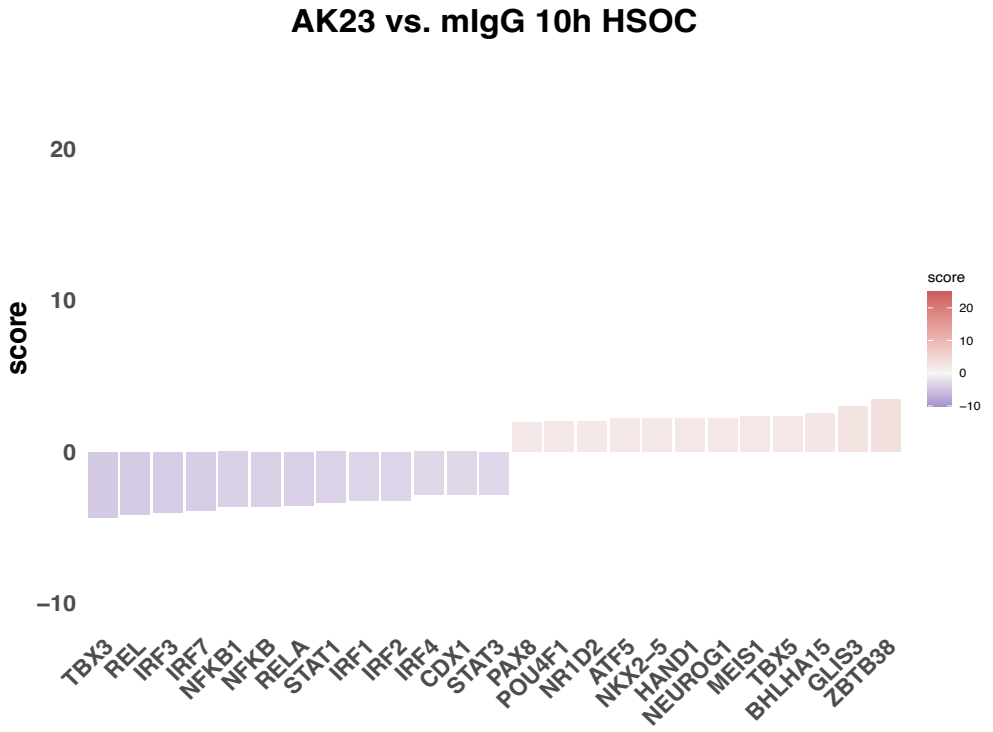
Supplemental Figure 11: Significantly enriched pathways of PX43 vs. hIgG 10h in human skin organ culture (HSOC) model. Pathways are from the HALLMRK collections, identified by the gage package, and the significance threshold is $q < 0.05$.



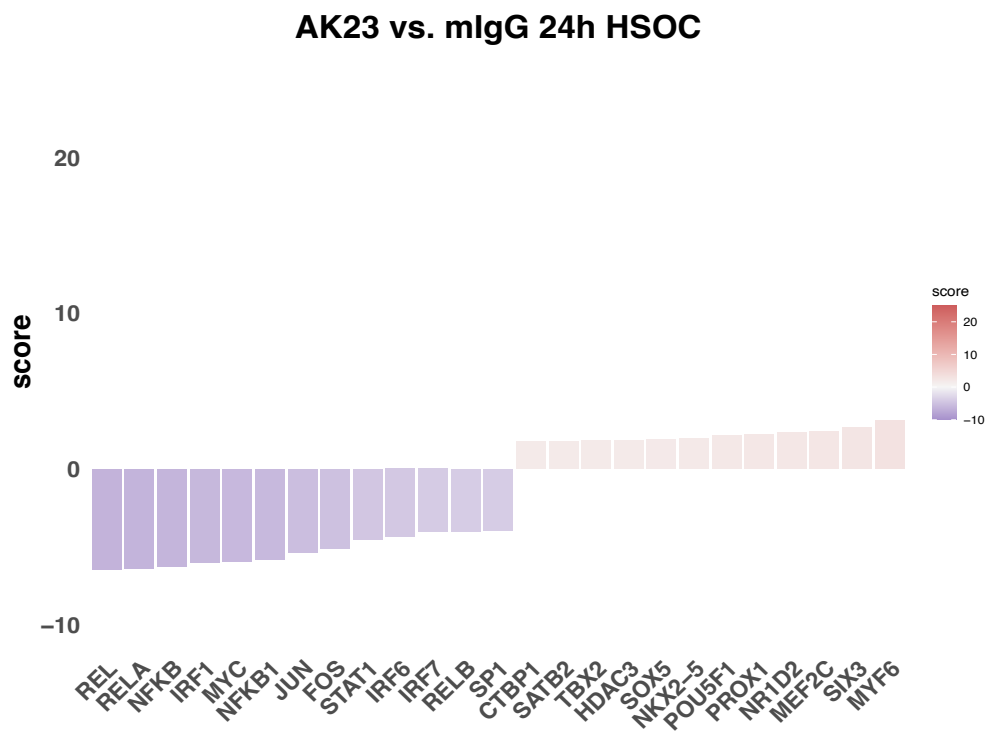
Supplemental Figure 12: Significantly enriched pathways of PX43 vs. hIgG 24h in human skin organ culture (HSOC) model. Pathways are from the HALLMRK collections, identified by the gage package, and the significance threshold is $q < 0.05$.



Supplemental Figure 13: Key differentially expressed transcription factors (TFs) in human skin organ culture (HSOC) at 5h for AK23 vs. mIgG. Top transcription factors were inferred using decoupleR (q-value < 0.05), based on differentially expressed genes and their corresponding log2 fold-change values. Predicted transcription factor activity scores are color-coded, with purple indicating negative activity scores and red indicating positive activity scores.



Supplemental Figure 14: Key differentially expressed transcription factors (TFs) in human skin organ culture (HSOC) at 10h for AK23 vs. mIgG. Top transcription factors were inferred using decoupleR (q-value < 0.05), based on differentially expressed genes and their corresponding log2 fold-change values. Predicted transcription factor activity scores are color-coded, with purple indicating negative activity scores and red indicating positive activity scores.



Supplemental Figure 15: Key differentially expressed transcription factors (TFs) in human skin organ culture (HSOC) at 24h for AK23 vs. mIgG. Top transcription factors were inferred using decoupleR (q-value < 0.05), based on differentially expressed genes and their corresponding log2 fold-change values. Predicted transcription factor activity scores are color-coded, with purple indicating negative activity scores and red indicating positive activity scores.

7.2 Supplementary Tables

This section presents the supplementary tables referenced in the main chapters.

Supplementary Table 1: Summary of the HALLMARK gene sets

HALLMARK NAME	PROCESS CATEGORY	DESCRIPTION	NUMBER OF FOUNDER SETS	NUMBER OF GENES
APICAL_JUNCTION	cellular component	Apical junction complex consisting of adherens and tight junctions	37	200
APICAL_SURFACE	cellular component	Membrane proteins in the apical domain	12	44
PEROXISOME	cellular component	Peroxisomes	28	107
ADIPOGENESIS	development	Adipocyte development	36	200
ANGIOGENESIS	development	Blood vessel formation	14	36
EPITHELIAL_MESENCHYMAL_TRANSITION	development	Epithelial mesenchymal transition	107	200
MYOGENESIS	development	Muscle differentiation	64	200

SPERMATOGENESIS	development	Sperm development and male fertility	24	135
PANCREAS_BETA_CELL	development	Genes specific to pancreatic beta cells	24	40
DNA_REPAIR	DNA damage	DNA repair	44	150
UV_RESPONSE_DOWN	DNA damage	UV response: down-regulated genes	17	144
UV_RESPONSE_UP	DNA damage	UV response: up-regulated genes	16	158
ALLOGRAFT_REJECTION	immune	Allograft rejection	190	200
COAGULATION	immune	Coagulation cascade	71	138
COMPLEMENT	immune	Complement cascade	71	200
INTERFERON_ALPHA_RESPONSE	immune	Interferon alpha response	82	97
INTERFERON_GAMMA_RESPONSE	immune	Interferon gamma response	82	200
IL6_JAK_STAT3_SIGNALING	immune	IL6 STAT3 signaling during acute phase response	24	87
INFLAMMATORY_RESPONSE	immune	Inflammation	120	200
BILE_ACID_METABOLISM	metabolic	Biosynthesis of bile acids	28	112
CHOLESTEROL_HOMEOSTASIS	metabolic	Cholesterol homeostasis	28	74
FATTY_ACID_METABOLISM	metabolic	Fatty acid metabolism	53	158
GLYCOLYSIS	metabolic	Glycolysis and gluconeogenesis	87	200
HEME_METABOLISM	metabolic	Heme metabolism	36	200
OXIDATIVE_PHOSPHORYLATION	metabolic	Oxidative phosphorylation and citric acid cycle	93	200
XENOBIOTIC_METABOLISM	metabolic	Metabolism of xenobiotics	124	200
APOPTOSIS	pathway	Programmed cell death; caspase pathway	80	161
HYPOXIA	pathway	Response to hypoxia; HIF1A targets	87	200
PROTEIN_SECRETION	pathway	Protein secretion	74	96
UNFOLDED_PROTEIN_RESPONSE	pathway	Unfolded protein response; ER stress	22	113
REACTIVE_OXYGEN_SPECIES_PATHWAY	pathway	Reactive oxygen species (ROS) pathway	13	49
E2F_TARGETS	proliferation	E2F targets	420	200
G2M_CHECKPOINT	proliferation	Cell cycle G2/M checkpoint	420	200
MYC_TARGETS_V1	proliferation	MYC targets variant 1	404	200
MYC_TARGETS_V2	proliferation	MYC targets variant 2	6	58
P53_PATHWAY	proliferation	p53 pathway	85	200

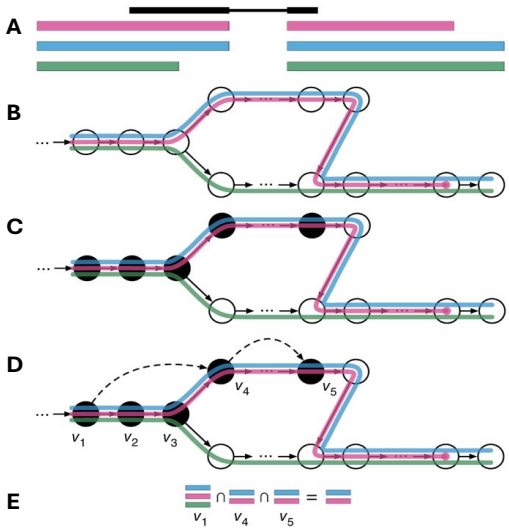
MITOTIC_SPINDLE	proliferation	Mitotic spindle assembly	108	200
ANDROGEN_RESPONSE	signaling	Androgen response	8	117
ESTROGEN_RESPONSE_EARLY	signaling	Early estrogen response	61	200
ESTROGEN_RESPONSE_LATE	signaling	Late estrogen response	61	200
IL2_JAK_STAT5_SIGNALING	signaling	IL2 STAT5 signaling	13	200
KRAS_SIGNALING_UP	signaling	KRAS signaling, up-regulated genes	14	200
KRAS_SIGNALING_DOWN	signaling	KRAS signaling, down-regulated genes	16	200
MTORC1_SIGNALING	signaling	mTORC1 signaling	487	200
NOTCH_SIGNALING	signaling	Notch signaling	49	32
PI3K_AKT_MTOR_SIGNALING	signaling	PI3K signaling via AKT to mTORC1	591	105
HEDGEHOG_SIGNALING	signaling	Hedgehog signaling	79	36
TGF_BETA_SIGNALING	signaling	TGF beta signaling	29	54
TNFA_SIGNALING_VIA_NFKB	signaling	TNFA signaling via NFkB	132	200
WNT_BETA_CATENIN_SIGNALING	signaling	Cannonical beta catenin pathway	49	42

7.3 Supplementary Methods

This section presents the supplementary methods referenced in the main chapters.

7.3.1 Supplementary Method 1: Transcript Quantification using Kallisto

An overview of Kallisto methodology is presented in Supplementary Figure 16, extracted from paper by Bray et al. (2016) [51].



Supplemental Figure 16: Overview of kallisto pseudoalignment algorithm. The input consists of a reference transcriptome and reads from an RNA-seq experiment. (**A**) An example of a read (in black) and three overlapping transcripts with exonic regions as shown. (**B**) An index is constructed by creating the transcriptome de Bruijn Graph (T-DBG) where nodes (v_1, v_2, v_3, ...) are *k*-mers, each transcript corresponds to a coloured path as shown and the path cover of the transcriptome induces a *k*-compatibility class for each *k*-mer. (**C**) Conceptually, the *k*-mers of a read are hashed (black nodes) to find the *k*-compatibility class of a read. (**D**) Skipping (black dashed lines) uses the information stored in the T-DBG to skip *k*-mers that are redundant because they have the same *k*-compatibility class. (**E**) The *k*-compatibility class of the read is determined by taking the intersection of the *k*-compatibility classes of its constituent *k*-mers. The figure is extracted from the paper of kallisto by Bray et al. (2016) [51].

