

# **DISSERTATION**

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the

Ruperto-Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

**M. Sc. Llorenç Solé Boldo**

born in Tortosa (Tarragona), Spain

Oral examination: 15th September 2021



# **Epigenomic characterization of human keratinocyte cancers and their cellular origin**

Referees:

Prof. Dr. Frank Lyko

Prof. Dr. Ana Martin-Villalba



# Zusammenfassung

Keratinozytenkarzinome (KC) sind, mit Millionen von diagnostizierten Fällen pro Jahr, die häufigste bösartige Erkrankung in hellhäutigen Bevölkerungsgruppen. Obwohl sie in der Vergangenheit von den Krebsregistern übersehen wurden, haben ihre kontinuierlich steigende Inzidenz und die wirtschaftliche Belastung, die sie für die Gesundheitssysteme darstellen, die KC ins Rampenlicht gerückt. KC entstehen aus epidermalen Keratinozyten und umfassen zwei Haupttumortypen: Basalzellkarzinome (BCC) und kutane Plattenepithelkarzinome (cSCC). Letztere entstehen meist aus präkanzerösen Dysplasien, die als aktinische Keratose (AK) bezeichnet werden oder aus *in situ* Karzinomen, bekannt als Morbus Bowen (BD). Auf Grundlage von DNA-Methylierungsprofilen wurden zwei Unterklassen von AK/cSCC identifiziert, die auf zwei unterschiedlichen Ursprungszellen zurückzuführen sind.

Diese Beobachtung wurde in dieser Arbeit durch die epigenomische Charakterisierung von 102 epidermalen Proben mit Infinium MethylationEPIC BeadChips weiter untersucht. Die Kohorte umfasste gesunde Kontrollen, AK, cSCC, BD, BCC und nicht-kanzeröse senile Warzen (seborrhoische Keratose; SK) und deckt damit das gesamte Spektrum der Malignität epidermaler Tumore ab. Methylierungsmuster an Keratinozyten-spezifischen Enhancern stratifizierten die Proben in zwei Subklassen, solche mit epidermalen stammzellähnlichen oder keratinozytenähnlichen Profilen. Während SK- und BCC-Proben überwiegend Keratinozyten-ähnliche Profile aufwiesen, zeigten cSCC und seine Vorläufer beide Profile, was auf mögliche klinische Implikationen hinweist. Weitere Analysen wiesen auf unterschiedliche Zellteilungsraten und invasive Merkmale zwischen den Zellursprungs-Unterklassen hin. Schließlich konnte durch -Methylomik und Transkriptomik die auf den Ursprungszellen basierende Stratifizierung der epidermalen Tumore validiert werden.

Dermale Fibroblasten sind eine heterogene Zellpopulation, die eine wesentliche Rolle bei der Aufrechterhaltung der Hautarchitektur und -funktion spielt. Ihre Vielfalt in der menschlichen Haut und ihre funktionelle Bedeutung sind jedoch noch relativ unklar. Im Rahmen dieser Arbeit wurde die Fibroblasten-Heterogenität systematisch anhand der Einzelzell-Transkriptomik von mehr als 5.000 Fibroblasten aus sonnengeschützter gesunder menschlicher Haut analysiert. Es wurden vier verschiedene Subpopulationen mit spezifischer Lokalisation identifiziert, die unterschiedliche sekretorische, mesenchymale und pro-inflammatorische Rollen ausüben. Hervorzuheben ist, dass dieses Priming bei intrinsischer Alterung deutlich reduziert wurde.

Insgesamt gewährt diese Arbeit wichtige Einblicke in den zellulären Ursprung von KC, eröffnet neue Möglichkeiten für eine robuste Risikobewertung von Patienten und stellt eine detaillierte Analyse der Heterogenität menschlicher dermaler Fibroblasten und ihrer intrinsischen Alterung dar.



# Abstract

Keratinocyte cancers (KC) are the most common malignancy in fair-skinned populations, with millions of cases diagnosed yearly. Although historically overlooked by cancer registries, their continuously rising incidence and the economic burden they pose on health systems have put KC in the spotlight. KC arise from epidermal keratinocytes and comprise two main tumor types: basal cell carcinoma (BCC) and cutaneous squamous cell carcinoma (cSCC). The latter mostly originates from pre-cancerous dysplasias named actinic keratosis (AK) or *in situ* carcinomas known as Bowen's disease (BD). Two subclasses of AK/cSCC were previously identified based on their DNA methylation profiles, reflecting two distinct cells-of-origin.

This notion was further explored in this thesis by epigenomically characterizing 102 epidermal samples using Infinium MethylationEPIC BeadChips. This cohort included healthy controls, AK, cSCC, BD, BCC, and non-cancerous senile warts (seborrheic keratosis; SK), thus comprising the full range of malignancy in epidermal tumors. Methylation patterns at keratinocyte-specific enhancers stratified the samples into two subclasses with epidermal stem cell-like or keratinocyte-like profiles. While SK and BCC samples predominantly displayed keratinocyte-like profiles, cSCC and its precursors displayed both, indicating potential clinical implications. Further analyses indicated distinct cell division rates and invasive features between cell-of-origin subclasses. Lastly, single-cell methylomics and transcriptomics validated the cell-of-origin-based stratification of epidermal tumors.

Dermal fibroblasts are a heterogeneous cell population with essential roles in maintaining skin architecture and function. However, their diversity in human skin and its functional significance are still relatively unclear. In this work, fibroblast heterogeneity was systematically analyzed using the single-cell transcriptomes of more than 5,000 fibroblasts from sun-protected healthy human skin. Four distinct subpopulations with specific localization and primed to exert differential secretory, mesenchymal, and pro-inflammatory roles were identified. Importantly, this priming was substantially reduced upon intrinsic aging.

Collectively, this work provides important insight into the cellular origin of KC, establishes new opportunities for robust patient risk assessment, and represents a detailed analysis of human dermal fibroblast heterogeneity and their intrinsic aging.



# Table of Contents

<b>List of Abbreviations</b> .....	V
<b>List of Figures</b> .....	VII
<b>List of Tables</b> .....	IX
<b>1. Introduction</b> .....	1
1.1. Human skin architecture and cellular diversity .....	1
1.1.1. Epidermal structure and differentiation.....	1
1.1.2. Dermal fibroblast diversity.....	3
1.1.3. Human skin at single-cell resolution.....	4
1.1.4. Effects of aging on the human skin .....	6
1.2. Keratinocyte cancers: a rising public health issue. ....	8
1.2.1. Epidemiology, risk factors and pathology of KC. ....	8
1.2.2. Defining the cellular origin of KC.....	11
1.3. Role of DNA methylation in keratinocyte cancers.....	13
1.3.1. The mammalian DNA (de)methylation machinery .....	14
1.3.2. DNA methylation aberrations in KC.....	15
1.3.3. Cell-of-origin-related DNA methylation patterns in cancer.....	17
<b>2. Aims</b> .....	21
<b>3. Results</b> .....	23
3.1. Methylation-based stratification of keratinocyte cancers .....	23
3.1.1. Epidermal tumors display aberrant methylomes.....	23
3.1.2. Keratinocyte enhancers stratify KC based on their cell-of-origin .....	25
3.1.3. Validation of the cell-of-origin-based stratification of epidermal tumors in independent datasets.....	27
3.1.4. KC subclasses display distinct division rates and invasive features .....	30
3.1.5. EpSC-like cSCC tumors display a higher metastatic potential.....	32
3.1.6. Single-cell transcriptomes specify KC cell-of-origin subclasses .....	36
3.2. Single-cell whole-genome bisulfite sequencing to identify the cells-of-origin of KC. ....	41
3.2.1. Establishment of a working sci-MET protocol.....	42
3.2.2. sci-MET protocol optimization for primary human epidermis .....	47
3.2.3. Single-cell methylomics identifies the KC cells-of-origin.....	49
3.3. Single-cell transcriptomes of the aging human skin reveal loss of fibroblast priming ...	51
3.3.1. Identification of the main cell populations present in the human skin .....	51

3.3.2.	Functional and spatial annotation of fibroblast subpopulations.....	55
3.3.3.	Validation of fibroblast subpopulations in skin sections.....	61
3.3.4.	Aging leads to loss of dermal fibroblast functional priming.....	65
3.3.5.	Aged fibroblast transcriptomes recapitulate known old skin features.....	67
3.3.6.	Fibroblast subpopulations display distinct skin aging-associated secreted proteins (SAASP) expression patterns.....	68
3.3.7.	Aging leads to a loss of cell-cell interactions in dermal fibroblasts.....	70
<b>4.</b>	<b>Discussion.....</b>	<b>73</b>
4.1.	Cell-of-origin-based stratification can be expanded to other epidermal entities and has potential clinical implications.....	73
4.2.	Single-cell genomics defined the cellular origin of KC.....	76
4.3.	Single-cell (epi)genomics contribution to precision oncology.....	78
4.4.	Human dermal fibroblast display functional priming.....	80
4.5.	Intrinsic aging leads to loss of fibroblast identity and cellular communication.....	83
4.6.	A framework for the study of fibroblast roles in skin disease.....	85
<b>5.</b>	<b>Conclusion.....</b>	<b>87</b>
<b>6.</b>	<b>Materials and Methods.....</b>	<b>89</b>
6.1.	Materials.....	89
6.1.1.	Chemicals, reagents, and enzymes.....	89
6.1.2.	Commercial Kits.....	90
6.1.3.	Consumables.....	91
6.1.4.	Equipment and devices.....	92
6.1.5.	Buffers and solutions.....	93
6.1.6.	Antibodies and mRNA probes.....	94
6.1.7.	Cell lines.....	94
6.1.8.	Cell Culture reagents.....	94
6.1.9.	Primers and oligos.....	95
6.1.10.	Software and computational tools.....	95
6.2.	Methods.....	96
6.2.1.	Clinical Samples.....	96
6.2.2.	Cell Culture.....	98
6.2.3.	DNA methylation profiling and data analysis.....	99
6.2.4.	Keratinocyte enhancer methylation analysis.....	99
6.2.5.	Targeted amplicon bisulfite sequencing.....	100
6.2.6.	scRNA-seq of human epidermis.....	101

6.2.7.	Bulk DNA methylation data deconvolution .....	104
6.2.8.	Single-cell combinatorial indexing for methylation analysis .....	104
6.2.9.	scRNA-seq of sun-protected healthy whole human skin .....	109
6.2.10.	mRNA fluorescence <i>in situ</i> hybridization (mRNA-FISH) .....	110
6.2.11.	Immunofluorescence assays (IF) .....	111
6.2.12.	Statistical analyses .....	112
<b>7.</b>	<b>Appendix</b> .....	<b>113</b>
7.1.	Supplementary Figures .....	113
7.2.	Supplementary Tables .....	121
7.3.	Additional results .....	122
7.3.1.	sci-MET protocol optimization for human epidermis isolated from frozen material. . .....	122
7.4.	Supplementary Methods .....	126
7.4.1.	Nuclear isolation protocols for sci-MET (fresh human epidermis).....	126
7.4.2.	Nuclear isolation protocols for sci-MET (FF human epidermis) .....	127
7.5.	List of publications .....	129
<b>8.</b>	<b>References</b> .....	<b>130</b>
<b>9.</b>	<b>Acknowledgments</b> .....	<b>149</b>