Kallisto operates in three main stages: index construction, pseudoalignment, and quantification. During index construction, kallisto builds a coloured de Bruijn graph [246] from the transcriptome, grouping k-mers into contigs with identical colourings and storing a hash table for a fast lookup of k-mer positions. Pseudoalignment maps sequencing reads to transcripts by identifying shared k-mers without requiring full alignment, leveraging the de Bruijn graph structure to minimize computational overhead. Quantification uses an efficient EM algorithm to estimate transcript abundances based on equivalence class counts, outputting results in transcripts per million (TPM) units [247]. Kallisto also includes optional steps for bias correction, which adjusts transcript effective lengths based on sequence-specific biases, and bootstrapping, which provides uncertainty estimates by resampling equivalence class counts to compute expression variability. Including bias correction and bootstrapping enhances the accuracy and reliability of kallisto's quantification.

The analysis workflow utilizing Kallisto follows the guidelines outlined in the official Kallisto manual. The process begins with constructing a human transcriptome index, which serves as the reference for subsequent quantification. For the quantification step, paired-end FASTQ files corresponding to each sample are provided as input.

During the quantification process, Kallisto generates three output files for each sample: abundance.h5, abundance.tsv and run_info.json. The abundance.h5 file is a binary format containing detailed quantification data, while the abundance.tsv file provides tab-delimited results, including transcript-level abundances in transcripts per million (TPM). The run_info.json file captures metadata about the quantification process, including the Kallisto version and the command-line parameters used.

7.3.2 Supplementary Method 2: Principal Component Analysis (PCA)

The following outlines the step-by-step process for performing PCA:

1) Data Preprocessing: Centering the Data

- Compute the mean for each variable in the dataset.
- Subtract the mean from each observed value of the corresponding variable. This step centers the data around zero, ensuring that PCA focuses on variance rather than absolute values.

2) Covariance Matrix Computation

- Calculate the covariance matrix of the centered data. The covariance matrix captures the relationships between variables and their joint variances, providing the foundation for identifying directions of maximum variance.

3) Eigenvalues and Eigenvectors Calculation

- Compute the eigenvalues and corresponding eigenvectors of the covariance matrix. The eigenvalues quantify the variance explained by their associated eigenvectors, while eigenvectors define the directions of maximum variance (principal components).

4) Normalization of Eigenvectors

- Normalize each eigenvector to unit length. These normalized eigenvectors represent the axes of the fitted ellipsoid and form a new orthonormal basis for the dataset.

5) Principal Components Selection

- Rank the eigenvectors by their associated eigenvalues in descending order. The eigenvector with the largest eigenvalue corresponds to the first principal component, explaining the maximum variance.
- Select the top principal components that capture the desired proportion of the total variance in the data.

6) Data Transformation

- Project the centered data onto the selected principal components by calculating the dot product of the data with the eigenvectors. This results in a transformed dataset with reduced dimensions.

7.3.3 Supplementary Method 3: Generalized Linear Model (GLM) in RNA-Seq Data Analysis

The generalized linear model (GLM) provides a flexible framework for modeling count-based data. In the case of RNA-seq, GLMs are commonly used with the negative binomial (NB) distribution to model the count data. The negative binomial distribution is ideal because it can model overdispersion, which occurs when the gene expression variance is larger than the mean, a common feature in RNA-seq data.

In the GLM framework, the log of the expected count for gene i in condition j, $E\left[Y_{ij}\right]$ is modelled as a linear function of covariates (e.g., experimental conditions or treatments):

$$\log\left(\mathrm{E}\left[\mathrm{Y_{ij}}\right]\right) = x_j^T \beta_i$$

where:

log generates the logarithm symbol.

$E\left[Y_{ij}\right]$ is the count of reads for gene i in sample j.

x_j is the vector of covariates (such as treatment or time).

β_i represents the regression coefficients for gene i.

T generates the transpose symbol for x_j^T.

In RNA-seq analysis, the GLM is often extended by incorporating the negative binomial distribution for the variance structure, as this distribution accounts for both the mean and variance of the counts. The likelihood function for the negative binomial model is:

$$f(Y_{ij}|\mu_{ij},\phi) = \frac{\Gamma(Y_{ij}+\phi)}{\Gamma(Y_{ij}+1)\Gamma(\phi)}\left(\frac{\mu_{ij}}{\mu_{ij}+\phi}\right)^{Y_{ij}}\left(\frac{\phi}{\mu_{ij}+\phi}\right)^{\phi}$$

where:

Y_{ij} is the observed count for gene i in sample j.

μ_{ij} is the expected count (mean) for gene i in sample j.

ϕ is the dispersion parameter, capturing overdispersion.

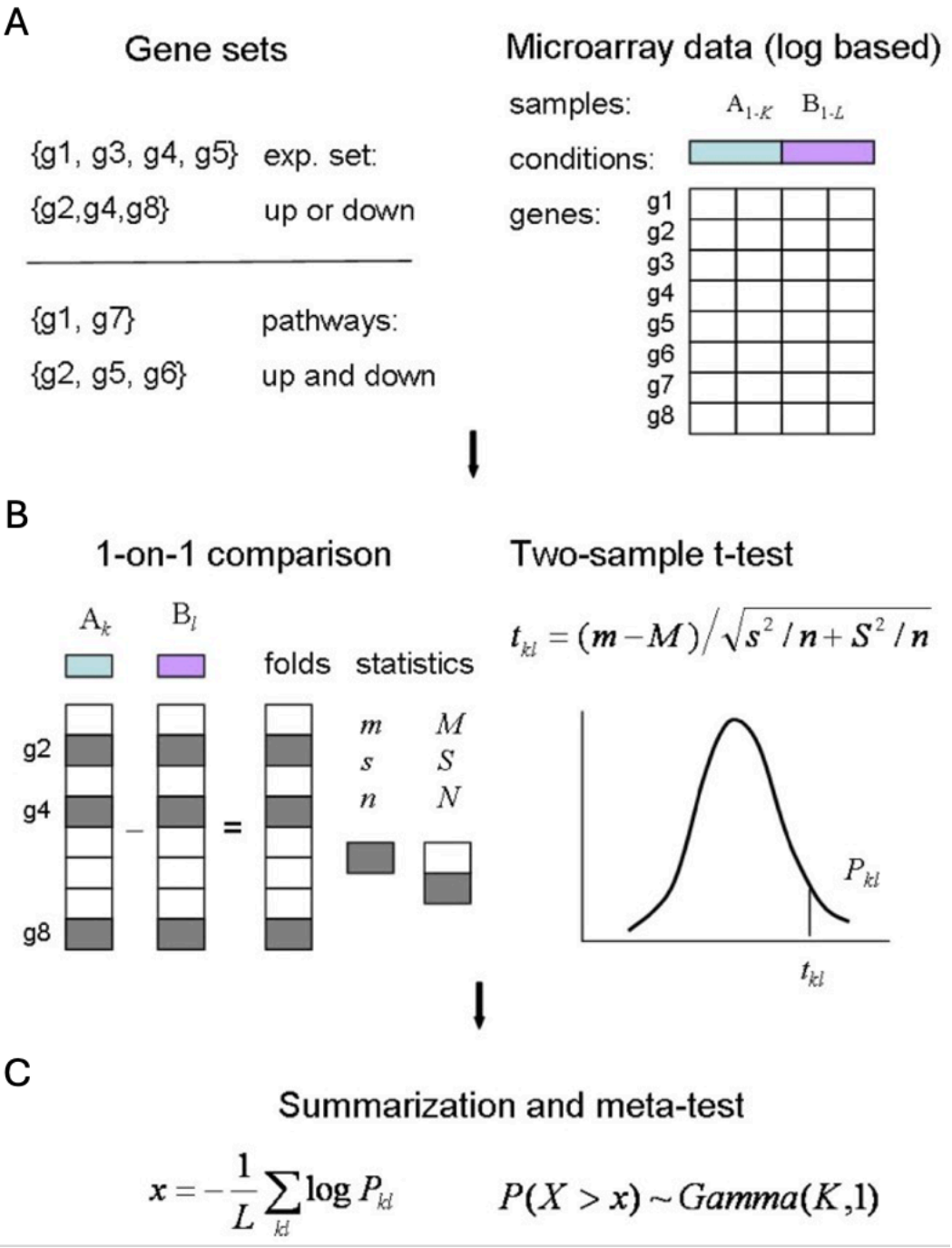
Γ is the gamma function, a continuous extension of the factorial function, for an integer n, $\Gamma(n) = (n-1)!$.

The first part, $\frac{\Gamma(Y_{ij}+\phi)}{\Gamma(Y_{ij}+1)\Gamma(\phi)}$, serves as a normalization factor to ensure that the probabilities sum to 1. The second part, $\left(\frac{\mu_{ij}}{\mu_{ij}+\phi}\right)^{Y_{ij}}$, represents the probability of observing Y_{ij} successes (counts) given the expected count and dispersion. The third part, $\left(\frac{\phi}{\mu_{ij}+\phi}\right)^{\phi}$, accounts for the probability of the remaining counts, reflecting the distribution's shape influenced by the dispersion parameter.

7.3.4 Supplementary Method 4: Gene Set Enrichment Analysis (GAGE) for Pathway Enrichment Analysis

GAGE employs a rank-based approach to enrichment analysis and incorporates methodological enhancements that make it particularly suitable for datasets with heterogeneous sample sizes and complex experimental designs [59]. Differentially expressed genes (DEGs) identified using DESeq2, along with their corresponding $\log_2$ fold change values, were used as input for the pathway enrichment analysis by GAGE. The Hallmark gene set collection from the MSigDB database was used as the reference for pathway annotation. For each pathway, enrichment scores, p-values, q-values, and additional statistical metrics were calculated to determine significant pathway-level changes.

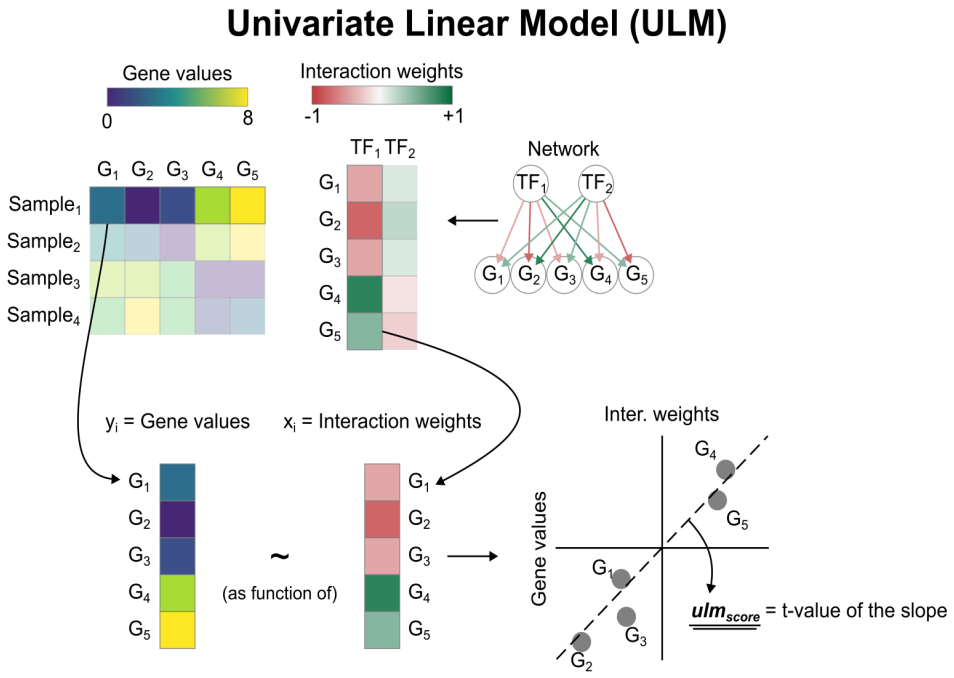
A schematic overview of the GAGE algorithm, including its underlying methodology and key computational features, is presented in Supplemental Figure 17.



Supplemental Figure 17: A schematic overview of the gage algorithm.Three major steps are applied in gage. Three major steps are applied in gage**.** (A) Step 1: input preparation. Separate gene sets into two categories: experimental sets and canonical pathways, for differential treatment in significant test. (B) Step 2: gene set differential expression tests based on one-on-one comparison between samples from the two experimental conditions. For each experiment-control pair, calculate differential expression in log-based fold change for all genes. Test whether specific gene sets are significantly differentially expressed relative to the background whole set using two-sample t-test. (C) Step 3: summarization. For each gene set, derive a global p-value based on a meta-test on the negative log sum of p-values from all one-on-one comparisons. More details of GAGE are given in the Methods. Variables m, s and n are the mean fold change, standard deviation and number of genes in a gene set, M, S and N are those for the whole set. Figure is extracted from the original paper for Luo et al. (2009) [59].

7.3.5 Supplementary Method 5: Univariate Linear Model (ULM) for Transcription Factor Activity Prediction Using decoupleR

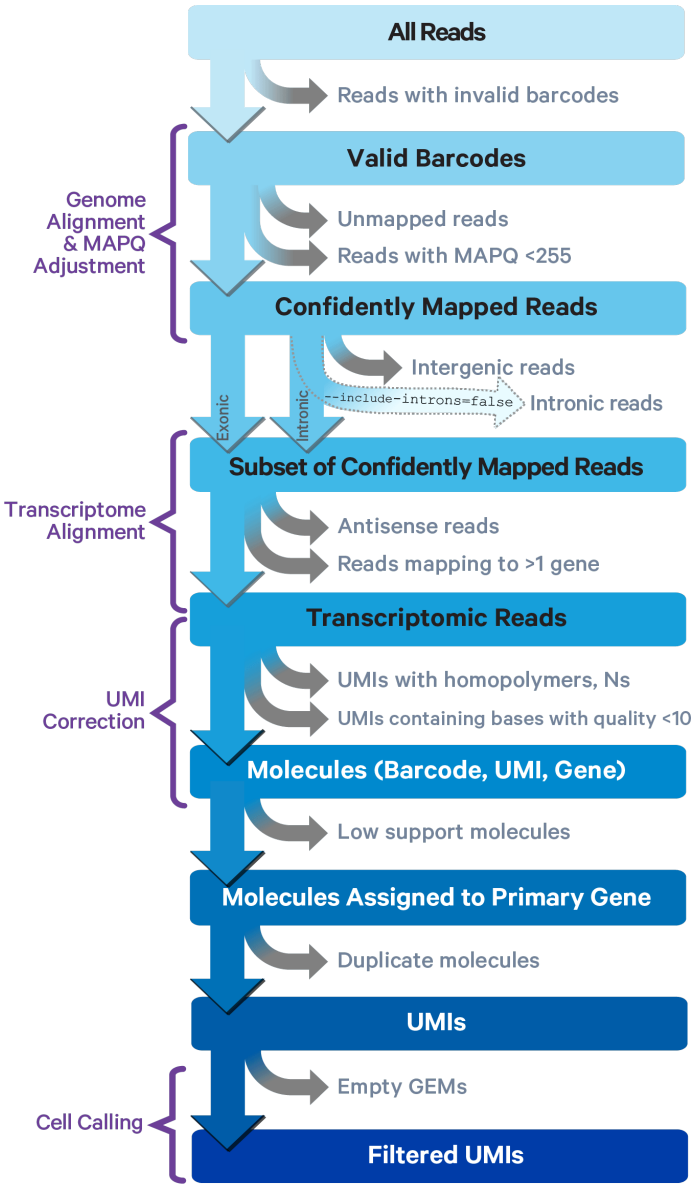
The univariate linear model (ULM) method is employed in decoupleR [68] to infer transcription factor (TF) enrichment scores for each sample in the dataset and each TF within the TF-gene interaction network. This approach fits a linear model that predicts observed gene expression based solely on the TF's interaction weights with its target genes. Once the model is fitted, the t-value of the slope serves as the enrichment score. A positive t-value indicates TF activation, while a negative t-value suggests TF inactivity. A schematic representation of this method is provided in Supplementary Figure 18.



Supplemental Figure 18: Inferring transcription factor activity using the univariate linear model (ULM). Gene expression values are modeled as a function of TF interaction weights, with the t-value of the regression slope serving as the enrichment score. A positive t-value indicates TF activation, while a negative t-value suggests inactivation. The figure was extracted from the decoupleR manual website by Pau Mompel (2024) [70].

7.3.6 Supplementary Method 6: Single-Cell RNA-seq Data Quantification Using Cell Ranger

The primary steps involved in raw sequencing reads processing by Cell Ranger [101] are summarized in Supplementary Figure 19.



Supplemental Figure 19: Key steps in reads processing by Cell Ranger. The main steps in read processing by Cell Ranger include aligning raw reads to the reference genome, filtering out low-quality reads, correcting unique molecular identifiers (UMIs), and assigning reads to individual cells based on UMI barcodes. The figure is extracted from the 10x genomics website by Jackman et al. (2023) [248].

7.3.7 Supplementary Method 7: Single-Cell RNA-seq Data Analysis Using Seurat

Seurat is used for downstream analyses, filtered barcode, feature, and matrix files generated by Cell Ranger were imported into Seurat. Downstream analysis including quality control (default parameters for the number of unique genes detected in each cell, total number of molecules detected within a cell and percentage of reads that map to the mitochondrial genome), normalization (LogNormalize), feature selection (VST, variance stabilizing transformation method), dimensionality reduction by UMAP (Uniform Manifold Approximation and Projection), clustering, differential gene expression analysis (Wilcoxon Rank Sum test) and visualization [129].

The key steps included in Seurat are as follows:

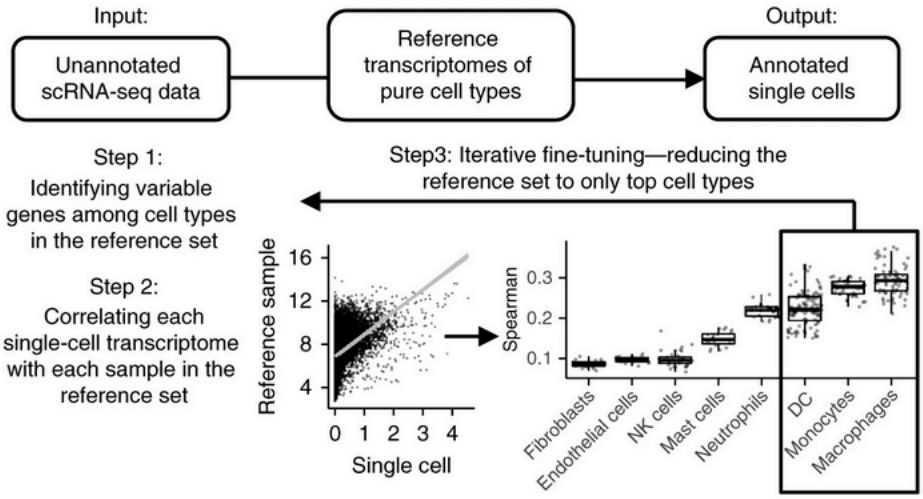
- Quality Control (QC) and Filtering. The number of unique genes detected in each cell, the total number of molecules detected within a cell, and the percentage of reads mapping to the mitochondrial genome are three important parameters for the initial QC steps to filter out low-quality cells, empty droplets, and dying cells that often exhibit extensive mitochondrial contamination. This can be effectively visualized using the VlnPlot function in Seurat, where the threshold can be set based on the values in the plot.
- Normalizing the data. By default, the global-scaling normalization method 'LogNormalize' normalizes feature expression for each cell by dividing by the total transcript count, multiplying by a scaling factor (10,000 by default), and applying a log transformation. Normalization can be performed using the NormalizeData function in Seurat.
- Identification of highly variable features (feature selection). A subset of features exhibiting high cell-to-cell variation within the dataset is then identified, characterized by high expression in some cells and low expression in others. Focusing on these highly variable genes in downstream analysis has been shown to enhance the detection of biological signals in single-cell datasets. Feature selection is performed using the FindVariableFeatures function in Seurat.

- Scaling the data. This is a standard pre-processing step before dimensional reduction techniques like PCA. This step gives equal weight in downstream analyses so that highly expressed genes do not dominate. Scaling the data is performed using the ScaleData function in Seurat.
- Linear dimensional reduction. The previously identified variable features serve as input for principal component analysis (PCA). For the leading principal components, Seurat generates a list of genes with the highest positive and negative loadings, representing gene modules that exhibit correlation or anti-correlation across single cells in the dataset. Linear dimensionality reduction is conducted using the RunPCA function in Seurat.
- Determine the dimensionality of the data. To mitigate the impact of technical noise in single-cell RNA sequencing (scRNA-seq) data, Seurat clusters cells based on their principal component (PC) scores, where each PC serves as a 'metafeature' that integrates information from a correlated set of features. The leading principal components thus provide a robust compression of the dataset. A commonly used heuristic approach for selecting the optimal number of PCs is the 'Elbow Plot,' which ranks PCs based on the percentage of variance explained. The inflection point, where the curve begins to plateau, indicates the number of principal components that capture the majority of the biological signal.
- Cluster the cells. Seurat employs a graph-based clustering approach by constructing a K-nearest neighbor (KNN) graph, where edges connect cells with similar feature expression patterns. The graph is first built using Euclidean distances in PCA space and refined by adjusting edge weights based on shared local neighborhoods (Jaccard similarity) via the FindNeighbors function, using the top 10 principal components. Clustering is then performed using modularity optimization techniques, such as the Louvain algorithm (default), to partition the graph into highly interconnected communities. The FindClusters function implements this process, with the resolution parameter controlling clustering granularity, where higher values yield more clusters.

- Non-linear dimensional reduction. This step aims to identify the underlying structure within the dataset, enabling the spatial organization of similar cells in a low-dimensional representation. Cells assigned to the same graph-based clusters should exhibit co-localization within these dimensionality reduction plots. Uniform Manifold Approximation and Projection (UMAP) is one such algorithm that facilitates this process, and it can be implemented using the RunUMAP function in Seurat.
- Finding differentially expressed features. By default, this method identifies both positive and negative markers for a given cluster in comparison to all other cells. The FindAllMarkers function automates this process across all clusters; however, it also allows for differential expression testing between specific groups of clusters or against the entire dataset.
- Assigning cell type to clusters. This step requires canonical markers to accurately associate unbiased clustering results with known cell types. In Seurat, well-established reference markers are available for certain tissues, such as peripheral blood mononuclear cells (PBMCs) and liver cells. However, there is currently no universally accepted gold standard for skin cell markers, necessitating caution when utilizing markers from external sources.

7.3.8 Supplementary Method 8: Single-Cell RNA-seq Data Cell Type Prediction Using SingleR

Supplementary Figure 20 shows a schematic representation of the SingleR [103] workflow.



Supplemental Figure 20: Schematic of SingleR for cell type annotation. The primary steps of the SingleR workflow include initially identifying variable genes that distinguish cell types within the reference dataset. Next, each single-cell transcriptome is correlated with individual samples from the reference set. This is followed by an iterative fine-tuning process that refines the reference set by retaining only the top-matching cell types, thereby improving annotation accuracy. The figure is extracted from Aran et al. (2019) [103].

7.3.9 Supplementary Method 9: Differentially Abundant Cell Groups Identification Using DA-seq

Identifying differentially abundant (DA) subpopulations between biological states, like healthy versus diseased, is challenging in single-cell RNA sequencing (scRNA-seq). Traditional methods rely on clustering cells, but this often misses DA subpopulations that don't align with clear clusters. DA-seq is a multiscale approach that calculates a local DA measure for each cell based on its k nearest neighbours across various k values [104]. This allows DA-seq to identify contiguous DA subpopulations in transcriptomic space without being restricted to clusters.

The DA-seq algorithm consists of several steps:

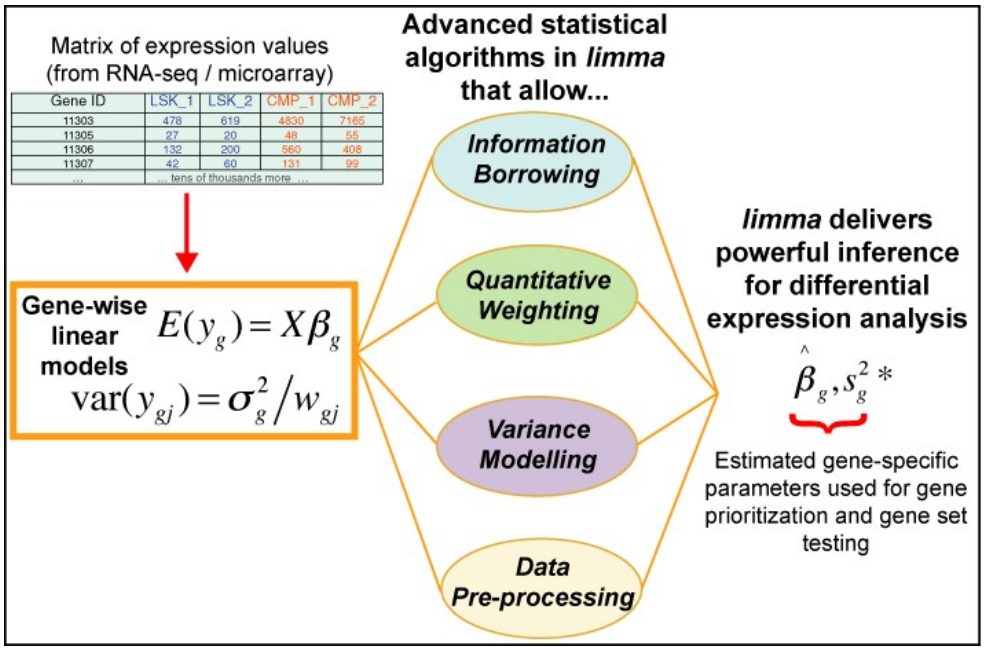
- Compute a score vector for each cell based on the relative prevalence of cells from both biological states in the cell's neighborhood. This measure is calculated for neighborhoods of varying sizes, providing a multiscale view of differential abundance for each cell.

- Combine the score vectors into a single DA measure by training a multivariate logistic regression classifier. This classifier predicts each cell's biological state (state 1 or state 2) based on its score vector. The resulting prediction probability is transformed into a DA measure, quantifying how much a cell's neighborhood is dominated by cells from one biological state.
- Group cells with similar DA measures into localized regions, based on their gene expression profiles. These regions represent subpopulations with significant differences in abundance between the two biological states. Each DA subpopulation is assigned a DA score, along with a P-value to assess reproducibility if sufficient biological replicates are available.
- Identify genes that distinguish each DA subpopulation from the rest of the cells or its immediate neighborhood. This step highlights the genetic features underlying the observed differential abundance in each subpopulation.