# List of abbreviations

<b>5caC</b>	5-carboxylcytosine
<b>5fC</b>	5-formylcytosine
<b>5hmC</b>	5-hydroxymethylcytosine
<b>5mC</b>	5-methylcytosine
<b>AK</b>	Actinic keratosis
<b>BCC</b>	Basal cell carcinoma
<b>BD</b>	Bowen's disease
<b>BSA</b>	Bovine serum albumin
<b>CCA</b>	Canonical Correlation Analysis
<b>cDNA</b>	coding DNA
<b>CPDs</b>	Cyclobutane pyrimidine dimers
<b>cSCC</b>	Cutaneous squamous cell carcinoma
<b>CTCL</b>	Cutaneous T-cell lymphoma
<b>DC</b>	Dendritic cells
<b>ddH<sub>2</sub>O</b>	Double-distilled Water
<b>DEJ</b>	Dermal-epidermal junction
<b>DMBA</b>	7,12-dimethyl-1,2-benzanthracene
<b>DMEM</b>	Dulbecco's Modified Eagle Medium
<b>DMF</b>	N,N-Dimethylformamide
<b>DMP</b>	Differentially methylated probes
<b>DNA</b>	Deoxyribonucleic Acid
<b>DNMT</b>	DNA methyltransferase
<b>dNTP</b>	deoxy Nucleotidetriphosphate Triphosphate
<b>DP</b>	Dermal papilla
<b>DS</b>	Dermal sheath
<b>DTT</b>	DL-Dithiothriol
<b>ECM</b>	Extracellular Matrix
<b>EDTA</b>	Ethylenediaminetetraacetic Acid
<b>EMT</b>	Epithelial-to-mesenchymal transition
<b>EpSC</b>	Epidermal stem cell
<b>ESCs/mESC</b>	Embryonic stem cells/mouse Embryonic stem cells
<b>FACS/FANS</b>	Fluorescence-Activated Cell/Nuclei sorting
<b>FBS</b>	Fetal Bovine Serum
<b>FF</b>	Fresh-frozen
<b>FFPE</b>	Formalin-fixed paraffin-embedded
<b>FISH</b>	Fluorescence in situ hybridization
<b>GEMs</b>	Gel Bead-in-emulsions
<b>GO</b>	Gene Ontology
<b>H&amp;E</b>	Hematoxylin and eosin
<b>HEPES</b>	Hydroxyethyl-Piperazineethane-Sulfonic Acid
<b>HF</b>	Hair follicle

<b>HPV</b>	Human papillomavirus
<b>IF</b>	Immunofluorescence
<b>IFE</b>	Interfollicular epidermis
<b>KC</b>	Keratinocyte cancers
<b>LIF</b>	Leukaemia inhibitory factor
<b>LNA</b>	Locked nucleic acid
<b>MCC</b>	Merkel cell carcinoma
<b>MDS</b>	Multidimensional scaling
<b>mRNA</b>	Messenger RNA
<b>ncRNA</b>	non-coding RNA
<b>NER</b>	Nucleotide excision repair
<b>NGS</b>	Next Generation Sequencing
<b>nGS</b>	Normal Goat Serum
<b>NHEK</b>	Normal human epidermal keratinocytes
<b>NMSC</b>	Non-melanoma skin cancer
<b>PBS</b>	Phosphate-buffered Saline
<b>PCA</b>	Principal Component analysis
<b>PCR</b>	Polymerase Chain Reaction
<b>PEG</b>	Polyethylene glycol
<b>PFA</b>	Paraformaldehyde
<b>RNA</b>	Ribonucleic Acid
<b>ROS</b>	Reactive oxygen species
<b>SAASP</b>	Skin aging-associated secreted proteins
<b>SASP</b>	Senescence-associated secretory phenotype
<b>SCDR</b>	Stem cell division rate
<b>sci-MET</b>	Single-cell combinatorial indexing for methylation analysis
<b>scRNA-seq</b>	Single-cell RNA sequencing
<b>scWGBS</b>	Single-cell whole-genome bisulfite sequencing
<b>SDS</b>	Sodium dodecyl sulfate
<b>SK</b>	Seborrheic keratosis
<b>SOTR</b>	Solid organ transplant recipient
<b>SPRI</b>	Solid-phase reversible immobilization
<b>TAE</b>	Tris-acetate-EDTA
<b>TCGA</b>	The Cancer Genome Atlas
<b>TET</b>	Ten-eleven translocation
<b>TF</b>	Transcription factor
<b>TME</b>	Tumor microenvironment
<b>TPA</b>	12-O-tetradecanoylphorbol-13-acetate
<b>UMAP</b>	Uniform Manifold Approximation and Projection
<b>UMI</b>	Unique Molecular Identifier
<b>UVR</b>	Ultraviolet radiation
<b>WGBS</b>	whole-genome bisulfite sequencing
<b>y/o</b>	years old

# List of Figures

Figure 1. Schematic model of the human skin and epidermis architecture .....	2
Figure 2. Droplet-based single-cell transcriptomics workflow .....	5
Figure 3. Global epidemiological data regarding KC .....	9
Figure 4. DNA methylation alterations with cancer-promoting properties.....	16
Figure 5. DNA methylation patterns in cancer cells reflect their cell-of-origin .....	18
Figure 6. DNA methylation profiling identifies two cell-of-origin subclasses of AK and cSCC ....	20
Figure 7. Epidermal tumors display archetypical cancer-specific methylation features.....	24
Figure 8. EpSC-specific enhancer methylation patterns stratify KC in two cell-of-origin subclasses. ....	26
Figure 9. Cell-of-origin-based stratification of independent KC methylation datasets .....	28
Figure 10. Combined analysis of BCC and sBCC samples .....	29
Figure 11. KC subclasses show distinct proliferative and invasive features.....	31
Figure 12. Mitotic age and EMT-associated methylation in validation datasets .....	32
Figure 13. Amplicon-sequencing approach for cSCC stratification. ....	34
Figure 14. Stratification of primary metastasizing cSCC and cSCC metastases.....	35
Figure 15. Single-cell RNA sequencing analysis of sun-protected healthy human epidermis ....	38
Figure 16. Keratinocyte differentiation trajectory inferred by RNA velocity.....	39
Figure 17. Deconvolution of bulk DNA methylation from epidermal tumors based on scRNA-seq data.....	40
Figure 18. sci-MET library preparation protocol optimization using mESCs.....	43
Figure 19. sci-MET recapitulates methylation patterns obtained in bulk experiments. ....	46
Figure 20. Optimization of human epidermis sample preparation for sci-MET .....	48
Figure 21. sci-MET analysis from human epidermis validates the cell-of-origin-based stratification of KC.....	50
Figure 22. Single-cell transcriptomics of sun-protected human skin identifies nine main cell types. ....	53
Figure 23. Individual analyses display similar cell clustering as the integrated analysis. ....	54
Figure 24. Minor skin cell populations cluster with main cell types. ....	55
Figure 25. Dermal fibroblast subpopulations display distinct functional transcriptomic signatures .....	56
Figure 26. Dermal fibroblast subpopulations display distinct spatial signatures .....	57

Figure 27. Second-level analyses reveal further human dermal fibroblast heterogeneity .....	60
Figure 28. Dermal papilla fibroblasts are contained within the mesenchymal fibroblast subpopulation.....	60
Figure 29. Fibroblast subpopulations validation in young skin using mRNA-FISH .....	63
Figure 30. Fibroblast subpopulations validation using immunofluorescence assays.....	65
Figure 31. Aging leads to loss of dermal fibroblast functional priming .....	66
Figure 32. Old fibroblasts display a global increase in immune response and decreased oxidative metabolism.....	68
Figure 33. SAASP-related gene expression in young and old dermal fibroblast subpopulations	69
Figure 34. Age-dependent loss of cell-cell interactions in dermal fibroblasts .....	70
Figure 35. Single-cell genomics in translational cancer research .....	79
Figure M1. PCR products for targeted amplicon bisulfite PCR .....	101
Figure M2. Assessment of homemade Tn5 transposase activity.....	106
Figure S1. Epidermal DNA methylation datasets display high tumor purity .....	113
Figure S2. Comparison of EPIC DNA methylation profiles in FF-FFPE matched samples .....	114
Figure S3. Amplicon-based targeted sequencing strategy for KC stratification.....	115
Figure S4. Representative differentially methylated regions between EpSC and differentiated keratinocytes in the sci-MET dataset.....	116
Figure S5. Fibroblast subpopulations validation in old skin using mRNA-FISH.....	118
Figure S6. Fibroblast interactome analyses in individual skin samples.....	119
Figure S7. Comparison of the four fibroblast subpopulations defined in this thesis with previously published datasets .....	120
Figure S8. sci-MET sample preparation protocol optimization for FF human epidermis .....	123
Figure S9. sci-MET libraries from FF epidermis samples display high rates of PCR duplication .....	125

# List of Tables

Table 1. Cell-of-origin-based stratification of human epidermal tumors.....	25
Table 2. Summary of the sequencing parameters for sci-MET libraries.....	45
Table 3. Most representative gene markers of human dermal fibroblast subpopulations.....	61
Table M1. Chemicals, reagents, and enzymes.....	89
Table M2. Commercial Kits .....	90
Table M3. Consumables .....	91
Table M4. Equipment and devices .....	92
Table M5. Buffers and solutions.....	93
Table M6. Antibodies and mRNA probes .....	94
Table M7. Cell lines .....	94
Table M8. Cell Culture reagents.....	94
Table M9. Primers and oligos.....	95
Table M10. Software and computational tools.....	95
Table S1. Overview of the human epidermal analyzed with sci-MET .....	121
Table S2. Overview of the sun-protected healthy skin samples used in the scRNA-seq analysis .....	121
Table S3. Nuclear isolation buffers from 10X Genomics demonstrated protocol .....	127
Table S4. Nuclear isolation buffers from Krishnaswami et al., 2016 protocol.....	128



# 1. Introduction

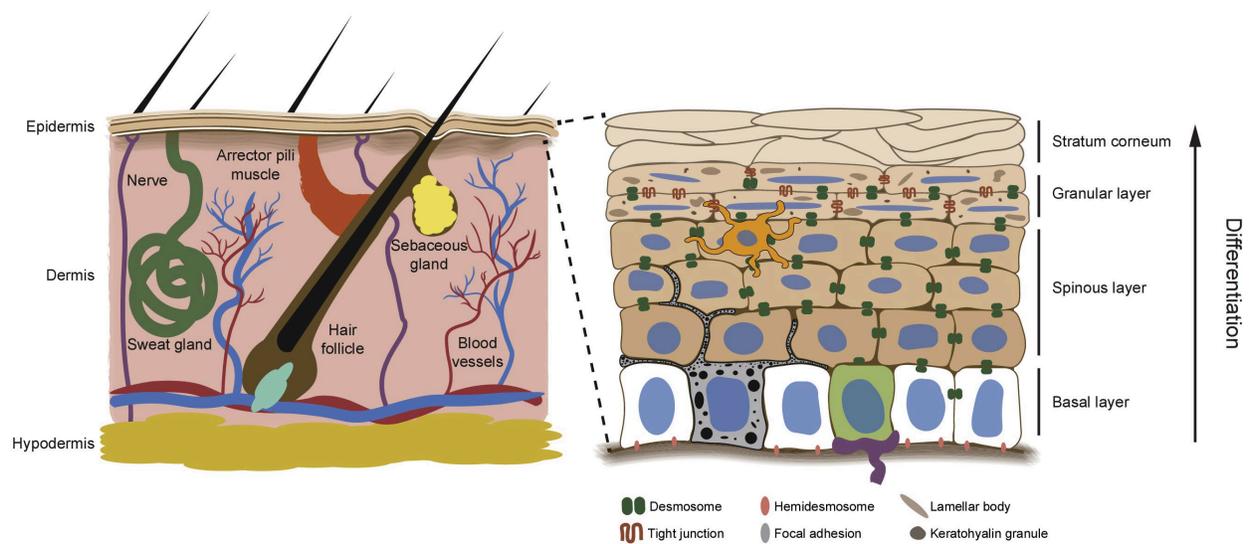
## 1.1. Human skin architecture and cellular diversity

The skin is the largest organ in the human body, covering most of its external surface. Due to its location, the skin constitutes the body's first line of defense, thus providing protection against several external aggressions such as pathogens, chemicals, ultraviolet radiation, or mechanical harm (Kolarsick, Kolarsick and Goodwin, 2011). The skin also plays essential physiological roles by preventing excessive water loss, regulating body temperature, and sensing stimuli from the environment (Kolarsick, Kolarsick and Goodwin, 2011; Wikramanayake, Stojadinovic and Tomic-Canic, 2014). Furthermore, this organ also exerts important endocrine and exocrine functions (Zouboulis, 2009). In agreement with this functional diversity, the skin is a complex organ that harbors several cell types and comprises three distinct layers: the epidermis, the dermis, and the hypodermis (Figure 1).