7.3.10 Supplementary Method 10: Differentially Abundant Proteins Identification Using Limma

Limma (Linear Models for Microarray Analysis) is a widely utilized statistical framework for analyzing high-throughput expression data, including microarrays, RNA sequencing (RNA-seq), and proteomics datasets [112]. It employs linear modeling to identify differentially expressed genes or proteins across experimental conditions, offering a robust approach for statistical inference. A key feature of Limma is its empirical Bayes (eBayes) moderation, which shrinks variance estimates toward a pooled estimate, enhancing statistical power—particularly in studies with limited sample sizes. The algorithm first applies a linear model to each feature (gene or protein) to estimate expression differences between conditions, followed by eBayes moderation to improve the reliability of standard error estimates. Additionally, Limma incorporates multiple testing correction methods, such as the Benjamini-Hochberg false discovery rate (FDR), to control for false positives. For proteomics data, log2 transformation of intensity values is required to approximate normality. The flexibility, computational efficiency, and statistical rigor of Limma make it a powerful tool for differential expression analysis in both transcriptomics and proteomics.

A schematic representation of the Limma analysis workflow is provided in Supplementary Figure 21.



Supplemental Figure 21: Schematic of the major components that are central to any limma analysis. For each gene g, we have a vector of gene expression values (β_g) and a design matrix X that relates these values to some coefficients of interest (g). The limma package includes statistical methods that (i) facilitate information borrowing using empirical Bayes methods to obtain posterior variance estimators (s_g2^*), (ii) incorporate observation weights (w_{gj} where j refers to sample) to allow for variations in data quality, (iii) allow variance modelling to accommodate technical or biological heterogeneity that may be present and (iv) pre-processing methods such as variance stabilization to reduce noise. These methods all help improve inference at both the gene and gene set level in small experiments. The figure is extracted from Ritchie et al. (2015) [112].

List of Figures

List of Tables

List of Supplemental Figures

List of Supplemental Tables

References

[1] Utoimmune A, Iseases D, Nne A, Avidson D, Iamond ED. Autoimmune Diseases. New England Journal of Medicine 2001;345. https://doi.org/10.1056/NEJM200108023450506.

[2] Knip M, Siljander H. Autoimmune mechanisms in type 1 diabetes. Autoimmun Rev 2008;7:550–7. https://doi.org/10.1016/J.AUTREV.2008.04.008.

[3] Tsokos GC. Autoimmunity and organ damage in systemic lupus erythematosus. Nature Immunology 2020 21:6 2020;21:605–14. https://doi.org/10.1038/s41590-020-0677-6.

[4] Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. Autoimmune responses are directed against self antigens 2001.

[5] A-Z Autoimmune Disease List | Global Autoimmune Institute n.d. https://www.autoimmuneinstitute.org/resources/autoimmune-disease-list/ (accessed October 10, 2024).

[6] Wang L, Wang F-S, Gershwin & ME, Wang A, Wang L, Me G. Human autoimmune diseases: a comprehensive update. J Intern Med 2015;278:369–95. https://doi.org/10.1111/JOIM.12395.

[7] Hato T, Dagher PC. How the innate immune system senses trouble and causes trouble. Clinical Journal of the American Society of Nephrology 2015;10:1459–69. https://doi.org/10.2215/CJN.04680514.

[8] Turvey SE, Broide DH. Innate immunity. Journal of Allergy and Clinical Immunology 2010;125:S24–32. https://doi.org/10.1016/J.JACI.2009.07.016.

[9] Litman GW, Cannon JP, Dishaw LJ. Reconstructing immune phylogeny: new perspectives. Nature Reviews Immunology 2005 5:11 2005;5:866–79. https://doi.org/10.1038/nri1712.

[10] Restifo NP, Gattinoni L. Lineage relationship of effector and memory T cells. Curr Opin Immunol 2013;25:556–63. https://doi.org/10.1016/J.COI.2013.09.003.

[11] Kurosaki T, Kometani K, Ise W. Memory B cells. Nature Reviews Immunology 2015 15:3 2015;15:149–59. https://doi.org/10.1038/nri3802.

[12] Pancer Z, Cooper MD. The evolution of adaptive immunity. Annu Rev Immunol 2006;24:497–518. https://doi.org/10.1146/ANNUREV.IMMUNOL.24.021605.090542/CITE/REFWORKS.

[13] Charles A Janeway J, Travers P, Walport M, Shlomchik MJ. Immunobiology. Immunobiology 2001:1–10.

[14] Holtmeier W, Kabelitz D. gammadelta T cells link innate and adaptive immune responses. Chem Immunol Allergy 2005;86:151–83. https://doi.org/10.1159/000086659.

[15] Venturi S, Venturi M. Iodine, thymus, and immunity. Nutrition 2009;25:977–9. https://doi.org/10.1016/J.NUT.2009.06.002.

[16] Radoja S, Frey AB, Vukmanović S. T-Cell Receptor Signaling Events Triggering Granule Exocytosis. Critical Reviews™ in Immunology 2006;26:265–90. https://doi.org/10.1615/CRITREVIMMUNOL.V26.I3.40.

[17] Cholera M, Chainani-Wu N. Management of Pemphigus Vulgaris. Adv Ther 2016;33:910–58. https://doi.org/10.1007/S12325-016-0343-4/METRICS.

[18] Mohammad Beigi PK. History of Pemphigus Vulgaris. A Clinician's Guide to Pemphigus Vulgaris 2018:13–13. https://doi.org/10.1007/978-3-319-67759-0_3.

[19] King DF, Holubar K. History of pemphigus. Clin Dermatol 1983;1:6–12. https://doi.org/10.1016/0738-081X(83)90019-6.

[20] Gupta VK, Kelbel TE, Nguyen D, Melonakos KC, Murrell DF, Xie Y, et al. A Globally Available Internet-Based Patient Survey of Pemphigus Vulgaris: Epidemiology and Disease Characteristics. Dermatol Clin 2011;29:393–404. https://doi.org/10.1016/J.DET.2011.03.016.
[21] Ramadon D, McCrudden MTC, Courtenay AJ, Donnelly RF. Enhancement strategies for transdermal drug delivery systems: current trends and applications. Drug Deliv Transl Res 2022;12:758–91. https://doi.org/10.1007/S13346-021-00909-6/FIGURES/22.
[22] Schmidt E, Kasperkiewicz M, Joly P. Pemphigus. The Lancet 2019;394:882–94. https://doi.org/10.1016/S0140-6736(19)31778-7/ASSET/75B82EF7-1792-4868-835A-11227A87991F/MAIN.ASSETS/GR5.JPG.
[23] Pathogenesis, clinical manifestations, and diagnosis of pemphigus - UpToDate n.d. https://www.uptodate.com/contents/pathogenesis-clinical-manifestations-and-diagnosis-of-pemphigus?source=search_result&search=pemphigus%20vulgaris&selectedTitle=1~36#H181679726 (accessed October 14, 2024).
[24] Delva E, Tucker DK, Kowalczyk AP. The Desmosome. Cold Spring Harb Perspect Biol 2009;1:a002543. https://doi.org/10.1101/CSHPERSPECT.A002543.
[25] Ishii K, Green KJ. Cadherin function: Breaking the barrier. Current Biology 2001;11:R569–72. https://doi.org/10.1016/S0960-9822(01)00346-3.
[26] Müller L, Hatzfeld M, Keil R. Desmosomes as Signaling Hubs in the Regulation of Cell Behavior. Front Cell Dev Biol 2021;9:745670. https://doi.org/10.3389/FCELL.2021.745670/BIBTEX.
[27] Amagai M, Karpati S, Prussick R, Klaus-Kovtun V, Stanley JR. Autoantibodies against the amino-terminal cadherin-like binding domain of pemphigus vulgaris antigen are pathogenic. J Clin Invest 1992;90:919–26. https://doi.org/10.1172/JCI115968.
[28] Kitajima Y, Inoue S, Yaoita H. Effects of Pemphigus Antibody on the Regeneration of Cell-Cell Contact in Keratinocyte Cultures grown in Low to Normal Ca+ + Concentration. Journal of Investigative Dermatology 1987;89:167–71. https://doi.org/10.1111/1523-1747.EP12470554.
[29] Kitajima Y, Aoyama Y, Seishima M. Transmembrane Signaling for Adhesive Regulation of Desmosomes and Hemidesmosomes, and for Cell-Cell Detachment Induced by Pemphigus IgG in Cultured Keratinocytes: Involvement of Protein Kinase C. Journal of Investigative Dermatology Symposium Proceedings 1999;4:137–44. https://doi.org/10.1038/SJ.JIDSP.5640197.
[30] Pelacho B, Natal C, España A, Sánchez-Carpintero I, Iraburu MJ, López-Zabalza MJ. Pemphigus vulgaris autoantibodies induce apoptosis in HaCaT keratinocytes. FEBS Lett 2004;566:6–10. https://doi.org/10.1016/J.FEBSLET.2004.03.107.
[31] Berkowitz P, Hu P, Liu Z, Diaz LA, Enghild JJ, Chua MP, et al. Desmosome Signaling: INHIBITION OF p38MAPK PREVENTS PEMPHIGUS VULGARIS IgG-INDUCED CYTOSKELETON REORGANIZATION. Journal of Biological Chemistry 2005;280:23778–84. https://doi.org/10.1074/JBC.M501365200.
[32] Sajda T, Sinha AA. Autoantibody signaling in pemphigus vulgaris: Development of an integrated model. Front Immunol 2018;9:358714. https://doi.org/10.3389/FIMMU.2018.00692/XML/NLM.
[33] Yamamoto Y, Aoyama Y, Shu E, Tsunoda K, Amagai M, Kitajima Y. Anti-desmoglein 3 (Dsg3) Monoclonal Antibodies Deplete Desmosomes of Dsg3 and Differ in Their Dsg3-depleting Activities Related to Pathogenicity. Journal of Biological Chemistry 2007;282:17866–76. https://doi.org/10.1074/JBC.M607963200.

[34] Kitajima Y. New insights into desmosome regulation and pemphigus blistering as a desmosome-remodeling disease. Kaohsiung J Med Sci 2013;29:1–13. https://doi.org/10.1016/J.KJMS.2012.08.001.
[35] Tavakolpour S, Mahmoudi HR, Mirzazadeh A, Balighi K, Darabi-Monadi S, Hatami SB, et al. Pathogenic and protective roles of cytokines in pemphigus: A systematic review. Cytokine 2020;129:155026. https://doi.org/10.1016/J.CYTO.2020.155026.
[36] Haeberle S, Wei X, Bieber K, Goletz S, Ludwig RJ, Schmidt E, et al. Regulatory T-cell deficiency leads to pathogenic bullous pemphigoid antigen 230 autoantibody and autoimmune bullous disease. Journal of Allergy and Clinical Immunology 2018;142:1831-1842.e7. https://doi.org/10.1016/J.JACI.2018.04.006/ASSET/F5D6F53F-F18A-4950-B8B9-E52E80824A71/MAIN.ASSETS/FX5.JPG.
[37] Zhu Y, Su J, Zhang P, Deng M, Wu R, Liu Y, et al. The dysregulation of circulating innate lymphoid cells is related to autoantibodies in pemphigus vulgaris. Int Immunopharmacol 2023;117:109921. https://doi.org/10.1016/J.INTIMP.2023.109921.
[38] Pollmann R, Walter E, Schmidt T, Waschke J, Hertl M, Möbs C, et al. Identification of autoreactive B cell subpopulations in peripheral blood of autoimmune patients with pemphigus vulgaris. Front Immunol 2019;10:459258. https://doi.org/10.3389/FIMMU.2019.01375/BIBTEX.
[39] Schmitt T, Waschke J. Autoantibody-Specific Signalling in Pemphigus. Front Med (Lausanne) 2021;8:701809. https://doi.org/10.3389/FMED.2021.701809/XML/NLM.
[40] FDA Approves Genentech's Rituxan (Rituximab) for Pemphigus Vulgaris n.d. https://www.gene.com/media/press-releases/14727/2018-06-07/fda-approves-genentechs-rituxan-rituxima (accessed October 16, 2024).
[41] Ahmed AR, Kaveri S, Spigelman Z. Long-Term Remissions in Recalcitrant Pemphigus Vulgaris. New England Journal of Medicine 2015;373:2693–4. https://doi.org/10.1056/NEJMC1508234/SUPPL_FILE/NEJMC1508234_DISCLOSURES.PDF.
[42] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57. https://doi.org/10.1038/NRG2484.
[43] Chu Y, Corey DR. RNA Sequencing: Platform Selection, Experimental Design, and Data Interpretation. Nucleic Acid Ther 2012;22:271. https://doi.org/10.1089/NAT.2012.0367.
[44] Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, et al. Transcriptome Sequencing to Detect Gene Fusions in Cancer. Nature 2009;458:97. https://doi.org/10.1038/NATURE07638.
[45] Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc 2012;7:1534. https://doi.org/10.1038/NPROT.2012.086.
[46] Kukurba KR, Montgomery SB. RNA Sequencing and Analysis. Cold Spring Harb Protoc 2015;2015:951. https://doi.org/10.1101/PDB.TOP084970.
[47] Alpern D, Gardeux V, Russeil J, Mangeat B, Meireles-Filho ACA, Breysse R, et al. BRB-seq: ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. Genome Biol 2019;20:71. https://doi.org/10.1186/S13059-019-1671-X.
[48] Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, et al. Highly multiplexed subcellular RNA sequencing in situ. Science 2014;343:1360. https://doi.org/10.1126/SCIENCE.1250212.

[49] Thind AS, Monga I, Thakur PK, Kumari P, Dindhoria K, Krzak M, et al. Demystifying emerging bulk RNA-Seq applications: the application and utility of bioinformatic methodology. Brief Bioinform 2021;22:1–16. https://doi.org/10.1093/BIB/BBAB259.
[50] Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. PLoS Comput Biol 2017;13:e1005457. https://doi.org/10.1371/JOURNAL.PCBI.1005457.
[51] Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology 2016 34:5 2016;34:525–7. https://doi.org/10.1038/nbt.3519.
[52] Differential gene expression n.d. https://biocorecrg.github.io/CRG_Bioinformatics_for_Biologists/differential_gene_expression.html (accessed February 13, 2025).
[53] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9.
[54] Gene-level differential expression analysis with DESeq2 | Introduction to DGE - ARCHIVED n.d. https://hbctraining.github.io/DGE_workshop_salmon/lessons/04_DGE_DESeq2_analysis.html (accessed February 13, 2025).
[55] Maleki F, Ovens K, Hogan DJ, Kusalik AJ. Gene Set Analysis: Challenges, Opportunities, and Future Research. Front Genet 2020;11:531777. https://doi.org/10.3389/FGENE.2020.00654/BIBTEX.
[56] Zyla J, Marczyk M, Weiner J, Polanska J. Ranking metrics in gene set enrichment analysis: Do they matter? BMC Bioinformatics 2017;18:1–12. https://doi.org/10.1186/S12859-017-1674-0/FIGURES/5.
[57] Ihnatova I, Popovici V, Budinska E. A critical comparison of topology-based pathway analysis methods. PLoS One 2018;13:e0191154. https://doi.org/10.1371/JOURNAL.PONE.0191154.
[58] Zhao K, Rhee SY. Interpreting omics data with pathway enrichment analysis. Trends in Genetics 2023;39:308–19. https://doi.org/10.1016/J.TIG.2023.01.003/ASSET/5DFF7C8C-DE5D-4FC3-9859-BCEC2D04D537/MAIN.ASSETS/GR2.SML.
[59] Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: Generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics 2009;10:1–17. https://doi.org/10.1186/1471-2105-10-161/FIGURES/5.
[60] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015;1:417. https://doi.org/10.1016/J.CELS.2015.12.004.
[61] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nature Genetics 2000 25:1 2000;25:25–9. https://doi.org/10.1038/75556.
[62] Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 2000;28:27–30. https://doi.org/10.1093/NAR/28.1.27.
[63] Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res 2011;39:D691–7. https://doi.org/10.1093/NAR/GKQ1018.
[64] Latchman DS. Transcription factors: An overview. Int J Biochem Cell Biol 1997;29:1305–12. https://doi.org/10.1016/S1357-2725(97)00085-X.

[65] Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell 2018;172:650–65. https://doi.org/10.1016/J.CELL.2018.01.029.
[66] Lac repressor - Wikipedia n.d. https://en.wikipedia.org/wiki/Lac_repressor (accessed February 14, 2025).
[67] Weidemüller P, Kholmatov M, Petsalaki E, Zaugg JB. Transcription factors: Bridge between cell signaling and gene regulation. Proteomics 2021;21:2000034. https://doi.org/10.1002/PMIC.202000034.
[68] Badia-I-Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. Bioinformatics Advances 2022;2. https://doi.org/10.1093/BIOADV/VBAC016.
[69] Müller-Dott S, Tsirvouli E, Vazquez M, Ramirez Flores RO, Badia-I-Mompel P, Fallegger R, et al. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. Nucleic Acids Res 2023;51:10934–49. https://doi.org/10.1093/NAR/GKAD841.
[70] Transcription factor activity inference in bulk RNA-seq • decoupleR n.d. https://saezlab.github.io/decoupleR/articles/tf_bk.html (accessed February 15, 2025).
[71] Linnarsson S, Teichmann SA. Single-cell genomics: coming of age. Genome Biol 2016;17. https://doi.org/10.1186/S13059-016-0960-X.
[72] Reinius B, Sandberg R. Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. Nat Rev Genet 2015;16:653–64. https://doi.org/10.1038/NRG3888.
[73] Reinius B, Mold JE, Ramsköld D, Deng Q, Johnsson P, Michaëlsson J, et al. Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. Nat Genet 2016;48:1430–5. https://doi.org/10.1038/NG.3678.
[74] Kolodziejczyk AA, Kim JK, Tsang JCH, Ilicic T, Henriksson J, Natarajan KN, et al. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. Cell Stem Cell 2015;17:471–85. https://doi.org/10.1016/J.STEM.2015.09.011.
[75] Martinez-Jimenez CP, Eling N, Chen HC, Vallejos CA, Kolodziejczyk AA, Connor F, et al. Aging increases cell-to-cell transcriptional variability upon immune stimulation. Science 2017;355:1433–6. https://doi.org/10.1126/SCIENCE.AAH4115.
[76] Björklund AK, Forkel M, Picelli S, Konya V, Theorell J, Friberg D, et al. The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. Nat Immunol 2016;17:451–60. https://doi.org/10.1038/NI.3368.
[77] Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science (1979) 2017;356. https://doi.org/10.1126/SCIENCE.AAH4573/SUPPL_FILE/AAH4573_VILLANI_SM.PDF.
[78] Stubbington MJT, Rozenblatt-Rosen O, Regev A, Teichmann SA. Single-cell transcriptomics to explore the immune system in health and disease. Science 2017;358:58–63. https://doi.org/10.1126/SCIENCE.AAN6828.
[79] Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 2014;343:776–9. https://doi.org/10.1126/SCIENCE.1247651.
[80] Poulin JF, Tasic B, Hjerling-Leffler J, Trimarchi JM, Awatramani R. Disentangling neural cell diversity using single-cell transcriptomics. Nat Neurosci 2016;19:1131–41. https://doi.org/10.1038/NN.4366.

[81] Zeisel A, M͡oz-Manchado AB, Codeluppi S, Lönnerberg P, Manno G La, Juréus A, et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science 2015;347:1138–42. https://doi.org/10.1126/SCIENCE.AAA1934.
[82] La Manno G, Gyllborg D, Codeluppi S, Nishimura K, Salto C, Zeisel A, et al. Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. Cell 2016;167:566-580.e19. https://doi.org/10.1016/J.CELL.2016.09.027.
[83] Gokce O, Stanley GM, Treutlein B, Neff NF, Camp JG, Malenka RC, et al. Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. Cell Rep 2016;16:1126–37. https://doi.org/10.1016/J.CELREP.2016.06.059.
[84] Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci 2016;19:335–46. https://doi.org/10.1038/NN.4216.
[85] Lein E, Borm LE, Linnarsson S. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. Science 2017;358:64–9. https://doi.org/10.1126/SCIENCE.AAN6827.
[86] Haber AL, Biton M, Rogel N, Herbst RH, Shekhar K, Smillie C, et al. A single-cell survey of the small intestinal epithelium. Nature 2017;551:333–9. https://doi.org/10.1038/NATURE24489.
[87] Grün D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature 2015;525:251–5. https://doi.org/10.1038/NATURE14966.
[88] Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, et al. A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst 2016;3:385-394.e3. https://doi.org/10.1016/J.CELS.2016.09.002.
[89] Angerer P, Haghverdi L, Büttner M, Theis FJ, Marr C, Buettner F. destiny: diffusion maps for large-scale single-cell data in R. Bioinformatics 2016;32:1241–3. https://doi.org/10.1093/BIOINFORMATICS/BTV715.
[90] Moignard V, Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, Wilkinson AC, et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat Biotechnol 2015;33:269–76. https://doi.org/10.1038/NBT.3154.
[91] Haghverdi L, Büttner M, Wolf FA, Buettner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. Nat Methods 2016;13:845–8. https://doi.org/10.1038/NMETH.3971.
[92] Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell 2016;167:1853-1866.e17. https://doi.org/10.1016/J.CELL.2016.11.038.
[93] Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell 2016;167:1867-1882.e21. https://doi.org/10.1016/J.CELL.2016.11.048.
[94] Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, et al. Pooled CRISPR screening with single-cell transcriptome readout. Nat Methods 2017;14:297–301. https://doi.org/10.1038/NMETH.4177.
[95] The neXt generation of single cell RNA-seq: An introduction to GEM-X technology - 10x Genomics n.d. https://www.10xgenomics.com/blog/the-next-generation-of-single-cell-rna-seq-an-introduction-to-gem-x-technology (accessed October 28, 2024).

[96] Yan L, Yang M, Guo H, Yang L, Wu J, Li R, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nature Structural & Molecular Biology 2013 20:9 2013;20:1131–9. https://doi.org/10.1038/nsmb.2660.
[97] Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X, et al. Tracing the Derivation of Embryonic Stem Cells from the Inner Cell Mass by Single-Cell RNA-Seq Analysis. Cell Stem Cell 2010;6:468–78. https://doi.org/10.1016/J.STEM.2010.03.015.
[98] Ramsköld D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nature Biotechnology 2012 30:8 2012;30:777–82. https://doi.org/10.1038/nbt.2282.
[99] Islam S, Kjällquist U, Moliner A, Zajac P, Fan JB, Lönnerberg P, et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res 2011;21:1160–7. https://doi.org/10.1101/GR.110882.110.
[100] Sandberg R. Entering the era of single-cell transcriptomics in biology and medicine. Nature Methods 2014 11:1 2013;11:22–4. https://doi.org/10.1038/nmeth.2764.
[101] 10XGenomics/cellranger: 10x Genomics Single Cell Analysis n.d. https://github.com/10XGenomics/cellranger (accessed October 28, 2024).
[102] Analysis, visualization, and integration of Visium HD spatial datasets with Seurat • Seurat n.d. https://satijalab.org/seurat/articles/pbmc3k_tutorial (accessed February 15, 2025).
[103] Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nature Immunology 2019 20:2 2019;20:163–72. https://doi.org/10.1038/s41590-018-0276-y.
[104] Zhao J, Jaffe A, Li H, Lindenbaum O, Sefik E, Jackson R, et al. Detection of differentially abundant cell subpopulations in scrna-seq data. Proc Natl Acad Sci U S A 2021;118:e2100293118. https://doi.org/10.1073/PNAS.2100293118/SUPPL_FILE/PNAS.2100293118.SD01.XLSX.
[105] Alves P, Arnold RJ, Novotny M V, Radivojac P, Reilly JP, Tang H. ADVANCEMENT IN PROTEIN INFERENCE FROM SHOTGUN PROTEOMICS USING PEPTIDE DETECTABILITY n.d.
[106] Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature Biotechnology 2001 19:3 2001;19:242–7. https://doi.org/10.1038/85686.
[107] Wolters DA, Washburn MP, Yates JR. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. Anal Chem 2001;73:5683–90. https://doi.org/10.1021/AC010617E.
[108] Hu L, Ye M, Jiang X, Feng S, Zou H. Advances in hyphenated analytical techniques for shotgun proteome and peptidome analysis—A review. Anal Chim Acta 2007;598:193–204. https://doi.org/10.1016/J.ACA.2007.07.046.
[109] Fournier ML, Gilmore JM, Martin-Brown SA, Washburn MP. Multidimensional separations-based shotgun proteomics. Chem Rev 2007;107:3654–86. https://doi.org/10.1021/CR068279A/ASSET/CR068279A.FP.PNG_V03.
[110] Nesvizhskii AI. Protein Identification by Tandem Mass Spectrometry and Sequence Database Searching. Methods Mol Biol 2007;367:87–119. https://doi.org/10.1385/1-59745-275-0:87.
[111] Nättinen J, Aapola U, Nukareddy P, Uusitalo H. Looking deeper into ocular surface health: an introduction to clinical tear proteomics analysis. Acta Ophthalmol 2022;100:486–98. https://doi.org/10.1111/AOS.15059.