### 1.1.1. Epidermal structure and differentiation

The epidermis is the outermost layer of the skin and consists of a stratified squamous epithelium and several appendages, including hair follicles (HF), sebaceous glands, and sweat glands (Figure 1) (Gonzales and Fuchs, 2017; Moreci and Lechler, 2020). The interfollicular epidermis (IFE), located between hair follicles, mainly comprises keratinocytes, which arise from the epidermal stem cells (EpSCs) located at the basal layer of the epithelium (Figure 1) (Gonzales and Fuchs, 2017; Moreci and Lechler, 2020). EpSCs and other progenitors in this layer are proliferative until they start differentiating. Then, keratinocytes switch to a post-mitotic phenotype and detach from the basement membrane that separates the epidermis from the dermis, a process known as delamination (Miroshnikova *et al.*, 2018). As differentiation progresses, keratinocytes migrate upwards, forming three distinct layers (spinous, granular, and cornified) with concomitant changes in gene expression, cellular shape, and cell-cell adhesion patterns (Figure 1) (Gonzales and Fuchs, 2017; Moreci and Lechler, 2020). Thus, suprabasal spinous keratinocytes gradually acquire a more spindle shape as they increase the expression of strong intercellular junctions upon migration. Spinous cells transition to granular keratinocytes, which eventually present a complete flattened shape, enrichment of tight junctions, and display visible keratohyalin granules and lamellar bodies in their cytoplasm (Gonzales and Fuchs, 2017; Moreci

and Lechler, 2020). Granular keratinocytes represent the last layer of living cells, becoming enucleated corneocytes upon cellular death (Figure 1). At this stage, the keratohyalin granules and lamellar bodies secrete their contents, promoting the tight cross-linking of keratin filaments and forming a lipid layer (Wikramanayake, Stojadinovic and Tomic-Canic, 2014; Gonzales and Fuchs, 2017; Moreci and Lechler, 2020). This provides a strong barrier to protect the body from mechanical, chemical, and biological agents, known as stratum corneum or cornified envelope (Wikramanayake, Stojadinovic and Tomic-Canic, 2014). Corneocytes are constantly desquamated, therefore forcing the EpSCs to maintain a tight balance between self-renewal and differentiation to ensure proper and continuous replenishment of the epidermis (Blanpain and Fuchs, 2009).



**Figure 1. Schematic model of the human skin and epidermis architecture.** Left: scheme of the human skin displaying the epidermis, dermis, and hypodermis layers. The different skin appendages and structures present in the human skin are also depicted. Right: scheme showing human epidermal differentiation. EpSC and progenitor keratinocytes from the basal layer are colored in white, while differentiated keratinocytes from the spinous, granular, and cornified layers are depicted in shades of brown. Minor epidermal cell populations such as melanocytes (grey), Merkel cells (green; next to a nerve ending in purple), and Langerhans cells (orange) are also shown.

*Adapted from Moreci and Lechler (2020).*

The human epidermis is also home to several other non-keratinocyte minor cell populations such as melanocytes, Merkel cells, or Langerhans cells (Kolarsick, Kolarsick and Goodwin, 2011). Melanocytes are dendritic cells located at the basal layer of the epidermis and responsible for skin pigmentation by producing melanin in specialized organelles named melanosomes (Cichorek *et al.*, 2013). Melanocytes transfer these melanosomes to keratinocytes through their long extensions, thus providing them with photo-protective melanin (Cichorek *et al.*, 2013). Also

situated in the basal layer, Merkel cells act as mechanoreceptors by activating skin sensory terminals through relatively unknown mechanisms (Maksimovic *et al.*, 2014). Lastly, epidermis-resident, antigen-presenting Langerhans cells present a suprabasal location and play essential roles in sensing environmental pathogens and triggering innate and adaptive immune responses (Clayton *et al.*, 2017).

### **1.1.2. Dermal fibroblast diversity**

The dermis, right beneath the epidermis, is a thick layer of connective tissue, mainly composed of collagen bundles, elastic fibers, and other extracellular matrix (ECM) components secreted by fibroblasts. It plays essential roles in skin architecture by providing elasticity and resistance as well as protection against external mechanical harm (Kolarsick, Kolarsick and Goodwin, 2011). The dermis also harbors several structures such as vasculature, nerves, mechanoreceptors, or epidermal appendages, including hair follicles, sebaceous glands, or sweat glands (Kolarsick, Kolarsick and Goodwin, 2011; Gonzales and Fuchs, 2017). Furthermore, it maintains complex regulatory networks with both the epidermis and the hypodermis (Gonzales and Fuchs, 2017; Rognoni and Watt, 2018). Hence, the dermis also exerts thermoregulatory and somatosensory functions as well as supporting roles for keratinocyte and adipocyte homeostasis.

The dermis is subdivided into two histologically distinguishable layers: the papillary and the reticular dermis (Pawlina, W. & Ross, 2015). The superficial papillary dermis is a thin, densely populated and highly irrigated and innervated layer, whereas the deeper reticular dermis represents a much thicker but less populated region (Pawlina, W. & Ross, 2015; Sriram, Bigliardi and Bigliardi-Qi, 2015). Dermal fibroblasts are known to differ between different body regions (Chang *et al.*, 2002; Rinn *et al.*, 2006), but substantial differences have also been described between papillary and reticular fibroblasts (Driskell and Watt, 2015; Sriram, Bigliardi and Bigliardi-Qi, 2015; Korosec *et al.*, 2019). Thus, papillary fibroblasts are thinner and spindle-shaped in contrast to the more expanded and squarer reticular fibroblasts (Schafer *et al.*, 1985; Janson *et al.*, 2012). They also display different proliferation rates, contractility, and cytokine and growth factor production and response (Schafer *et al.*, 1985; Sorrell and Caplan, 2004; Janson *et al.*, 2012; Sriram, Bigliardi and Bigliardi-Qi, 2015). Lastly, ECM production and organization are also significantly different between both fibroblast types. For instance, papillary fibroblasts generate thin and poorly organized collagen fiber bundles with a lower collagen type I:III ratio, while reticular fibroblasts produce thick and well-organized collagen fibers with a higher type I:III ratio (Sorrell and Caplan, 2004; Sriram, Bigliardi and Bigliardi-Qi, 2015). In addition, other ECM components like proteoglycans such as Versican or Decorin are more predominantly expressed in the reticular

and papillary dermis, respectively (Schonherr *et al.*, 1993; Sorrell and Caplan, 2004; Sriram, Bigliardi and Bigliardi-Qi, 2015).

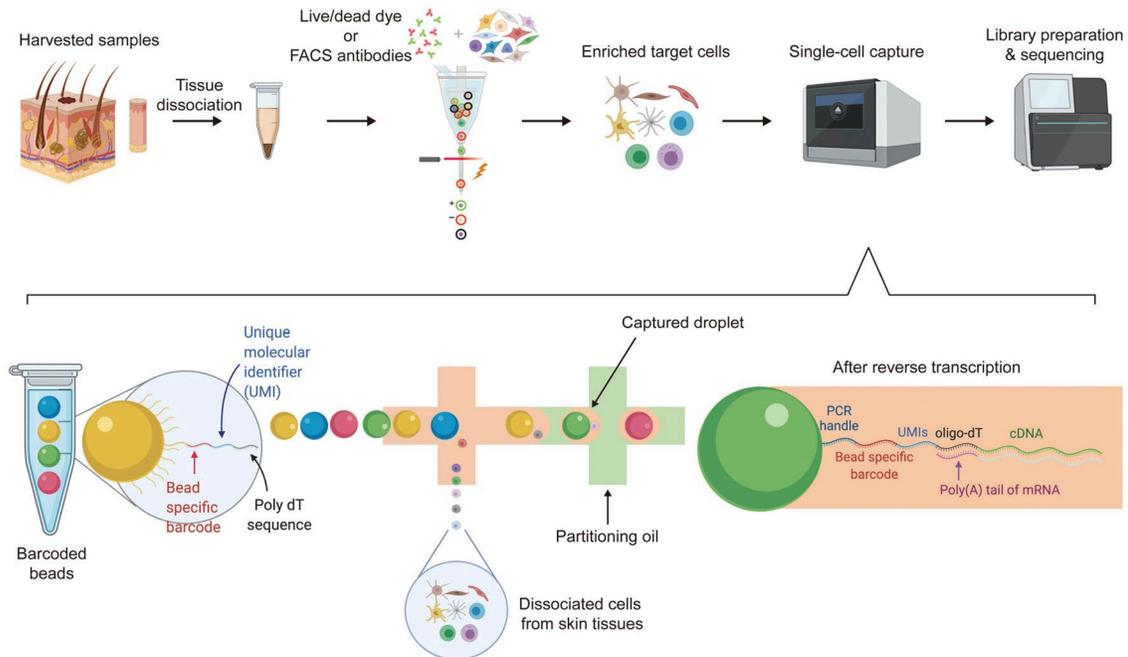
Besides their archetypical functions in maintaining skin architecture, dermal fibroblasts are also involved in several other processes in human skin homeostasis and regeneration. For instance, fibroblasts located at the papillary dermis are important components of the EpSC niche and regulate their maintenance and differentiation (Hsu, Li and Fuchs, 2014; Rognoni and Watt, 2018). Dermal fibroblasts have been shown to carry out immunoregulatory functions by secreting and responding to cytokines and chemokines (Haniffa *et al.*, 2007). But most importantly, fibroblasts play pivotal roles in wound healing processes by, for instance, promoting the initial inflammatory processes, migrating and providing a new ECM to the wound edge, supporting keratinocytes in the skin re-epithelization, or by transforming to highly contractile myofibroblasts that aid in closing the wound (Sriram, Bigliardi and Bigliardi-Qi, 2015; Stunova and Vistejnova, 2018).

Pre-adipocytes, a specialized subset of fibroblasts that differentiate into mature adipocytes, are also present in the reticular dermis (Rivera-Gonzalez, Shook and Horsley, 2014; Rognoni and Watt, 2018). Once they become mature adipocytes, they form the hypodermis. This layer of adipose tissue provides essential structural functions by acting as a kind of cushion that protects internal organs from external impacts (Rivera-Gonzalez, Shook and Horsley, 2014). In addition, the hypodermis acts as an energy source and a thermal insulator for the body (Rivera-Gonzalez, Shook and Horsley, 2014). Subcutaneous adipocytes have also been found to exert regulatory roles in the hair follicle cycle, fibrosis and scarring processes, wound healing, and immune defense (Hsu, Li and Fuchs, 2014; Rivera-Gonzalez, Shook and Horsley, 2014; Chen, Zhang and Gallo, 2019).

### **1.1.3. Human skin at single-cell resolution**

Single-cell approaches, particularly single-cell transcriptomics, have blossomed in the past decade, becoming an essential tool for understanding cellular composition and studying intercellular communication or lineage commitment in both healthy and diseased complex tissues. Several single-cell RNA-sequencing (scRNA-seq) methodologies have been developed with various cell capture approaches, including 96- or 384-well plate-based methods such as smart-seq2, or droplet-based microfluidics systems such as the 10X Genomics or the Fluidigm C1 platforms (Hwang, Lee and Bang, 2018). These methods differ in the number of captured single-cells and number of genes detected (Ziegenhain *et al.*, 2017; Hwang, Lee and Bang, 2018; Ding

*et al.*, 2020). Hence, droplet-based approaches allow for the simultaneous sequencing of thousands of cells with a lower gene detection (Figure 2), while plate-based systems can detect more genes per cell but are limited to lesser cell recovery (Ziegenhain *et al.*, 2017; Hwang, Lee and Bang, 2018; Ding *et al.*, 2020).



**Figure 2. Droplet-based single-cell transcriptomics workflow.** Collected samples are mechanically and/or enzymatically dissociated with tissue-specific protocols to obtain a single cell suspension. Optionally, dead cells can be removed, or particular target cells can be enriched using, for instance, column-based assays or FACS sorting. Samples can then be introduced into a microfluidic system such as the 10X Genomics platform for cell capture in droplets, as depicted in the figure. Each droplet contains a bead with a unique barcode and a poly-dT hybridization sequence. After cellular lysis, the mRNA from a single cell will attach by complementation with their poly(A) tails. In addition, beads also include a unique molecular identifier (UMI) for mRNA quantification.