[112] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47. https://doi.org/10.1093/NAR/GKV007.
[113] Cohen P. The origins of protein phosphorylation. Nature Cell Biology 2002 4:5 2002;4:E127–30. https://doi.org/10.1038/ncb0502-e127.
[114] Vlastaridis P, Kyriakidou P, Chaliotis A, Van de Peer Y, Oliver SG, Amoutzias GD. Estimating the total number of phosphoproteins and phosphorylation sites in eukaryotic proteomes. Gigascience 2017;6:1–11. https://doi.org/10.1093/GIGASCIENCE/GIW015.
[115] Pan QL, Xie ZM, Dai XN, Zhang Y, Shen XC, Li QQ, et al. Effects of Pemphigus Vulgaris Serum on the Expression of ATP2C1 and PKP3 in HaCaT Cells. Int J Dermatol Venereol 2021;4:224–8. https://doi.org/10.1097/JD9.0000000000000181.
[116] Xiaoyun X, Chaofei H, Weiqi Z, Chen C, Lixia L, Queping L, et al. Possible Involvement of F1F0-ATP synthase and Intracellular ATP in Keratinocyte Differentiation in normal skin and skin lesions. Scientific Reports 2017 7:1 2017;7:1–10. https://doi.org/10.1038/srep42672.
[117] Xie Z, Dai X, Li Q, Lin S, Ye X. Tacrolimus reverses pemphigus vulgaris serum-induced depletion of desmoglein in HaCaT cells via inhibition of heat shock protein 27 phosphorylation. BMC Immunol 2023;24:1–9. https://doi.org/10.1186/S12865-023-00582-Z/TABLES/1.
[118] Rheinwald JG, Hahn WC, Ramsey MR, Wu JY, Guo Z, Tsao H, et al. A Two-Stage, p16INK4A- and p53-Dependent Keratinocyte Senescence Mechanism That Limits Replicative Potential Independent of Telomere Status. Mol Cell Biol 2002;22:5157–72. https://doi.org/10.1128/MCB.22.14.5157-5172.2002.
[119] Dickson MA, Hahn WC, Ino Y, Ronfard V, Wu JY, Weinberg RA, et al. Human Keratinocytes That Express hTERT and Also Bypass a p16INK4a-Enforced Mechanism That Limits Life Span Become Immortal yet Retain Normal Growth and Differentiation Characteristics. Mol Cell Biol 2000;20:1436–47. https://doi.org/10.1128/MCB.20.4.1436-1447.2000.
[120] Smits JGA, Qu J, Niehues H, Zhou H. Characterization of In Vitro Differentiation of Human Primary Keratinocytes by RNA-Seq Analysis. J Vis Exp 2020;2020. https://doi.org/10.3791/60905.
[121] Steinert L, Fuchs M, Sigmund AM, Didona D, Hudemann C, Möbs C, et al. Desmosomal Hyper-Adhesion Affects Direct Inhibition of Desmoglein Interactions in Pemphigus. Journal of Investigative Dermatology 2024;144:2682-2694.e10. https://doi.org/10.1016/J.JID.2024.03.042.
[122] Tadesse Egu D, Schmitt T, Waschke J, Schmidt E, Kitajima Y, Memorial Hospital K, et al. Mechanisms Causing Acantholysis in Pemphigus-Lessons from Human Skin. Front Immunol 2022;13:884067. https://doi.org/10.3389/FIMMU.2022.884067.
[123] Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res 2009;38:1767-1771. https://doi.org/10.1093/nar/gkp1137.
[124] FASTA format description n.d. https://nrps.igs.umaryland.edu/fasta.html (accessed April 11, 2025).
[125] kevinblighe/PCAtools: PCAtools: everything Principal Components Analysis n.d. https://github.com/kevinblighe/PCAtools?tab=readme-ov-file (accessed October 23, 2024).

[126] Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res 2016;4:1521. https://doi.org/10.12688/F1000RESEARCH.7563.2.
[127] pheatmap: Pretty Heatmaps. CRAN: Contributed Packages 2010. https://doi.org/10.32614/CRAN.PACKAGE.PHEATMAP.
[128] kevinblighe/EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling n.d. https://github.com/kevinblighe/EnhancedVolcano (accessed October 24, 2024).
[129] Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell 2021;184:3573-3587.e29. https://doi.org/10.1016/J.CELL.2021.04.048.
[130] Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data n.d. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed April 14, 2025).
[131] Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016;32:3047–8. https://doi.org/10.1093/BIOINFORMATICS/BTW354.
[132] Long Read RNA-seq Data Standards and Processing Pipeline – ENCODE n.d. https://www.encodeproject.org/rna-seq/long-read-rna-seq/#standards (accessed April 14, 2025).
[133] Chong CE, Lim KP, Gan CP, Marsh CA, Zain RB, Abraham MT, et al. Over-expression of MAGED4B increases cell migration and growth in oral squamous cell carcinoma and is associated with poor disease outcome. Cancer Lett 2012;321:18–26. https://doi.org/10.1016/J.CANLET.2012.03.025.
[134] Fujiwara R, Takenaka S, Hashimoto M, Narawa T, Itoh T. Expression of human solute carrier family transporters in skin: possible contributor to drug-induced skin disorders. Scientific Reports 2014 4:1 2014;4:1–8. https://doi.org/10.1038/srep05251.
[135] Ye F, Zeng Q, Dan G, Dong X, Chen M, Sai Y, et al. Nitrogen mustard prevents transport of Fra-1 into the nucleus to promote c-Fos- and FosB-dependent IL-8 induction in injured mouse epidermis. Toxicol Lett 2020;319:256–63. https://doi.org/10.1016/J.TOXLET.2019.10.006.
[136] Wedel S, Martic I, Guerrero Navarro L, Ploner C, Pierer G, Jansen-Dürr P, et al. Depletion of growth differentiation factor 15 (GDF15) leads to mitochondrial dysfunction and premature senescence in human dermal fibroblasts. Aging Cell 2023;22. https://doi.org/10.1111/ACEL.13752.
[137] Ehnert S, Rinderknecht H, Liu C, Voss M, Konrad FM, Eisler W, et al. Increased Levels of BAMBI Inhibit Canonical TGF-β Signaling in Chronic Wound Tissues. Cells 2023, Vol 12, Page 2095 2023;12:2095. https://doi.org/10.3390/CELLS12162095.
[138] Dong C, Liu Y, Yu G, Li X, Chen L. Study of LBHD1 Expression with Invasion and Migration of Bladder Cancer. Open Life Sci 2019;14:440–7. https://doi.org/10.1515/BIOL-2019-0049/MACHINEREADABLECITATION/RIS.
[139] Fu W, Liang D, Wu X, Chen H, Hong X, Wang J, et al. Long noncoding RNA LINC01435 impedes diabetic wound healing by facilitating YY1-mediated HDAC8 expression. IScience 2022;25. https://doi.org/10.1016/j.isci.2022.104006.
[140] Lin MH, Khnykin D. Fatty acid transporters in skin development, function and disease. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids 2014;1841:362–8. https://doi.org/10.1016/J.BBALIP.2013.09.016.
[141] Zhen Y, Yang J, Song J, Xing Z, Zheng J. Silencing ARL11 relieved atherosclerotic inflammation and lipid deposition via retraining JAK2/STAT1 pathway.

Atherosclerosis 2024;398:118564. https://doi.org/10.1016/J.ATHEROSCLEROSIS.2024.118564.
[142] Baron JM, Höller D, Schiffer R, Frankenberg S, Neis M, Merk HF, et al. Expression of Multiple Cytochrome P450 Enzymes and Multidrug Resistance-Associated Transport Proteins in Human Skin Keratinocytes. Journal of Investigative Dermatology 2001;116:541–8. https://doi.org/10.1046/J.1523-1747.2001.01298.X.
[143] Katiyar SK, Matsui MS, Mukhtar H. Ultraviolet-B Exposure of Human Skin Induces Cytochromes P450 1A1 and 1B1. Journal of Investigative Dermatology 2000;114:328–33. https://doi.org/10.1046/J.1523-1747.2000.00876.X.
[144] Leijs MM, Esser A, Amann PM, Schettgen T, Heise R, Fietkau K, et al. Expression of CYP1A1, CYP1B1 and IL-1β in PBMCs and skin samples of PCB exposed individuals. Science of The Total Environment 2018;642:1429–38. https://doi.org/10.1016/J.SCITOTENV.2018.06.136.
[145] Kyoreva M, Li Y, Hoosenally M, Hardman-Smart J, Morrison K, Tosi I, et al. CYP1A1 Enzymatic Activity Influences Skin Inflammation Via Regulation of the AHR Pathway. Journal of Investigative Dermatology 2021;141:1553-1563.e3. https://doi.org/10.1016/J.JID.2020.11.024.
[146] Baron JM, Wiederholt T, Heise R, Merk HF, Bickers DR. Expression and Function of Cytochrome P450-Dependent Enzymes in Human Skin Cells. Curr Med Chem 2008;15:2258–64. https://doi.org/10.2174/092986708785747535.
[147] Xu Y, Johansson M, Karlsson A. Human UMP-CMP kinase 2, a novel nucleoside monophosphate kinase localized in mitochondria. J Biol Chem 2008;283:1563–71. https://doi.org/10.1074/JBC.M707997200.
[148] El-Diwany R, Soliman M, Sugawara S, Breitwieser F, Skaist A, Coggiano C, et al. CMPK2 and BCL-G are associated with type 1 interferon-induced HIV restriction in humans. Sci Adv 2018;4. https://doi.org/10.1126/SCIADV.AAT0843.
[149] Zhu M, Lv J, Wang W, Guo R, Zhong C, Antia A, et al. CMPK2 is a host restriction factor that inhibits infection of multiple coronaviruses in a cell-intrinsic manner. PLoS Biol 2023;21. https://doi.org/10.1371/JOURNAL.PBIO.3002039.
[150] Pawlak JB, Hsu JCC, Xia H, Han P, Suh HW, Grove TL, et al. CMPK2 restricts Zika virus replication by inhibiting viral translation. PLoS Pathog 2023;19. https://doi.org/10.1371/JOURNAL.PPAT.1011286.
[151] Lai JH, Wu DW, Wu CH, Hung LF, Huang CY, Ka SM, et al. Mitochondrial CMPK2 mediates immunomodulatory and antiviral activities through IFN-dependent and IFN-independent pathways. IScience 2021;24:102498. https://doi.org/10.1016/j.isci.2021.102498.
[152] McConnell BB, Yang VW. Mammalian Krüppel-Like factors in health and diseases. Physiol Rev 2010;90:1337–81. https://doi.org/10.1152/PHYSREV.00058.2009/ASSET/IMAGES/LARGE/Z9J00410256300 07.JPEG.
[153] Lili LN, Klopot A, Readhead B, Baida G, Dudley JT, Budunova I. Transcriptomic Network Interactions in Human Skin Treated with Topical Glucocorticoid Clobetasol Propionate. Journal of Investigative Dermatology 2019;139:2281–91. https://doi.org/10.1016/J.JID.2019.04.021.
[154] SMIM11 - Antibodies - The Human Protein Atlas n.d. https://www.proteinatlas.org/ENSG00000205670-SMIM11/summary/antibody (accessed November 11, 2024).
[155] Moya-Molina M, Dorgau B, Flood E, Letteboer SJF, Lorentzen E, Coxhead J, et al. Deciphering the impact of PROM1 alternative splicing on human photoreceptor

development and maturation. Cell Death & Disease 2024 15:10 2024;15:1–17. https://doi.org/10.1038/s41419-024-07105-7.
[156] Guo Y, Gao Z, LaGory EL, Kristin LW, Gupte J, Gong Y, et al. Liver-specific mitochondrial amidoxime–reducing component 1 (Mtarc1) knockdown protects the liver from diet-induced MASH in multiple mouse models. Hepatol Commun 2024;8. https://doi.org/10.1097/HC9.0000000000000419.
[157] Angonin D, van Raay TJ. Nkd1 Functions as a Passive Antagonist of Wnt Signaling. PLoS One 2013;8:e74666. https://doi.org/10.1371/JOURNAL.PONE.0074666.
[158] Castellucci LC, Almeida L, Cherlin S, Fakiola M, Francis RW, Carvalho EM, et al. A Genome-wide Association Study Identifies SERPINB10, CRLF3, STX7, LAMP3, IFNG-AS1, and KRT80 As Risk Loci Contributing to Cutaneous Leishmaniasis in Brazil. Clinical Infectious Diseases 2021;72:e515–25. https://doi.org/10.1093/CID/CIAA1230.
[159] Tagore M, Hergenreder E, Perlee SC, Cruz NM, Menocal L, Suresh S, et al. GABA Regulates Electrical Activity and Tumor Initiation in Melanoma. Cancer Discov 2023;13:2270. https://doi.org/10.1158/2159-8290.CD-23-0389.
[160] Arnoux V, Nassour M, L'Helgoualc'h A, Hipskind RA, Savagner P. Erk5 Controls Slug Expression and Keratinocyte Activation during Wound Healing. Mol Biol Cell 2008;19:4738. https://doi.org/10.1091/MBC.E07-10-1078.
[161] Li W, Nadelman C, Henry G, Fan J, Muellenhoff M, Medina E, et al. The p38-MAPK/SAPK pathway is required for human keratinocyte migration on dermal collagen. J Invest Dermatol 2001;117:1601–11. https://doi.org/10.1046/J.0022-202X.2001.01608.X.
[162] Kennedy-Crispin M, Billick E, Mitsui H, Gulati N, Fujita H, Gilleaudeau P, et al. Human keratinocytes' response to injury upregulates CCL20 and other genes linking innate and adaptive immunity. J Invest Dermatol 2012;132:105–13. https://doi.org/10.1038/JID.2011.262.
[163] Nuutila K, Siltanen A, Peura M, Bizik J, Kaartinen I, Kuokkanen H, et al. Human skin transcriptome during superficial cutaneous wound healing. Wound Repair and Regeneration 2012;20:830–9. https://doi.org/10.1111/J.1524-475X.2012.00831.X.
[164] Yu J, Qu L, Xia Y, Zhang X, Feng J, Duan M, et al. TMEM189 negatively regulates the stability of ULK1 protein and cell autophagy. Cell Death & Disease 2022 13:4 2022;13:1–10. https://doi.org/10.1038/s41419-022-04722-y.
[165] Kishibe M, Baida G, Bhalla P, Lavker RM, Schlosser B, Iinuma S, et al. Important role of kallikrein 6 for the development of keratinocyte proliferative resistance to topical glucocorticoids. Oncotarget 2016;7:69479. https://doi.org/10.18632/ONCOTARGET.9926.
[166] Lin Z, Zou J, Sui X, Yao S, Lin L, Wang J, et al. Necroptosis-related lncRNA signature predicts prognosis and immune response for cervical squamous cell carcinoma and endocervical adenocarcinomas. Scientific Reports 2022 12:1 2022;12:1–14. https://doi.org/10.1038/s41598-022-20858-5.
[167] Cheng NL, Chen X, Kim J, Shi AH, Nguyen C, Wersto R, et al. MicroRNA-125b modulates inflammatory chemokine CCL4 expression in immune cells and its reduction causes CCL4 increase with age. Aging Cell 2015;14:200. https://doi.org/10.1111/ACEL.12294.
[168] Kasela T, Dąbala M, Mistarz M, Wieczorek W, Wierzbik-Strońska M, Boroń K, et al. Effects of Cyclosporine A and Adalimumab on the expression profiles histaminergic system-associated genes and microRNAs regulating these genes in HaCaT cells. Cell Cycle 2022;21:2499–516. https://doi.org/10.1080/15384101.2022.2103342.