*Adapted from Doyoung, K. et al. (2020).*

Multiple adult stem cell populations have been described in mouse IFE and HF, displaying different skin homeostasis and regeneration contributions (Solanas and Benitah, 2013; Rognoni and Watt, 2018). However, a detailed characterization of the potential human EpSC pools is still missing, and single-cell approaches offer an excellent opportunity to explore this issue (Kim, Chung and Kim, 2020). Thus, the most extensive scRNA-seq study focused on the human epidermis to date comprised almost 100,000 single-cell transcriptomes from nine healthy donors, collected from three body sites (trunk, scalp, and neonatal foreskin) as well as three psoriatic skin

samples (Cheng *et al.*, 2018). Up to eight keratinocyte populations were identified in healthy epidermis, displaying different proportions among locations. These eight populations mainly represented classic epidermal differentiation stages, including two basal undifferentiated populations, but also revealed functionally specialized keratinocytes (Cheng *et al.*, 2018). Further analyses of the neonatal foreskin keratinocytes established transcription factor (TF) modules with differential expression during epidermal differentiation, and identified ETV4 and ZBED2 as novel stemness regulators (Finnegan *et al.*, 2019). Also using human neonatal foreskin samples, a more recent study on approximately 18,000 epidermal cells uncovered further heterogeneity in the basal layer, as four distinct populations of undifferentiated keratinocytes with specific localization were identified (S. Wang *et al.*, 2020). Therefore, supporting the existence of multiple EpSC and other progenitors in the human epidermis.

Dermal fibroblast heterogeneity has also been investigated at the single-cell level, identifying transcriptionally distinct subpopulations. For instance, single-cell transcriptomics performed on 184 sorted dermal cells from a 64 y/o female donor using Smart-seq2 identified five subpopulations, with two prominent clusters separating papillary and reticular fibroblasts (Philippeos *et al.*, 2018). A more extensive study comprising 2,742 fibroblasts obtained from sun-exposed skin and using the 10X Genomics platform identified two main fibroblast subpopulations, characterized by *SFRP2/DPP4* and *FMO1/LSP1* expression (Tabib *et al.*, 2018). These subpopulations presented distinct morphology and were located in particular dermal regions, although a clear stratification of papillary and reticular fibroblasts was not found. In addition, the authors described five minor subpopulations. However, the functional contribution of the distinct fibroblast subpopulations remained largely unclear (Tabib *et al.*, 2018).

### **1.1.4. Effects of aging on the human skin**

Chronological aging refers to the gradual and deleterious biological changes that affect tissue structure and functionality over time. Due to the external location of the skin, its aging process is highly influenced by extrinsic factors, whose effects are superimposed to the ones of the intrinsic chronological aging. The ethnicity and gender of the individual as well as the anatomical region comprise the main intrinsic factors influencing skin aging, while sun exposure comprises the main external factor (Rittié and Fisher, 2015; Krutmann *et al.*, 2017). In fact, chronic sun exposure accelerates skin aging and leads to a particular set of effects known as photoaging (Rittié and Fisher, 2015). Other important extrinsic factors include air pollution, smoking, nutrition, stress and sleep deprivation, and the use of cosmetic products (Rittié and Fisher, 2015; Krutmann *et al.*,

2017). Both intrinsic and extrinsic aging affects all human skin layers, resulting in a thin, dry, and wrinkled appearance with loss of elasticity and increased fragility (Rittié and Fisher, 2015; Haydont, Bernard and Fortunel, 2019).

One of the most characteristic structural changes observed in aged skin is the overall epidermal thinning and the dermal-epidermal junction (DEJ) flattening (Giangreco *et al.*, 2010; Rittié and Fisher, 2015). This results in the shortening or loss of the typical invaginations of the human epidermis, known as rete ridges (Mine *et al.*, 2008; Giangreco *et al.*, 2010). Such structural alterations have been linked to the progressive reduction of the proliferative capacity of basal keratinocytes over time (Kligman, 1979; Giangreco *et al.*, 2010). The flattening of the DEJ implies a decreased interacting surface between the epidermis and dermis, thus reducing the communication and nutrient transfer between layers and the resistance of the epidermis to mechanical stimuli (Langton *et al.*, 2016; Haydont, Bernard and Fortunel, 2019). Corneocytes are not majorly affected by aging; therefore, the balance of water loss is not substantially affected (Roskos, Maibach and Guy, 1989; Luebberding, Krueger and Kerscher, 2013). However, an impaired barrier function of the epidermis has also been observed for certain substances in the aged skin (Roskos, Maibach and Guy, 1989; Luebberding, Krueger and Kerscher, 2013). Lastly, the old epidermis also presents a reduced regenerative capacity in wound healing processes (Gosain and DiPietro, 2004; Rittié and Fisher, 2015).

Progressive atrophy of the dermis is one of the most profound effects of skin aging. Such atrophy is due to the progressive loss of a well-organized ECM, particularly aggravated in photoaged skin (Rittié and Fisher, 2015; Haydont, Bernard and Fortunel, 2019). Thus, collagen production in aged skin is reduced concomitant with increased degradation leading to an overall reduction of collagen fibers (Haydont, Bernard and Fortunel, 2019; Shin *et al.*, 2019). Elastic fibers from the papillary dermis are also degraded over time in sun-protected skin (Langton *et al.*, 2012). However, non-functional and disorganized elastic fibers accumulate in the sun-exposed aged dermis, a phenomenon known as solar elastosis (Shin *et al.*, 2019). The expression of other ECM components such as proteoglycans or glycosaminoglycans is also altered upon aging but differs between intrinsically or extrinsically aged skin (Haydont, Bernard and Fortunel, 2019; Shin *et al.*, 2019). Increased expression of matrix metalloproteinase (MMP) members, the responsible enzymes for ECM degradation, are also found to be increased in old human skin in response to the age-dependent accumulation of reactive oxygen species (ROS) (Golden, Hinerfeld and Melov, 2002; Mine *et al.*, 2008; Fisher *et al.*, 2009). All these ECM remodeling effects are responsible for the generation of wrinkles, loss of elasticity, and gain of laxity of the aged skin (Haydont, Bernard and Fortunel, 2019; Shin *et al.*, 2019).

At the cellular level, fibroblasts become reduced in numbers and size as they collapse due to their reduced contacts with the ECM (Quan *et al.*, 2013; Fisher *et al.*, 2016). Furthermore, due to the long-lived nature of fibroblasts, cellular damages accumulate over time, particularly those inflicted by photoaging (Tigges *et al.*, 2014; Lee and Liu, 2020). Some research indicates that fibroblasts from the distinct dermal layers are differently affected by aging. Thus, aging has been proposed to primarily affect papillary fibroblasts but not the reticular ones in an *in vitro* system (Mine *et al.*, 2008). This was in agreement with a reduced papillary dermis thickness in aged skin (Mine *et al.*, 2008). Furthermore, aged papillary fibroblasts might acquire reticular fibroblast features (Sriram, Bigliardi and Bigliardi-Qi, 2015). Thus far, most knowledge about fibroblast aging in humans was generated using *in vitro* systems, uncovering that dermal fibroblasts display most of the hallmarks of aging (Tigges *et al.*, 2014). However, the assessment of such hallmarks *in vivo* is still limited and further research is needed.

### **1.2. Keratinocyte cancers: a rising public health issue.**

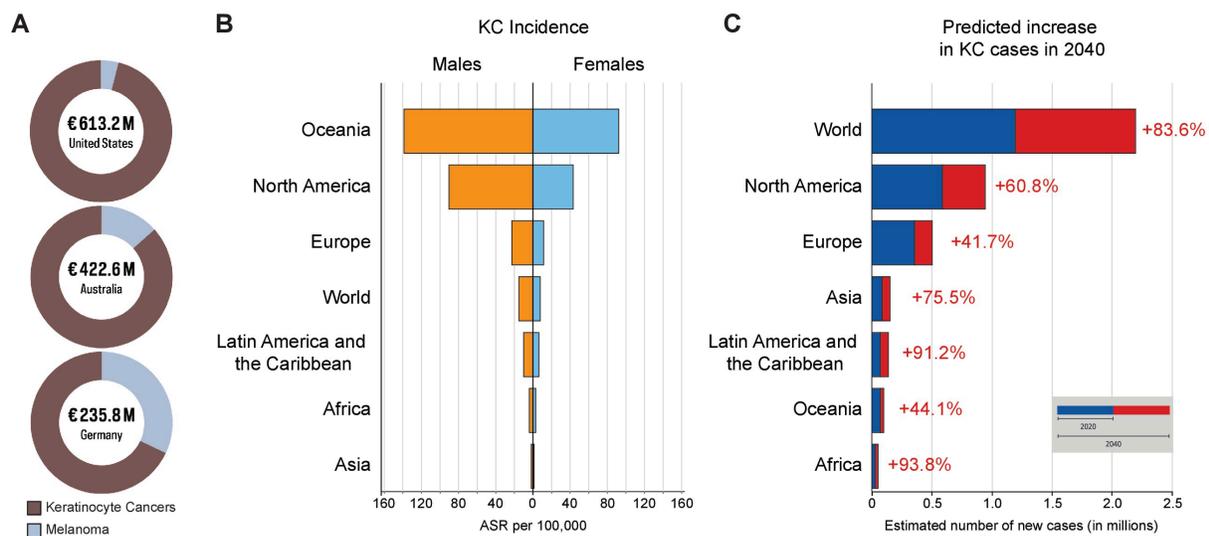
Skin cancer has been traditionally divided into two main groups: melanoma and keratinocyte cancers (KC), often referred to as non-melanoma skin cancer (NMSC). Other rare skin cancers include Merkel cell carcinoma (MCC) or cutaneous T-cell lymphoma (CTCL). The incidence of all these malignancies is alarmingly increasing worldwide, despite numerous prevention campaigns (Korgavkar, Xiong and Weinstock, 2013; Paulson *et al.*, 2018; Fitzmaurice *et al.*, 2019). This is of particular importance for KC as it represents the most common type of skin cancer, and the main cancer type in fair skin populations. The high numbers of yearly diagnosed KC cases and the associated morbidities explain the heavy sanitary and economic burden that KC poses on public health systems (Figure 3A) (Chen, Kempton and Rao, 2016; Mofidi *et al.*, 2018).

#### **1.2.1. Epidemiology, risk factors and pathology of KC.**

Keratinocyte cancers mainly comprise basal cell carcinoma (BCC) and cutaneous squamous cell carcinoma (cSCC), which account for 80% and 20% of the cases, respectively (Apalla, Lallas, *et al.*, 2017; Bartoš and Kullová, 2018). While melanoma originates from melanocytes, KC arises from epidermal keratinocytes and represents the most common skin cancer type, with an estimated incidence twenty times higher than melanoma (Apalla, Nashan, *et al.*, 2017; Stang, Khil and Kaj, 2019). The incidence of KC has risen in the last decades, as illustrated by the estimated 33% increase in the total number of cases globally between 2007 and 2017 (Fitzmaurice *et al.*,

2019), and it is predicted to continue to increase (Figure 3B and 3C). Only in 2017, around 7.7 million cases of KC were registered worldwide (Fitzmaurice *et al.*, 2019). Moreover, these alarming numbers are most probably an underestimation due to the lack of proper documentation of KC in national cancer registries in most countries.

Despite comprising the most common skin malignancy, BCC is generally associated with low mortality rates, as its metastatic potential has been estimated to be less than 0.05% (Apalla, Lallas, *et al.*, 2017). However, BCC presents high morbidity due to the many diagnosed cases and its frequent high invasiveness and associated tissue degeneration (Cameron *et al.*, 2019). While BCC does not arise from precursor lesions, cSCC arises in a multi-step carcinogenic process. Thus, most cSCC develop from a pre-cancerous dysplasia named actinic keratosis (AK) or an *in situ* carcinoma known as Bowen's disease (BD) (Ratushny *et al.*, 2012; Burton, Ashack and Khachemoune, 2016). The annual rate of progression to invasive cSCC is relatively low and was broadly estimated at 0.025%-16% for AK (Glogau, 2000) and 3%-5% for BD (Kao, 1986; Kossard and Rosen, 1992).



**Figure 3. Global epidemiological data regarding KC. (A)** Direct costs derived from the treatment and care of KC and melanoma in United States of America, Australia and Germany, three countries with high KC incidence (values in 2013 euros) **(B)** Estimated age-standardized incidence of BCC and cSCC per continent and gender in 2020 using all age-groups (0-85+). World values are also shown. **(C)** Estimated number of new cases of BCC and cSCC from 2020 to 2040 in each continent, calculated using both sexes and all ages (0-85+). World values are also shown and percentage of increase is depicted in red. ASR = age-standardized rates, M= millions.