[169] Huang YZ, Zheng YX, Zhou Y, Xu F, Cui YZ, Chen XY, et al. OAS1, OAS2, and OAS3 Contribute to Epidermal Keratinocyte Proliferation by Regulating Cell Cycle and Augmenting IFN-1–Induced Jak1–Signal Transducer and Activator of Transcription 1 Phosphorylation in Psoriasis. Journal of Investigative Dermatology 2022;142:2635-2645.e9. https://doi.org/10.1016/J.JID.2022.02.018.
[170] McDermott JE, Vartanian KB, Mitchell H, Stevens SL, Sanfilippo A, Stenzel-Poore MP. Identification and Validation of Ifit1 as an Important Innate Immune Bottleneck. PLoS One 2012;7:e36465. https://doi.org/10.1371/JOURNAL.PONE.0036465.
[171] Raposo RA, Gupta R, Abdel-Mohsen M, Dimon M, Debbaneh M, Jiang W, et al. Antiviral gene expression in psoriasis. Journal of the European Academy of Dermatology and Venereology 2015;29:1951–7. https://doi.org/10.1111/JDV.13091.
[172] Straube R, Shah M, Flockerzi D, Wolf DA. Trade-off and flexibility in the dynamic regulation of the cullin-RING ubiquitin ligase repertoire. PLoS Comput Biol 2017;13:e1005869. https://doi.org/10.1371/JOURNAL.PCBI.1005869.
[173] Qian Y, Diaz LA, Ye J, Clarke SH. Dissecting the Anti-Desmoglein Autoreactive B Cell Repertoire in Pemphigus Vulgaris Patients. The Journal of Immunology 2007;178:5982–90. https://doi.org/10.4049/JIMMUNOL.178.9.5982.
[174] Zhao X, Jin T, Ji X, Zhang Q, Zhang X, Wu Z, et al. Wdr17 Regulates Cell Proliferation, Cell Cycle Progression and Apoptosis in Mouse Spermatocyte Cell Line. Animals 2024, Vol 14, Page 1418 2024;14:1418. https://doi.org/10.3390/ANI14101418.
[175] Lv Z, Qi L, Hu X, Mo M, Jiang H, Fan B, et al. Zic Family Member 2 (ZIC2): a Potential Diagnostic and Prognostic Biomarker for Pan-Cancer. Front Mol Biosci 2021;8:631067. https://doi.org/10.3389/FMOLB.2021.631067/FULL.
[176] Desposito D, Chollet C, Taveau C, Descamps V, Alhenc-Gelas F, Roussel R, et al. Improvement of skin wound healing in diabetic mice by kinin B2 receptor blockade. Clin Sci (Lond) 2016;130:45–56. https://doi.org/10.1042/CS20150295.
[177] Yang C, Chen SJ, Chen BW, Zhang KW, Zhang JJ, Xiao R, et al. Gene Expression Profile of the Human Colorectal Carcinoma LoVo Cells Treated With Sporamin and Thapsigargin. Front Oncol 2021;11:621462. https://doi.org/10.3389/FONC.2021.621462/BIBTEX.
[178] Shi W, Li X, Wang Z, Li C, Wang D, Li C. CCL3 Promotes Cutaneous Wound Healing Through Recruiting Macrophages in Mice. Cell Transplant 2024;33. https://doi.org/10.1177/09636897241264912/ASSET/IMAGES/LARGE/10.1177_09636897241264912-FIG6.JPEG.
[179] Chou WW, Wang YS, Chen KC, Wu JM, Liang CL, Juo SHH. Tannic acid suppresses ultraviolet B-induced inflammatory signaling and complement factor B on human retinal pigment epithelial cells. Cell Immunol 2012;273:79–84. https://doi.org/10.1016/J.CELLIMM.2011.11.003.
[180] Bayer A, Wijaya B, Möbus L, Rademacher F, Rodewald M, Tohidnezhad M, et al. Platelet-Released Growth Factors and Platelet-Rich Fibrin Induce Expression of Factors Involved in Extracellular Matrix Organization in Human Keratinocytes. International Journal of Molecular Sciences 2020, Vol 21, Page 4404 2020;21:4404. https://doi.org/10.3390/IJMS21124404.
[181] Devos M, Mogilenko DA, Fleury S, Gilbert B, Becquart C, Quemener S, et al. Keratinocyte Expression of A20/TNFAIP3 Controls Skin Inflammation Associated with Atopic Dermatitis and Psoriasis. Journal of Investigative Dermatology 2019;139:135–45. https://doi.org/10.1016/J.JID.2018.06.191.
[182] Cataisson C, Ohman R, Patel G, Pearson A, Tsien M, Jay S, et al. Inducible Cutaneous Inflammation Reveals a Protumorigenic Role for Keratinocyte CXCR2 in Skin

Carcinogenesis. Cancer Res 2009;69:319–28. https://doi.org/10.1158/0008-5472.CAN-08-2490.
[183] Furue M, Furue K, Tsuji G, Nakahara T. Interleukin-17A and Keratinocytes in Psoriasis. International Journal of Molecular Sciences 2020, Vol 21, Page 1275 2020;21:1275. https://doi.org/10.3390/IJMS21041275.
[184] Furue K, Ito T, Tsuji G, Nakahara T, Furue M. Scratch wound-induced CXCL8 upregulation is EGFR-dependent in keratinocytes. J Dermatol Sci 2020;99:209–12. https://doi.org/10.1016/j.jdermsci.2020.07.002.
[185] Wang P, Zhou Y, Yang JQ, Landeck L, Min M, Chen XB, et al. The role of Sprouty1 in the proliferation, differentiation and apoptosis of epidermal keratinocytes. Cell Prolif 2018;51:e12477. https://doi.org/10.1111/CPR.12477.
[186] Nakanishi G, Fujimoto W, Arata J. IRF-1 expression in normal human epidermal keratinocytes. Arch Dermatol Res 1997;289:415–20. https://doi.org/10.1007/S004030050214/METRICS.
[187] Sun X, Li Y. Increase of ISG15 in psoriasis lesions and its promotion of keratinocyte proliferation via the Hif-1α signalling pathway. Exp Dermatol 2023;32:1971–81. https://doi.org/10.1111/EXD.14927.
[188] Han JH, Karki R, Malireddi RKS, Mall R, Sarkar R, Sharma BR, et al. NINJ1 mediates inflammatory cell death, PANoptosis, and lethality during infection conditions and heat stress. Nature Communications 2024 15:1 2024;15:1–17. https://doi.org/10.1038/s41467-024-45466-x.
[189] Thorn CF, Whitehead AS. Differential Glucocorticoid Enhancement of the Cytokine-Driven Transcriptional Activation of the Human Acute Phase Serum Amyloid A Genes, SAA1 and SAA2. The Journal of Immunology 2002;169:399–406. https://doi.org/10.4049/JIMMUNOL.169.1.399.
[190] Zhao S, Dong X, Ni X, Li L, Lu X, Zhang K, et al. Exploration of a Novel Prognostic Risk Signature and Its Effect on the Immune Response in Nasopharyngeal Carcinoma. Front Oncol 2021;11:709931. https://doi.org/10.3389/FONC.2021.709931/BIBTEX.
[191] Chen F, Elgaher WAM, Winterhoff M, Büssow K, Waqas FH, Graner E, et al. Citraconate inhibits ACOD1 (IRG1) catalysis, reduces interferon responses and oxidative stress, and modulates inflammation and cell metabolism. Nature Metabolism 2022 4:5 2022;4:534–46. https://doi.org/10.1038/s42255-022-00577-x.
[192] Shehwana H, Ijaz S, Fatima A, Walton S, Sheikh ZI, Haider W, et al. Transcriptome Analysis of Host Inflammatory Responses to the Ectoparasitic Mite Sarcoptes scabiei var. hominis. Front Immunol 2021;12:778840. https://doi.org/10.3389/FIMMU.2021.778840/BIBTEX.
[193] Leśniak W, Bohush A, Maksymowicz M, Piwowarczyk C, Karolak NK, Jurewicz E, et al. Involvement of CacyBP/SIP in differentiation and the immune response of HaCaT keratinocytes. Immunobiology 2023;228:152385. https://doi.org/10.1016/J.IMBIO.2023.152385.
[194] Pedrosa E, Carretero-Iglesia L, Boada A, Colobran R, Faner R, Pujol-Autonell I, et al. CCL4L Polymorphisms and CCL4/CCL4L Serum Levels Are Associated with Psoriasis Severity. Journal of Investigative Dermatology 2011;131:1830–7. https://doi.org/10.1038/JID.2011.127.
[195] Xu B, Musai J, Tan YS, Hile GA, Swindell WR, Klein B, et al. A critical role for IFN-β signaling for IFN-κ induction in keratinocytes. Frontiers in Lupus 2024;2:1359714. https://doi.org/10.3389/FLUPU.2024.1359714.
[196] Wolff F, Loipetzberger A, Gruber W, Esterbauer H, Aberger F, Frischauf AM. Imiquimod directly inhibits Hedgehog signalling by stimulating adenosine

receptor/protein kinase A-mediated GLI phosphorylation. Oncogene 2013 32:50 2013;32:5574–81. https://doi.org/10.1038/onc.2013.343.
[197] Zhang Q, Hu Y, Wei P, Shi L, Shi L, Li J, et al. Identification of hub genes for adult patients with sepsis via RNA sequencing. Scientific Reports 2022 12:1 2022;12:1–14. https://doi.org/10.1038/s41598-022-09175-z.
[198] Sun Z, Mao Y, Zhang X, Lu S, Wang H, Zhang C, et al. Identification of ARHGEF38, NETO2, GOLM1, and SAPCD2 Associated With Prostate Cancer Progression by Bioinformatic Analysis and Experimental Validation. Front Cell Dev Biol 2021;9:718638. https://doi.org/10.3389/FCELL.2021.718638/BIBTEX.
[199] Zhang Y, Tu C, Zhang D, Zheng Y, Peng Z, Feng Y, et al. Wnt/β-Catenin and Wnt5a/Ca2+ Pathways Regulate Proliferation and Apoptosis of Keratinocytes in Psoriasis Lesions. Cellular Physiology and Biochemistry 2015;36:1890–902. https://doi.org/10.1159/000430158.
[200] Gao Y, Zhan W, Guo D, Lin H, Farooq MA, Jin C, et al. GPR97 depletion aggravates imiquimod-induced psoriasis pathogenesis via amplifying IL-23/IL-17 axis signal pathway. Biomedicine & Pharmacotherapy 2024;179:117431. https://doi.org/10.1016/J.BIOPHA.2024.117431.
[201] Kanda N, Shimizu T, Tada Y, Watanabe S. IL-18 enhances IFN-γ-induced production of CXCL9, CXCL10, and CXCL11 in human keratinocytes. Eur J Immunol 2007;37:338–50. https://doi.org/10.1002/EJI.200636420.
[202] Ferrari SM, Ruffilli I, Colaci M, Antonelli A, Ferri C, Fallahi P. CXCL10 in psoriasis. Adv Med Sci 2015;60:349–54. https://doi.org/10.1016/J.ADVMS.2015.07.011.
[203] Mamoor S. The C-C motif chemokine ligand 3 like 3, CCL3L3, is differentially expressed in the blood in cases of viral co-infection. 2020. https://doi.org/10.31219/OSF.IO/Y8KUD.
[204] Bünemann E, Hoff NP, Buhren BA, Wiesner U, Meller S, Bölke E, et al. Chemokine ligand-receptor interactions critically regulate cutaneous wound healing. Eur J Med Res 2018;23:1–17. https://doi.org/10.1186/S40001-017-0299-0/TABLES/3.
[205] Zhu Z, Wang W, Zha Y, Wang X, Wang L, Han J, et al. Transcriptomic and Lipidomic Profiles in Nasal Polyps of Glucocorticoid Responders and Non-Responders: Before and After Treatment. Front Pharmacol 2022;12. https://doi.org/10.3389/FPHAR.2021.814953/PDF.
[206] Mommaas M, Tada J, Ponec M. Distribution of low-density lipoprotein receptors and apolipoprotein B on normal and on reconstructed human epidermis. J Dermatol Sci 1991;2:97–105. https://doi.org/10.1016/0923-1811(91)90018-S.
[207] Månsson B, Wenglén C, Mörgelin M, Saxne T, Heinegård D. Association of Chondroadherin with Collagen Type II. Journal of Biological Chemistry 2001;276:32883–8. https://doi.org/10.1074/jbc.M101680200.
[208] Liu Z, Yan W, Liu S, Liu Z, Xu P, Fang W. Regulatory network and targeted interventions for CCDC family in tumor pathogenesis. Cancer Lett 2023;565:216225. https://doi.org/10.1016/J.CANLET.2023.216225.
[209] Zhang YJ, Sun YZ, Gao XH, Qi RQ. Integrated bioinformatic analysis of differentially expressed genes and signaling pathways in plaque psoriasis. Mol Med Rep 2019;20:225–35. https://doi.org/10.3892/MMR.2019.10241/HTML.
[210] Chen G, Chen X, Duan X, Zhang R, Bai C. Unraveling the roles of IFIT3 gene and immune-metabolic pathways in psoriasis: a bioinformatics exploration for diagnostic markers and therapeutic targets. Front Mol Biosci 2024;11:1439837. https://doi.org/10.3389/FMOLB.2024.1439837/BIBTEX.