Sources: **(A,B)** International Agency for Research on Cancer (IARC); **(C)** The Cancer Atlas (3<sup>rd</sup> edition).

However, most patients present multiple lesions at the time of diagnosis, which increases the risk of developing an invasive cSCC (Ratushny *et al.*, 2012). Of note, around 30% of AK regress spontaneously (Marks, Rennie and Selwood, 1988), but the molecular mechanisms leading to the progression or regression of these lesions are still poorly understood.

Most cSCC, and its precursor lesions, are diagnosed and surgically removed in early stages, therefore presenting a good prognosis. Nonetheless, around 5% are able to metastasize each year and are associated with poor clinical outcomes as they display a five-year survival rate lower than 30% (Ratushny *et al.*, 2012; Burton, Ashack and Khachemoune, 2016). Thus, risk prediction is of particular interest for these malignancies, and several staging systems have been developed. Tumor diameter  $\geq 2$  cm, perineural invasion, and poor differentiation are considered high-risk factors (Burton, Ashack and Khachemoune, 2016).

The main risk factor for developing KC is chronic sun exposure, specifically to ultraviolet radiation (UVR) (Bowden, 2004). UVR induces DNA damage in epidermal keratinocytes, which is usually repaired by nucleotide excision repair (NER) (Yu and Lee, 2017). However, errors in this DNA damage repair system lead to an accumulation of somatic mutations with a specific pattern at dipyrimidine sites (C>T and CC>TT) (Ikehata and Ono, 2011; Yu and Lee, 2017). As a result, sun-exposed healthy epidermis harbors a high mutational burden, leading to the generation of patches with mutated keratinocytes with carcinogenic potential (Albibas *et al.*, 2018; Wei *et al.*, 2021). Of note, this overtime accumulation results in a late age of onset for KC, which primarily affects individuals over 60 y/o (Hernando *et al.*, 2021). An increase in the aged population and changes in the behavior towards recreational sun exposure largely explain the increased incidence of KC, and skin cancer in general, in the last years. Importantly, UVR from tanning salons also increases the risk for developing KC (Wehner *et al.*, 2012), as illustrated by the increased incidence of BCC in young women using such methods (Christenson *et al.*, 2005).

Even though both types of carcinoma arise from keratinocytes due to chronic UVR exposure, the mutagenic landscape of both entities displays substantial differences. Thus, while both entities present extremely high mutational burdens (cSCC: 50 mutations/Mb; BCC: 65 mutations/Mb), different genes and pathways are affected (Jayaraman *et al.*, 2014; Inman *et al.*, 2018). BCC tumorigenesis is driven mainly by activation of the Sonic Hedgehog (SHH) signaling pathway (Cameron *et al.*, 2019; Dika *et al.*, 2020). Thus, up to 85% of the human BCCs display mutations in genes related to this pathway (i.e. *PTCH1*, *SMO*, *SUFU*, and *TP53*) (Bonilla *et al.*, 2016). In contrast, the most frequent mutations in cSCC involve the inactivation of tumor suppressor *TP53*, which leads to high genomic instability (Dotto and Rustgi, 2016; Inman *et al.*, 2018). Other commonly affected genes in cSCC include the loss-of-function of *CDKN2A*, which leads to

increased proliferation; tumor-suppressors *NOTCH1/2*; and epigenetic factors such as *KMT2C*, *KMT2D*, and *TET2* (Pacifico *et al.*, 2008; South *et al.*, 2014; Chitsazzadeh *et al.*, 2016; Inman *et al.*, 2018). Importantly, similar mutational patterns have been found in pre-invasive AK, and only a few potential driver mutations for AK to cSCC progression have been described, including mutations affecting the TGF- $\beta$  pathway (Thomson *et al.*, 2021).

Additional risk factors for developing KC include intrinsic factors such as skin pigmentation, gender, genetic features, and immunosuppression, as well as extrinsic factors such as human papillomavirus (HPV) infection (Nagarajan *et al.*, 2019). Thus, the incidence of KC is higher in men than in women, as well as in people with less skin pigmentation, blond or red hair, and blue or green eyes (Heaton and Lawrence, 2019; Nagarajan *et al.*, 2019; Budden *et al.*, 2021). Immunosuppression is considered a high-risk factor for developing KC, especially in solid organ transplant recipients (SOTR) (Lowenstein *et al.*, 2017; Madeleine *et al.*, 2017). The incidence of cSCC in such patients has been estimated to be 60- to 200-fold higher than in their immunocompetent counterparts, resulting in a switch in the BCC/cSCC ratio (Lowenstein *et al.*, 2017; Madeleine *et al.*, 2017). Higher cSCC risk has also been described in patients with rare hereditary diseases such as Fanconi anemia, epidermolysis bullosa, or oculocutaneous albinism, while higher BCC risk has been observed in basal cell nevus syndrome patients (Nagarajan *et al.*, 2019; Cives *et al.*, 2020). Also, the genetic disorder xeroderma pigmentosum has been associated with a higher risk for cSCC, BCC, and melanoma (Nagarajan *et al.*, 2019; Cives *et al.*, 2020). Lastly, HPV infection and chronic inflammatory diseases of the skin such as psoriasis or atopic dermatitis have been linked to a higher KC risk (Cheng *et al.*, 2015; Chahoud *et al.*, 2016; Cives *et al.*, 2020).

### 1.2.2. Defining the cellular origin of KC

The cellular origin of cancers has been proposed to influence the evolution of the disease and some of the tumor features such as differentiation and aggressiveness (Visvader, 2011; Blanpain, 2013). Thus, the identification of such tumor-initiating cells might provide important diagnostic and therapeutic opportunities (Visvader, 2011; Blanpain, 2013). Substantial effort has been made to identify the cells-of-origin of KC, with several cell types being highlighted as possible origins. Most knowledge on this topic comes from genetically modified mouse models and lineage tracing approaches. Thus, several mouse models have been generated by inducing the activation of oncogenes or inactivation of tumor suppressor genes in particular epidermal populations using the Cre recombinase system under the control of specific promoters (Blanpain, 2013; Das and

Yaniv, 2020). Furthermore, coupling the Cre recombinase to the estrogen receptor (ER) allows for the activation of the Cre recombinase in a controlled manner by the addition of tamoxifen (Blanpain, 2013; Das and Yaniv, 2020).

The cellular origin of BCC is still highly debated, and several cell types have been proposed as candidates. Early studies suggested the interfollicular basal layer, hair follicle keratinocytes, and even eccrine glands as possible origins for human BCC (Zackheim, 1963; Heenan and Bogle, 1993; Krüger, Blume-Peytavi and Orfanos, 1999). More recent research also found contradictory results, with immunostaining assays suggesting both a hair follicle and an interfollicular origin (Sellheyer, 2011; Tan *et al.*, 2018). To shed some light on the matter, several studies used BCC mouse models that mimic the human condition by either depleting *Ptch1* or introducing a constitutively active mutant of Smo (SmoM2), therefore leading to activation of the SHH pathway. Thus, inducing SmoM2 expression in hair follicle stem cells and their progenies did not induce BCC formation (Youssef *et al.*, 2010). In contrast, BCC emerged upon expression of this oncogene in long-lived progenitors from the IFE and the infundibulum, the region that connects the IFE with the hair follicles (Youssef *et al.*, 2010). In a subsequent study, IFE progenitors were found to undergo a transcriptional reprogramming to an embryonic hair follicle progenitor-like phenotype after expression of SmoM2, thus highlighting that the expression of hair follicle markers in BCC does not necessarily imply a HF origin (Youssef *et al.*, 2012). Nevertheless, and in direct opposition, other studies described that BCC could arise from several hair follicle keratinocyte populations upon depletion of *Ptch1*, while IFE stem cells did not generate tumors (Kasper *et al.*, 2011; Peterson *et al.*, 2015). In addition, BCCs were also found to arise from touch dome cells, a subset of innervated interfollicular keratinocytes with mechanosensory functions, upon loss of *Ptch1* (Peterson *et al.*, 2015). Whether the differences observed in these studies are due to the different oncogenic drivers used or reflect distinct cells-of-origin of BCC is still unclear.

The most common approach to study cSCC in mice is through chemically induced carcinogenesis by topical application of the 7,12-dimethyl-1,2-benzanthracene (DMBA) mutagen and the proliferation and inflammation-promoting agent (TPA) (Huang and Balmain, 2014). This treatment promotes the development of benign papillomas (mice equivalents to AK) that can progress to invasive cSCC. DMBA-TPA treatment leads to cSCC initiation primarily by generating mutations in oncogenes from the RAS family, such as *Hras* and *Kras* (Huang and Balmain, 2014). Thus, several studies use the overexpression or the introduction of constitutively active forms of such oncogenes (i.e *Kras*<sup>G12D</sup>) to mimic the effects of DMBA-TPA treatment. For instance, the overexpression of *Hras* in differentiated keratinocytes under the control of the *K10* promoter, lead to the formation of cSCC, but only at wounding sites (Bailleul *et al.*, 1990).

Induction of *Kras*<sup>G12D</sup> expression in hair follicle stem cells promoted benign papilloma development, while the same oncogene did not result in tumor formation from transient amplifying keratinocytes of the hair follicle (Lapouge *et al.*, 2011; White *et al.*, 2011). Also, papillomas were found to develop from IFE basal cells and progenitors upon expression of *Kras*<sup>G12D</sup> (Lapouge *et al.*, 2011). Nevertheless, *Kras*<sup>G12D</sup> alone could not promote the development of invasive cSCC and the additional depletion of *Trp53* was needed (Lapouge *et al.*, 2011; White *et al.*, 2011). Thus, Latil *et al.* (2017) evaluated the capacity of IFE basal and HF keratinocytes to develop invasive cSCC upon combined oncogenic *Kras*<sup>G12D</sup> expression and *Trp53* deletion under the control of *K14* and *Lgr5* promoters, respectively. Although both cell types gave rise to malignant cSCC, those arising from IFE cells displayed a well-differentiated phenotype while those originating from HF populations showed poor differentiation and features of epithelial-to-mesenchymal transition (EMT) (Latil *et al.*, 2017). Collectively, these studies suggest that keratinocytes from both the interfollicular and the hair follicle compartment comprise the cells-of-origin of cSCC in mice. However, how these studies translate into the initiation of cSCC in humans is still largely unknown.

### 1.3. Role of DNA methylation in keratinocyte cancers

Epigenetics refers to the study of heritable changes in gene expression without modifying the DNA sequence but through the regulation of chromatin architecture and accessibility. Epigenetic gene regulation depends on the coordinated action of diverse mechanisms, including covalent histone modifications, DNA methylation, chromatin remodelers, and non-coding RNAs (ncRNAs). The epigenetically regulated expression patterns are crucial in establishing and maintaining cellular state and identity, yet are reversible and highly dynamic, thus granting cellular plasticity (Avgustinova and Benitah, 2016; Flavahan, Gaskell and Bernstein, 2017). Disruption of the epigenetic landscape is a hallmark of several human conditions and diseases such as aging and cancer (López-Otín *et al.*, 2013; Flavahan, Gaskell and Bernstein, 2017). Hence, the epigenome of cancer cells is characterized by abnormal DNA methylation and histone modification patterns as well as deregulation of ncRNAs expression (Rodríguez-Paredes and Esteller, 2011; Flavahan, Gaskell and Bernstein, 2017). Of note, DNA methylation aberrations are often detected in the early stages of tumorigenesis, thus playing important roles in cancer initiation and progression.

### 1.3.1. The mammalian DNA (de)methylation machinery

DNA methylation is a covalent epigenetic modification that results from transferring a methyl group from S-adenosylmethionine to the fifth carbon of the cytosine ring, thus forming 5-methylcytosine (5mC). A family of enzymes known as DNA methyltransferases (DNMTs) catalyzes this reaction (Lyko, 2018). In mammals, such modification occurs mainly in symmetrical CpG dinucleotides, which in humans have been estimated in roughly 28 million (Smith and Meissner, 2013). Of those, around 60-80% are usually methylated (Smith and Meissner, 2013). Importantly, most of the genome is depleted of CpG sites. However, small regions of ~1000 bp with a high CpG density can be found across the genome, known as CpG islands. These islands are often unmethylated and mostly located close to gene promoters (Deaton and Bird, 2011). In fact, approximately 70% of the annotated promoters are associated with CpG islands (Saxonov, Berg and Brutlag, 2006). DNA methylation is considered a repressive epigenetic mechanism as it is involved in the repression of transposable elements, X-chromosome inactivation, regulation of genomic imprinting, and generation of heterochromatin (Smith and Meissner, 2013; Greenberg and Bourc'his, 2019). However, it also dynamically regulates gene transcription. Thus, DNA methylation at promoters has been linked with the silencing of gene expression, while unmethylated promoters with concomitant methylated gene bodies are usually present in actively transcribed genes (Greenberg and Bourc'his, 2019).

Five DNMTs can be found in humans: DNMT1, DNMT2, DNMT3A, DNMT3B, and DNMT3L. However, only DNMT1, DNMT3A, and DNMT3B are responsible for establishing and maintaining DNA methylation as DNMT3L does not have catalytic activity (Lyko, 2018), and DNMT2 was shown not to methylate DNA but tRNAs instead (Goll *et al.*, 2006). Thus, DNMT3A and DNMT3B are responsible for the *de novo* methylation of DNA during embryogenesis and adult homeostasis (Lyko, 2018; Greenberg and Bourc'his, 2019), with the supporting role of the inactive DNMT3L, which acts as a cofactor for their activity (Ooi *et al.*, 2007). DNA methylation patterns are maintained during DNA replication through the action of DNMT1 and its essential cofactor, the E3 ubiquitin-protein ligase UHRF1 (Bostick *et al.*, 2007; Lyko, 2018). Thus, UHRF1 selectively binds to hemimethylated DNA strands at the replication fork and then recruits DNMT1, which methylates the nascent DNA strand (Bostick *et al.*, 2007; Lyko, 2018).

DNA can be demethylated in a passive or active manner. Passive DNA demethylation occurs as a result of successive rounds of DNA replication without the maintenance of DNA methylation (Hashimoto *et al.*, 2012). Conversely, active demethylation is mediated by the ten-eleven translocation (TET) enzyme family activity. Three different TET proteins are encoded in the

human genome: TET1, TET2, and TET3 (An, Rao and Ko, 2017). These enzymes catalyze the progressive 5mC oxidation into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (Ito *et al.*, 2011). DNMT1 does not recognize these modified cytosines, thus resulting in demethylation upon replication (Hashimoto *et al.*, 2012; An, Rao and Ko, 2017). The balance between DNA methylation and demethylation shapes the methylation landscape, particularly at regulatory regions such as gene promoters and enhancers (Hon *et al.*, 2014; Rinaldi *et al.*, 2016; Zhang *et al.*, 2016).

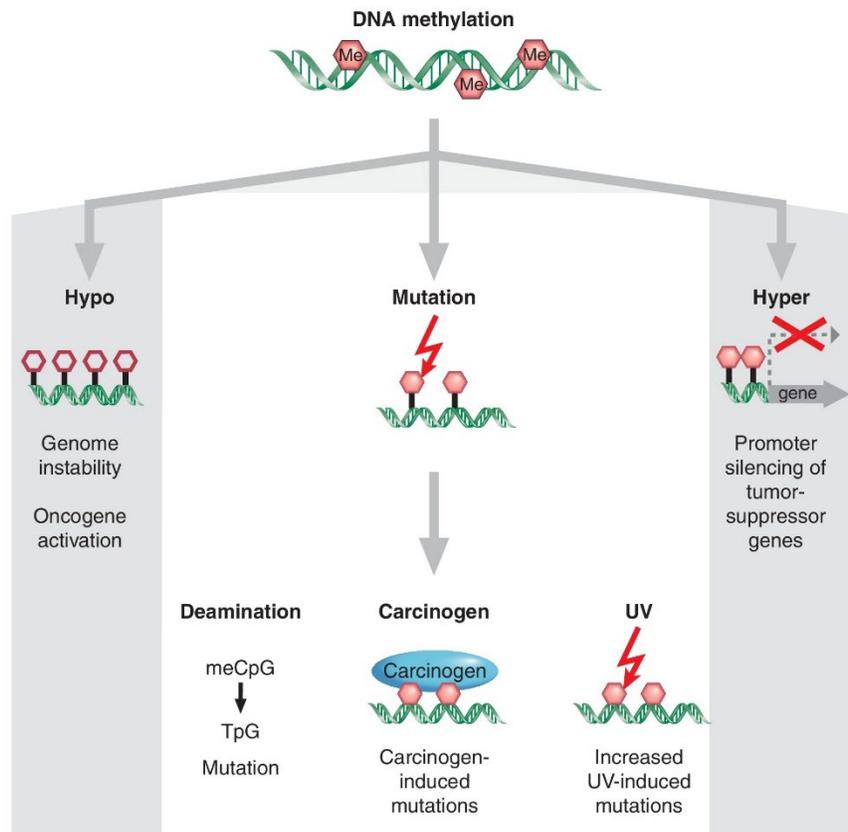
### 1.3.2. DNA methylation aberrations in KC

Disruption of the DNA methylation patterns in cancer cells results in a general rearrangement of the 5mC distribution. Thus, cancer cells typically present a global hypomethylation of the genome concomitant with local hypermethylation at CpG islands (Figure 4) (Baylin and Jones, 2016). Furthermore, DNA methylation also contributes to carcinogenesis by direct mutagenesis as 5mC residues have an increased mutagenic predisposition (Figure 4). Importantly, all these alterations occur simultaneously and lead to different cancer-promoting events (Baylin and Jones, 2016).

While in healthy tissues, roughly 80% of the CpGs are methylated, this percentage is substantially reduced in malignant tissues, ranging from 40% to 60% (Baylin and Jones, 2016). This loss is observed as large blocks of DNA hypomethylation and has been associated with activation of oncogenes and an increased genomic instability due to the activation of repetitive sequences and transposable elements (Berman *et al.*, 2012; Bert *et al.*, 2013; Ehrlich and Lacey, 2013). Nonetheless, how this global DNA methylation loss occurs is still largely unknown. This phenomenon has also been observed in KC. For instance, loss of DNA methylation in the repetitive long interspersed nuclear elements-1 (LINE-1) was observed in cSCC and psoriatic epidermis (Yooyongsatit *et al.*, 2015). Since LINE-1 elements represent approximately 17% of the genome, the authors suggested that its hypomethylation could be extrapolated to the rest of the genome (Yooyongsatit *et al.*, 2015). Notably, a genome-wide analysis observed broad hypomethylation blocks in cSCC, which were already present in photoaged healthy epidermis (Vandiver *et al.*, 2015). Hence, suggesting that disruptions in DNA methylation in the human epidermis occur prior to malignant transformation.

Hypermethylation at CpG islands is usually linked with the silencing of tumor suppressor genes and is one of the most studied epigenetic alterations in cancer (Shen and Laird, 2013; Baylin and Jones, 2016). Several genes are silenced through this mechanism in KC. For instance, the *FHIT* gene, involved in DNA damage responses, and genes related to the SHH and WNT signaling pathways (*SHH*, *APC*, *SFRP5*, and *RASSF1A*) were found to be downregulated by promoter

hypermethylation in BCC (Goldberg *et al.*, 2006; Brinkhuizen *et al.*, 2012). In contrast, low-level methylation was described in the promoter of *PTCH*, suggesting only a minor role for DNA methylation in regulating the expression of this particular gene in BCC (Heitzer *et al.*, 2010). Some of the most relevant genes found to be silenced in cSCC include genes associated with the RB1/P16 and p53 signaling pathways (i.e., *CDH1*, *CDKN2A*, *DAPK1*, *MGMT*, *RB1*, and *RASSF*), genes regulating epidermal proliferation and differentiation (i.e., *GOS2*, *ID4* and *IRF6*) as well as some miRNAs (i.e., *MIR204*) (Murao *et al.*, 2006; Li, Sawalha and Lu, 2009; Toll *et al.*, 2016). In addition, *CCND1* was found to be down-regulated in arsenic-induced BD, while *S100A4* was silenced in BD, BCC and cSCC (Li *et al.*, 2009).



**Figure 4. DNA methylation alterations with cancer-promoting properties.** The scheme depicts the three main mechanisms by which DNA methylation promotes cancer initiation and progression. First, the general DNA hypomethylation observed in cancer cells leads to higher genomic instability and activation of oncogenes. Moreover, local hypermethylation at CpG islands located in gene promoters influences carcinogenesis by silencing tumor-suppressor genes. Methylated cytosines are more prone to mutation than unmethylated cytosines as they spontaneously deaminate, resulting in C > T transitions. In addition, methylated CpGs facilitate the binding of certain carcinogens to DNA and are more sensitive to UVR, thus increasing its mutagenesis. Full hexagon: methylated cytosine; Empty hexagon: unmethylated hexagon.

*Adapted from Baylin, S.B. and Jones, P.A. (2016).*

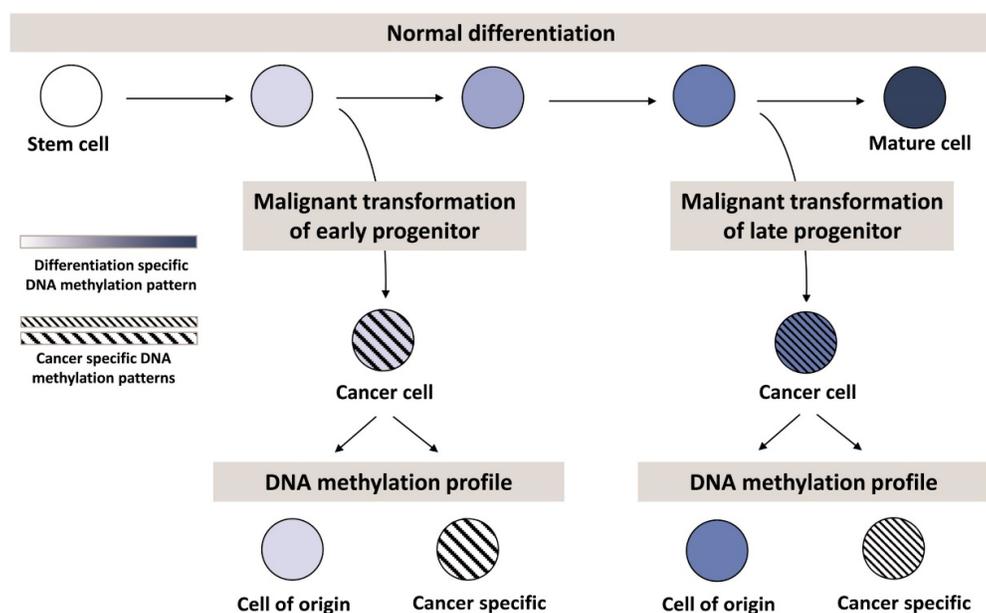
The cancer-promoting direct mutagenesis of methylated CpG sites can occur through various mechanisms (Figure 4). First, methylated cytosines have an inherently increased rate of deamination, resulting in C to T transitions that can lead to the inactivation of tumor suppressor genes (Holliday and Grigg, 1993). In addition, the methylation residue in CpG sites favors the binding of chemical carcinogens to DNA, resulting in increased mutagenesis (Greenblatt *et al.*, 1994). Lastly, and of particular interest for the skin cancer field, DNA methylation increases the mutagenesis of UV radiation (de Oliveira, de Souza and de Castro Coêlho, 2020). Methylated cytosines absorb more energy from UV light, favoring the generation of cyclobutane pyrimidine dimers (CPDs) (Ikehata and Ono, 2011). These CPDs are unstable and easily deaminate, thus inducing C to T substitutions. Hence, methylated sites comprise mutational hotspots in important driver genes for KC development, such as *TP53* (Greenblatt *et al.*, 1994; Pfeifer, Tang and Denissenko, 2000).