[211] Brody SL, Pan J, Huang T, Xu J, Xu H, Koenitizer J, et al. Loss of an extensive ciliary connectome induces proteostasis and cell fate switching in a severe motile ciliopathy. BioRxiv 2024:2024.03.20.585965. https://doi.org/10.1101/2024.03.20.585965.
[212] Liu B, Cheng L, Gao H, Zhang J, Dong Y, Gao W, et al. The biology of VSIG4: Implications for the treatment of immune-mediated inflammatory diseases and cancer. Cancer Lett 2023;553:215996. https://doi.org/10.1016/J.CANLET.2022.215996.
[213] Iyer S V., Ranjan A, Elias HK, Parrales A, Sasaki H, Roy BC, et al. Genome-wide RNAi screening identifies TMIGD3 isoform1 as a suppressor of NF-κB and osteosarcoma progression. Nature Communications 2016 7:1 2016;7:1–13. https://doi.org/10.1038/ncomms13561.
[214] Giuliodori A, Beffagna G, Marchetto G, Fornetto C, Vanzi F, Toppo S, et al. Loss of cardiac Wnt/β-catenin signalling in desmoplakin-deficient AC8 zebrafish models is rescuable by genetic and pharmacological intervention. Cardiovasc Res 2018;114:1082–97. https://doi.org/10.1093/CVR/CVY057.
[215] Mellor LF, Gago-Lopez N, Bakiri L, Schmidt FN, Busse B, Rauber S, et al. Keratinocyte-derived S100A9 modulates neutrophil infiltration and affects psoriasis-like skin and joint disease. Ann Rheum Dis 2022;81:1400–8. https://doi.org/10.1136/ANNRHEUMDIS-2022-222229.
[216] Kim MJ, Im MA, Lee JS, Mun JY, Kim DH, Gu A, et al. Effect of S100A8 and S100A9 on expressions of cytokine and skin barrier protein in human keratinocytes. Mol Med Rep 2019;20:2476–83. https://doi.org/10.3892/MMR.2019.10454/HTML.
[217] Hodgkinson LM, Wang L, Duncan G, Edwards DR, Michael Wormstone I. ADAM and ADAMTS gene expression in native and wound healing human lens epithelial cells. Mol Vis 2010;16:2765.
[218] Zhang Y, Wang L, Sun X, Li F. SERPINB4 Promotes Keratinocyte Inflammation via p38MAPK Signaling Pathway. J Immunol Res 2023;2023:3397940. https://doi.org/10.1155/2023/3397940.
[219] Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze A V., Mikheenko A, et al. The complete sequence of a human genome. Science 2022;376:44–53. https://doi.org/10.1126/SCIENCE.ABJ6987.
[220] Swindell WR, Remmer HA, Sarkar MK, Xing X, Barnes DH, Wolterink L, et al. Proteogenomic analysis of psoriasis reveals discordant and concordant changes in mRNA and protein abundance. Genome Med 2015;7. https://doi.org/10.1186/S13073-015-0208-5.
[221] Rietschel ET, Kirikae T, Schade FU, Mamat U, Schmidt G, Loppnow H, et al. Bacterial endotoxin: molecular relationships of structure to activity and function. The FASEB Journal 1994;8:217–25. https://doi.org/10.1096/FASEBJ.8.2.8119492.
[222] De Biase D, Sperandeo P, Kuehn MJ, Lynch J, Contreras-Rodríguez A, Daniel Avila-Calderón E, et al. Outer Membrane Vesicles of Gram-Negative Bacteria: An Outlook on Biogenesis. Front Microbiol 2021;12:557902. https://doi.org/10.3389/FMICB.2021.557902.
[223] HÖFTE H, de GREVE H, SEURINCK J, JANSENS S, MAHILLON J, AMPE C, et al. Structural and functional analysis of a cloned delta endotoxin of Bacillus thuringiensis berliner 1715. Eur J Biochem 1986;161:273–80. https://doi.org/10.1111/J.1432-1033.1986.TB10443.X.
[224] Regarding Endotoxin – AB Biosciences n.d. https://abbiosciences.com/blogs/blogs/regarding-endotoxin (accessed February 15, 2025).
[225] O'Neill LAJ. TLRs: Professor Mechnikov, sit on your hat. Trends Immunol 2004;25:687–93. https://doi.org/10.1016/J.IT.2004.10.005.

[226] Alexis NE, Lay JC, Almond M, Bromberg PA, Patel DD, Peden DB. Acute LPS inhalation in healthy volunteers induces dendritic cell maturation in vivo. Journal of Allergy and Clinical Immunology 2005;115:345–50. https://doi.org/10.1016/j.jaci.2004.11.040.
[227] Liebers V, Raulf-Heimsoth M, Brüning T. Health effects due to endotoxin inhalation (review). Arch Toxicol 2008;82:203–10. https://doi.org/10.1007/S00204-008-0290-1/FIGURES/2.
[228] Iwanaga S, Otsuka M. Biochemical principle of Limulus test for detecting bacterial endotoxins. Proc Jpn Acad Ser B Phys Biol Sci 2007;83:110. https://doi.org/10.2183/PJAB.83.110.
[229] Ding JL, Ho B. A new era in pyrogen testing. Trends Biotechnol 2001;19:277–81. https://doi.org/10.1016/S0167-7799(01)01694-8.
[230] US6645724B1 - Assays for endotoxin - Google Patents n.d. https://patents.google.com/patent/US6645724B1/en (accessed February 15, 2025).
[231] Liu Y, Zhou J, White KP. RNA-seq differential expression studies: more sequence or more replication? Bioinformatics 2014;30:301–4. https://doi.org/10.1093/BIOINFORMATICS/BTT688.
[232] Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA 2016;22:839–51. https://doi.org/10.1261/RNA.053959.115/-/DC1.
[233] Fuchs E, Raghavan S. Getting under the skin of epidermal morphogenesis. Nat Rev Genet 2002;3:199–209. https://doi.org/10.1038/NRG758.
[234] Van Laethem A, Van Kelst S, Lippens S, Declercq W, Vandenabeele P, Janssens S, et al. Activation of p38 MAPK is required for Bax translocation to mitochondria, cytochrome c release and apoptosis induced by UVB irradiation in human keratinocytes. FASEB J 2004;18:1946–8. https://doi.org/10.1096/FJ.04-2285FJE.
[235] Hu Y, Zhang X, Zhao Z, Chen X, Zhou Z, Yang X, et al. Keratinocytes apoptosis contributes to crizotinib induced-erythroderma. Toxicol Lett 2020;319:102–10. https://doi.org/10.1016/J.TOXLET.2019.11.007.
[236] Bektas M, Jolly P, Rubenstein DS. Apoptotic Pathways in Pemphigus. Dermatol Res Pract 2010;2010:456841. https://doi.org/10.1155/2010/456841.
[237] Lippens S, Denecker G, Ovaer P, Vandenabeele P, Declercq W. Death penalty for keratinocytes: apoptosis versus cornification. Cell Death & Differentiation 2005 12:2 2005;12:1497–508. https://doi.org/10.1038/sj.cdd.4401722.
[238] Hartmann V, Guo S, Rahimi S, Radine U, Malheiros D, Salviano-Silva A, et al. In pemphigus, cell detachment, but not autoantibody binding, induces cell-wide, long-lasting transcriptomic and proteomic changes n.d. https://doi.org/10.1101/2025.02.10.637416.
[239] Burmester IAK, Flaswinkel S, Thies CS, Kasprick A, Kamaguchi M, Bumiller-Bini V, et al. Identification of novel therapeutic targets for blocking acantholysis in pemphigus. Br J Pharmacol 2020;177:5114–30. https://doi.org/10.1111/BPH.15233.
[240] Fang H, Li Q, Wang G. The role of T cells in pemphigus vulgaris and bullous pemphigoid. Autoimmun Rev 2020;19:102661. https://doi.org/10.1016/J.AUTREV.2020.102661.
[241] Egami S, Watanabe T, Fukushima-Nomura A, Nomura H, Takahashi H, Yamagami J, et al. Desmoglein-Specific B-Cell−Targeted Single-Cell Analysis Revealing Unique Gene Regulation in Patients with Pemphigus. Journal of Investigative Dermatology 2023;143:1919-1928.e16. https://doi.org/10.1016/J.JID.2023.03.1661.

[242] Fetter T, Niebel D, Braegelmann C, Wenzel J. Skin-Associated B Cells in the Pathogenesis of Cutaneous Autoimmune Diseases—Implications for Therapeutic Approaches. Cells 2020, Vol 9, Page 2627 2020;9:2627. https://doi.org/10.3390/CELLS9122627.
[243] Barone F, Nayara S, Camposa J, Cloakea T, Withers DR, Toellner KM, et al. IL-22 regulates lymphoid chemokine production and assembly of tertiary lymphoid organs. Proc Natl Acad Sci U S A 2015;112:11024–9. https://doi.org/10.1073/PNAS.1503315112/-/DCSUPPLEMENTAL/PNAS.201503315SI.PDF.
[244] Cañete JD, Celis R, Yeremenko N, Sanmartí R, van Duivenvoorde L, Ramírez J, et al. Ectopic lymphoid neogenesis is strongly associated with activation of the IL-23 pathway in rheumatoid synovitis. Arthritis Res Ther 2015;17:1–12. https://doi.org/10.1186/S13075-015-0688-0/FIGURES/7.
[245] Strandmoe AL, Bremer J, Diercks GFH, Gostyński A, Ammatuna E, Pas HH, et al. Beyond the skin: B cells in pemphigus vulgaris, tolerance and treatment. British Journal of Dermatology 2024;191:164–76. https://doi.org/10.1093/BJD/LJAE107.
[246] Iqbal Z, Caccamo M, Turner I, Flicek P, McVean G. De novo assembly and genotyping of variants using colored de Bruijn graphs. Nature Genetics 2012 44:2 2012;44:226–32. https://doi.org/10.1038/ng.1028.
[247] Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011;12:1–16. https://doi.org/10.1186/1471-2105-12-323/TABLES/6.
[248] Cell Ranger's Gene Expression Algorithm - Official 10x Genomics Support n.d. https://www.10xgenomics.com/support/software/cell-ranger/latest/algorithms-overview/cr-gex-algorithm (accessed February 15, 2025).

Acknowledgements

I am profoundly grateful to Professor Hauke Busch for giving me the opportunity to study and grow in omics data analysis, particularly in the field of autoimmune diseases. I extend my heartfelt thanks to Dr. Axel Künstner for his invaluable mentorship in data analysis. Their immense patience in sharing scripts, assisting with debugging, and meticulously reviewing and refining my presentation slides, page by page, has been truly invaluable.

I deeply thank Dr. Anke Fähnrich, Fabian Ott, and Yamil Maluje for their guidance in single-cell RNA sequencing (scRNA-seq) data analysis. I also appreciate the warm support from colleagues Niklas Reimer and Verónica Calonga Solís when I joined the lab, and I am particularly grateful to Rochi Saurabh for helping me with various PhD-related activities.

I would like to express my gratitude to my lab members—Michael Olbrich, Tim Niklas Lenfers, Inken Wohlers, Mohammad Khodaygani, Vanessa Klauenberg, Tjorben Nawroth, Marius Möller, Sven Marcel Stefan, Katja Stefan, Svenja Denker, Vivian Wernau, Melina Schoenknecht, Saad Abdelwakeel —for their valuable feedback during group meetings and research discussions with great patience and kindness. Special thanks go to our lab secretary, Stephanie Latif, for her unwavering support and administrative guidance.

I am equally grateful to Professor Jennifer Hundt, Professor Ralf Ludwig, Veronika Hartmann, Kauderer Claudia, Professor Eliane J. Müller, Dr. William Harition, Siavash Rahimi, Patrizia Sauta, Taravat Shojaeian, Professor Tomaso Zambelli, Elaheh Zare Eelanjegh, Mengjia Xu, Professor Christoph Hammers, and other experts collaborating on the Pemphigus Vulgaris projects funded by the Swiss National Science Foundation. Their expertise and efforts have greatly enriched my learning and research experience.

I sincerely thank Professor Jennifer Hundt, Sarah Stenger, Zuzana Penxová, Steffen Pichlo, and Professor Tanja Lange for their invaluable collaboration on the lupus research project. Their detailed guidance and insightful discussions greatly contributed to the expansion of the research scope and the advancement of knowledge in the field of autoimmune diseases.

I am sincerely grateful to Dr. Laura Kirchhoff and Dr. Skadi Lange, our research coordinators, for their tremendous support throughout the study.

I am deeply grateful to the Omics Cluster team, including Frank Rühlemann, Gudrun Amedick, Kristin Grothe, and the committee members, for their invaluable support in providing computing resources for sequencing data analysis.

I would like to express my sincere gratitude to the administrative and HR offices of UKSH, the University of Lübeck, as well as the IT, graduate office, and CBBM facilities for their invaluable support.

Finally, I would like to express my gratitude to all the students and senior researchers, both on and off campus, who have supported me throughout my PhD journey. Although I cannot list all their names here, their guidance and encouragement have been essential to my personal and academic growth and success.