The role of DNA methylation in the progression of AK to cSCC has only recently been addressed. Thus, in a previous study from Prof. Dr. Lyko's group, the methylomes of 12 healthy epidermis, 16 AK, and 18 cSCC samples were profiled using the Infinium MethylationEPIC BeadChips (Illumina) (Rodríguez-Paredes *et al.*, 2018). These arrays represent the current gold standard method for DNA methylation analysis and allow the quantitative analysis of approximately 850,000 CpG and non-CpG sites, covering the vast majority of the genome with enrichment for regulatory elements such as gene promoters and enhancers (Moran, Arribas and Esteller, 2016). This study found major methylation changes in both AK and cSCC compared to healthy epidermis, with both entities displaying the classical cancer-specific methylation features. Nonetheless, no statistically significant variation could be observed in the progression from AK to cSCC, indicating that DNA methylation aberrations occur early in this disease continuum but do not play important roles in its progression (Rodríguez-Paredes *et al.*, 2018). In contraposition, another study using the same technology identified a set of 94 CpGs with variable methylation status between AK and metastatic cSCC (Hervás-Marín *et al.*, 2019).

### **1.3.3. Cell-of-origin-related DNA methylation patterns in cancer**

DNA methylation patterns are highly cell type-specific due to their essential role in establishing cellular identity. Furthermore, dynamic changes in DNA methylation during differentiation have been described in several lineages, including epidermal keratinocytes. Hence, a specific loss of DNA methylation in lineage-specific regulatory elements with a concomitant gain of methylation in regulatory elements of other lineages was observed upon murine keratinocyte differentiation

(Bock *et al.*, 2012). Similarly, human keratinocyte differentiation was associated with subsequent hypomethylation of epidermal differentiation-related promoter genes (Sen *et al.*, 2010). Consistently, the expression of *DNMT1* and its cofactor *UHRF1* is reduced upon differentiation while *DNMT3A* and *DNMT3B* expression increases (Sen *et al.*, 2010; Rinaldi *et al.*, 2016). Functional analyses using genetically engineered mouse models and human cell lines illustrate the relevance of DNA methylation in maintaining epidermal homeostasis. For instance, DNMT1-deficient human keratinocytes implanted in a mouse xenograft model showed premature differentiation and failed to generate a fully developed epidermis (Sen *et al.*, 2010). A similar effect was seen upon knock-down of *UHRF1*, thus indicating that the DNMT1-UHRF1 coordinated action maintains the EpSC phenotype (Sen *et al.*, 2010). Comparably, human EpSC self-renewal was impaired upon knock-down of *DNMT3A* or *DNMT3B*, and differentiation was also affected upon DNMT3A reduction (Rinaldi *et al.*, 2016). Mechanistically, DNMT3A and DNMT3B were found to bind and promote the activity of enhancers associated with genes driving EpSC self-renewal and differentiation (Rinaldi *et al.*, 2016). Nevertheless, conditional depletion of *Dnmt1* in the basal layer of the epidermis in mice did not result in gross alterations, nor did the depletion of *Dnmt3a*, *Dnmt3b*, or combined, which could reflect compensatory effects between DNMTs (Li *et al.*, 2012; Rinaldi *et al.*, 2017).



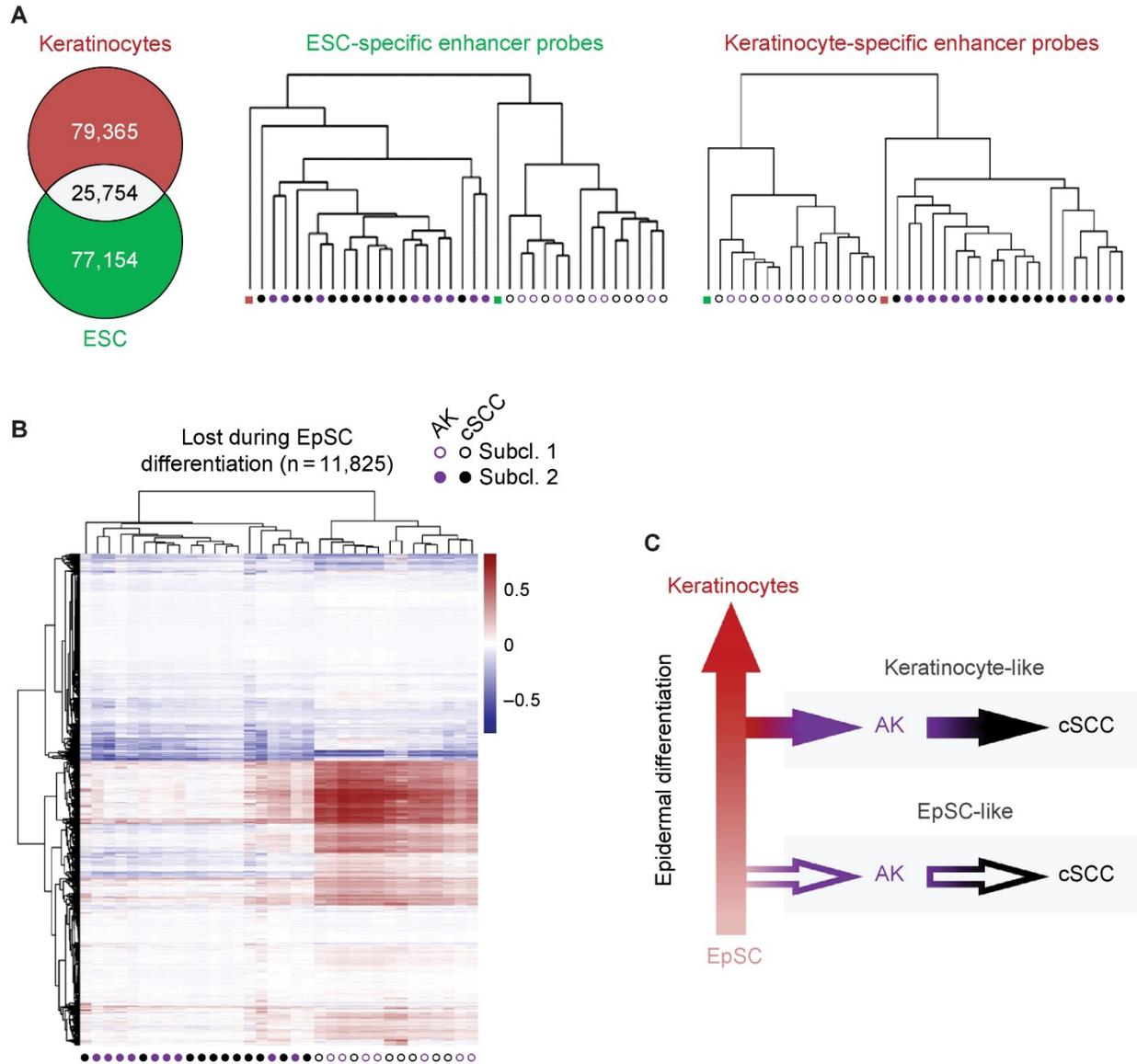
**Figure 5. DNA methylation patterns in cancer cells reflect their cell-of-origin.** Cell-type-specific DNA methylation patterns are acquired upon normal differentiation (depicted in a gradient of blue). Upon malignant transformation, DNA methylation in cancer cells becomes aberrant (depicted as black lines), but cell-of-origin patterns are substantially maintained and still detectable.

*Adapted from Kulis, M. et al. (2013).*

The DNA methylation patterns established during embryogenesis or cellular differentiation thus comprise a sort of molecular fingerprint for particular cell types and states. Such fingerprints can also be partially detected in cancer cells in addition to the specific methylation changes acquired during malignant transformation (Figure 5) (Kulis *et al.*, 2013; Moran *et al.*, 2016). As a matter of fact, the tumor stratification of the 33 types of human cancers contained in The Cancer Genome Atlas (TCGA) dataset was found to be influenced mainly by cell-of-origin patterns (Hoadley *et al.*, 2018). Furthermore, the clinical relevance of DNA-methylation-based cell-of-origin subclasses has been proposed in several human malignancies such as colorectal cancer (Bormann *et al.*, 2018), epithelial ovarian cancers (Io Riso *et al.*, 2020), or various types of leukemias (Oakes *et al.*, 2016; Wierzbinska *et al.*, 2020). Thus, DNA methylation can be used as a robust biomarker for tumor stratification based on their cell-of-origin, with potential benefits for patient risk assessment.

In this regard, two subclasses of AK and cSCC were identified based on their methylation profiles (Rodríguez-Paredes *et al.*, 2018). Such subclasses were firstly observed by analyzing the methylation patterns of keratins, a group of genes whose expression in epithelial tissues depends on cell type and differentiation stage (Moll, Divo and Langbein, 2008). To further investigate whether these subclasses reflected distinct cells-of-origin, the authors stratified the 16 AK and 18 cSCC samples included in the study based on the methylation status of enhancers defined in normal human epidermal keratinocytes (NHEK) or H1 human embryonic stem cells (ESC). This analysis identified the same tumor subclasses, one clustering with keratinocytes and the other with ESCs (Figure 6A). Lastly, tumor stratification based on the methylation patterns present in 11,825 enhancers known to be lost upon epidermal differentiation (Rinaldi *et al.*, 2016) classified AK and cSCC samples into the same two groups (Figure 6B). Thus, these subclasses were interpreted to reflect the epigenetic programs of two distinct cells of origin: one closely related to an EpSC and the other to a more differentiated keratinocyte (Figure 6C).

Recent advances have made it possible to study DNA methylation at a single-cell resolution (Karemaker and Vermeulen, 2018). Such approaches have the potential to critically improve the identification of the tumor-initiating cells in human malignancies. Nonetheless, most available techniques thus far are still limited to a low number of cells, low CpG coverage, or both (Karemaker and Vermeulen, 2018). Therefore, further research towards robust high-throughput single-cell methylomics methodologies is needed and will be crucial to bring this field forward into both fundamental and clinical research.



**Figure 6. DNA methylation profiling identifies two cell-of-origin subclasses of AK and cSCC. (A)** Left: Venn diagram defining the exclusive enhancer regions in normal human keratinocytes and human embryonic stem cells (ESC). Right: Unsupervised hierarchical clustering of 16 AK and 18 cSCC epidermal samples according to the methylation patterns in ESC or keratinocyte-specific enhancers. **(B)** Unsupervised hierarchical clustering of 16 AK and 18 cSCC samples based on the methylation patterns in enhancers known to be lost upon epidermal differentiation. **(C)** Schematic of the proposed bimodal cell-of-origin for AK and cSCC, arising from two distinct epidermal differentiation stages.

*Adapted from Rodríguez-Paredes, M. et al. (2018)*

## 2. Aims

Based on their DNA methylation profiles, AK and cSCC samples were stratified into two subclasses, which were interpreted to reflect two distinct cells-of-origin: one closely related to an EpSC and one to a more differentiated keratinocyte. However, whether this stratification strategy could also be applied to other epidermal malignancies, whether it had clinical implications and what were the exact cells-of-origin was still unclear.

The central aim of this thesis was to expand the epigenomic characterization and stratification of KC by analyzing a larger cohort (n=102), including healthy epidermis, AK, and cSCC but also BD, BCC, and non-cancerous senile warts (seborrheic keratosis; SK). This cohort was generated using Infinium MethylationEPIC BeadChips and represented the full spectrum of malignancy in epidermal tumors. The thesis also aimed to interrogate the potential prognostic value of the cell-of-origin-based stratification, as well as to further characterize the cells-of-origin of KC using cutting-edge single cell-omics techniques such as single-cell combinatorial indexing for methylation analysis (sci-MET) and single-cell RNA-sequencing.

In addition, single-cell analyses were extended to dermal fibroblasts in order to comprehensively analyze their diversity in human skin homeostasis and to investigate the changes they undergo upon intrinsic aging. To that end, a dataset containing more than 15,000 single-cell transcriptomes from sun-protected healthy whole human skin obtained from five young and old donors was generated and characterized in this work.



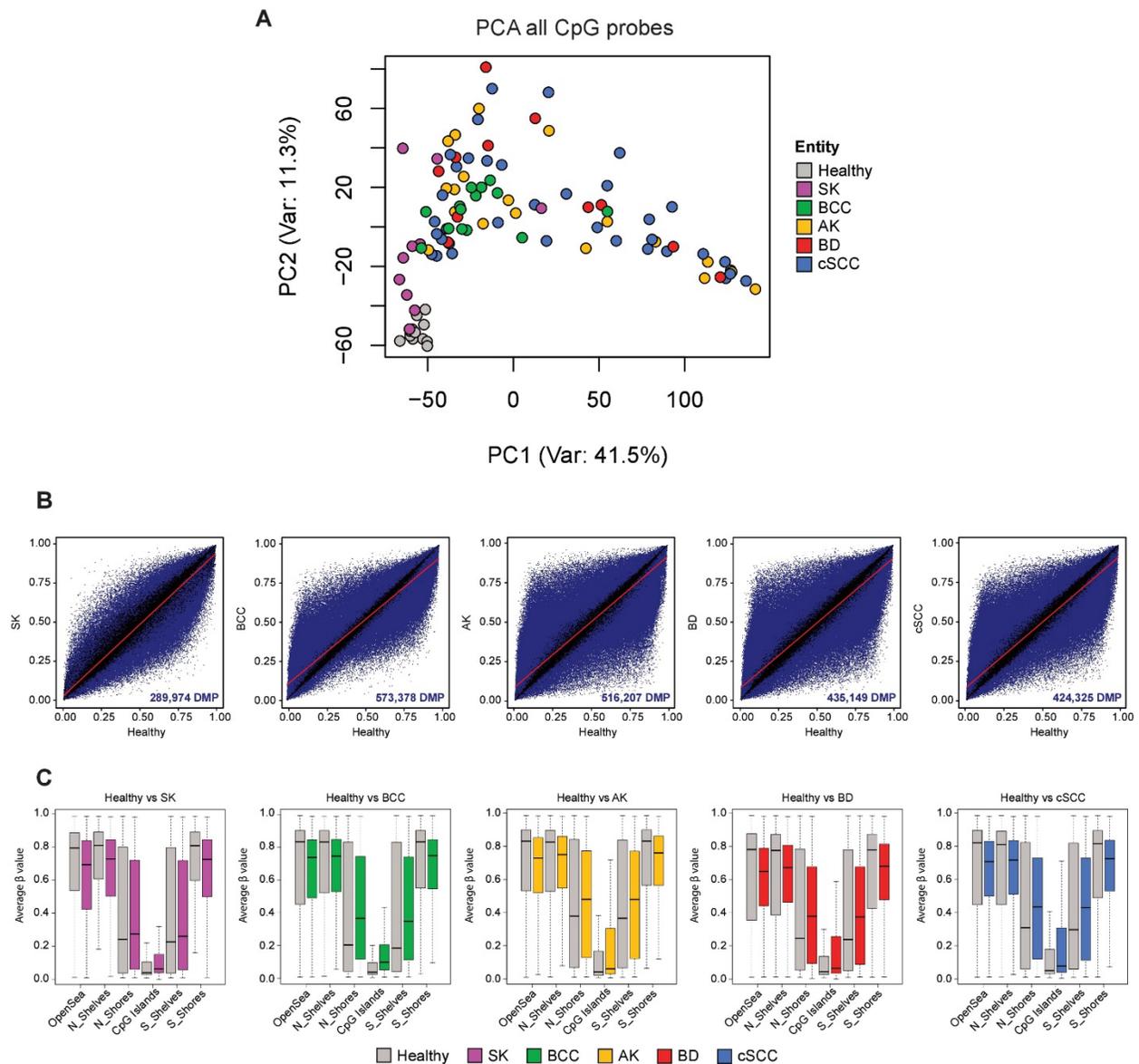
## 3. Results

### 3.1. Methylation-based stratification of keratinocyte cancers

To further explore and improve the proposed cell-of-origin stratification strategy for keratinocyte cancers (KC), the previously published Infinium Methylation EPIC dataset (Rodríguez-Paredes *et al.*, 2018) was enlarged in the scope of this thesis to 12 healthy epidermis, 20 AK, and 35 cSCC samples. Moreover, to assess whether this classification could be applied to other epidermal tumors, 11 *in situ* carcinoma (BD) samples and 14 rarely metastatic BCC were also analyzed, as well as ten non-cancerous senile warts (seborrheic keratosis; SK). Thus, the complete cohort comprised a total of 102 methylomes and examined the full spectrum of malignancy of epidermal tumors. To avoid confounding factors due to tissue heterogeneity, all samples were taken as punch biopsies from the center of the lesion, and the epidermis part of the tumor was isolated from the dermis (see Methods, section 6.2.1.1). Raw sequencing data was batch-corrected, normalized, and filtered to eliminate low-quality or confounding probes (see Methods, section 6.2.3) prior to downstream analysis with the R package Minfi (Aryee *et al.*, 2014).

#### 3.1.1. Epidermal tumors display aberrant methylomes

In a first analysis, principal component analysis (PCA) with all remaining CpG probes (n= 632,778) was performed. Healthy epidermis was clearly separated from tumor samples, which did not group by entity (Figure 7A). Interestingly, non-cancerous SK samples displayed an intermediate position (Figure 7A). Pairwise comparisons between each epidermal tumor entity and healthy specimens revealed that all KC types showed aberrant methylomes and presented cancer-specific methylation features (Figure 7B). These included a global hypomethylation concomitant with local hypermethylation at CpG islands (Figure 7C). These observations held true for non-cancerous SK samples, possibly reflecting its hyperproliferative phenotype (Figure 7C).



**Figure 7. Epidermal tumors display archetypal cancer-specific methylation features. (A)** Principal Component analysis of 12 Healthy, 10 SK, 14 BCC, 20 AK, 11 BD, and 35 cSCC epidermal samples using all CpGs after filtering ( $n=632,778$ ). Coloring is according to sample type. **(B)** Scatter Plots displaying pairwise comparisons between each epidermal tumor entity and healthy epidermis. Differentially methylated probes ( $P < 0.05$ , F-test) are depicted in blue. **(C)** Box plots displaying the average methylation values in the different genomic regions in relation to CpG islands for each epidermal tumor entity compared to healthy epidermis. SK: seborrheic keratosis, BCC: basal cell carcinoma, AK: actinic keratosis, BD: Bowen's disease, cSCC: cutaneous squamous cell carcinoma, DMP: differentially methylated probes, N: north, S: south.

### 3.1.2. Keratinocyte enhancers stratify KC based on their cell-of-origin

The proposed classification strategy in Rodríguez-Paredes *et al.*, 2018 was based on the methylation patterns of a particular set of enhancers (n=11,825) known to be specific of EpSCs and lost upon differentiation (Rinaldi *et al.*, 2016). Unsupervised clustering based on the methylation profiles of these enhancers resulted in the stratification of the KC cohort into two main groups, one closely related to EpSCs (EpSC-like) and another one closely related to a more differentiated keratinocyte (Keratinocyte-like; Figure 8A). Notably, healthy epidermis samples, which mainly contain differentiated keratinocytes, did not form an independent cluster from the tumors, thus, highlighting the closer association of the Keratinocyte-like tumors to more differentiated cells. Tumor stratification into two groups could also be observed in the PCA analysis with all CpG probes (Figure 8B). Importantly, tumor stratification was not due to differences in tumor purity or increased immune cell content (Figure S1).

Of note, the new AK and cSCC, as well as the *in situ* carcinoma (BD) samples, stratified into both cell-of-origin subclasses. These three entities were classified in an approximate 60% Keratinocyte-like – 40% EpSC-like ratio (Table 1 and Figure 8A). In contrast, rarely metastatic BCC and non-cancerous SK were almost exclusively classified as Keratinocyte-like, therefore, suggesting a bias towards a more differentiated keratinocyte as the cell-of-origin for epidermal tumors with a lower metastatic potential.

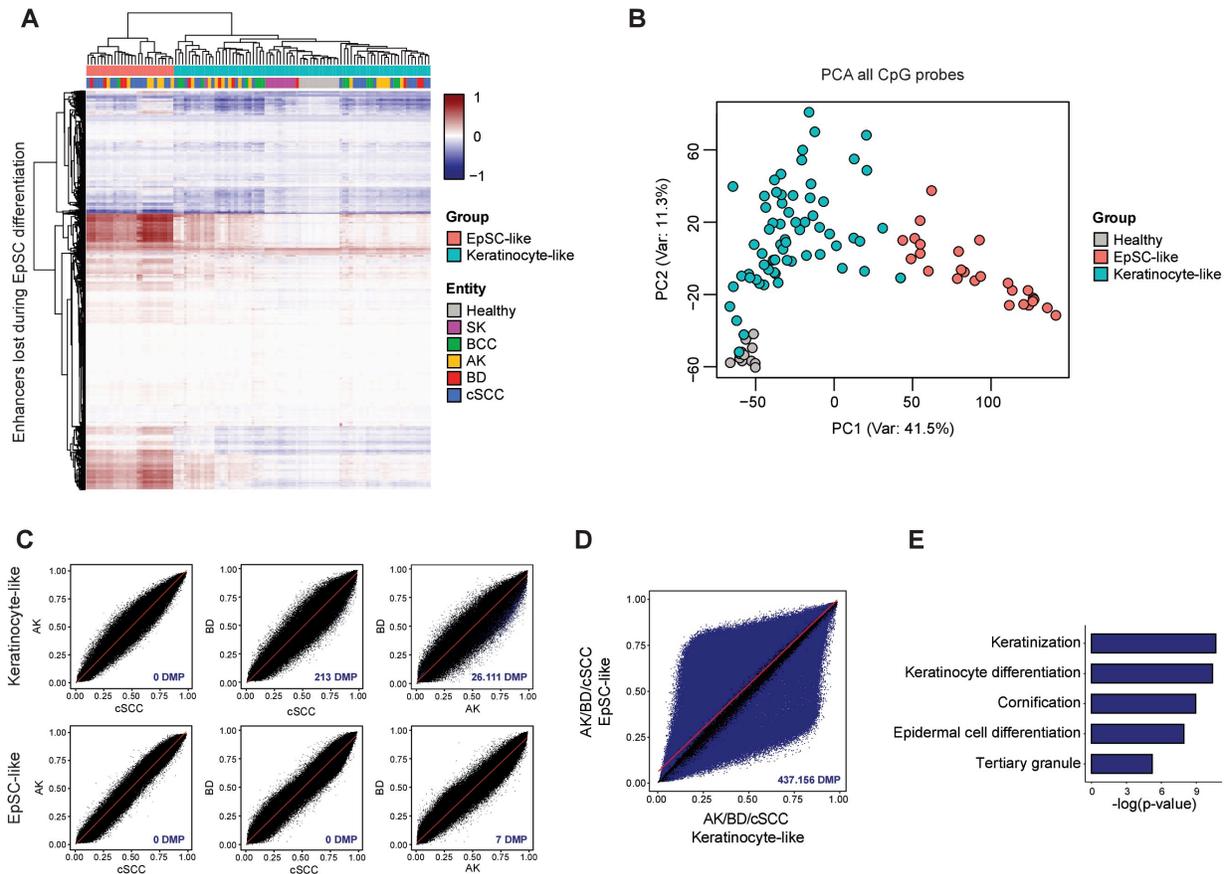
	Keratinocyte-like	EpSC-like
SK	10/10 (100%)	0/10 (0%)
BCC	13/14 (92.9%)	1/14 (7.1%)
AK	11/20 (55%)	9/20 (45%)
BD	7/11 (63.6%)	4/11 (36.4%)
cSCC	21/35 (60%)	14/35 (40%)

**Table 1. Cell-of-origin-based stratification of human epidermal tumors.** Ratio of samples from each entity stratified as Keratinocyte-like or EpSC-like tumors based on the methylation profiles at enhancers known to be lost upon epidermal differentiation. SK: seborrheic keratosis, BCC: basal cell carcinoma, AK: actinic keratosis, BD: Bowen’s disease, cSCC: cutaneous squamous cell carcinoma.

The methylation differences between epidermal tumors arising from distinct cells-of-origin were investigated in further analyses. A particular focus was placed on cSCC and the two precursor lesions included in the study, AK and BD, as they are part of the same disease continuum and the main tumor entities displaying a bimodal cell-of-origin stratification. Thus, pairwise

### 3. Results

comparisons between pre-invasive lesions (BD and AK) and cSCC samples from the same cell-of-origin subclass uncovered highly similar methylomes as almost no differentially methylated probes were identified (Figure 8C). Furthermore, comparing both pre-invasive lesions between them revealed no significant differences in the EpSC-like group and only a few DMPs ( $n=26,111$ ; adjusted  $P < 0.05$ , F-test) in the Keratinocyte-like group (Figure 8C).



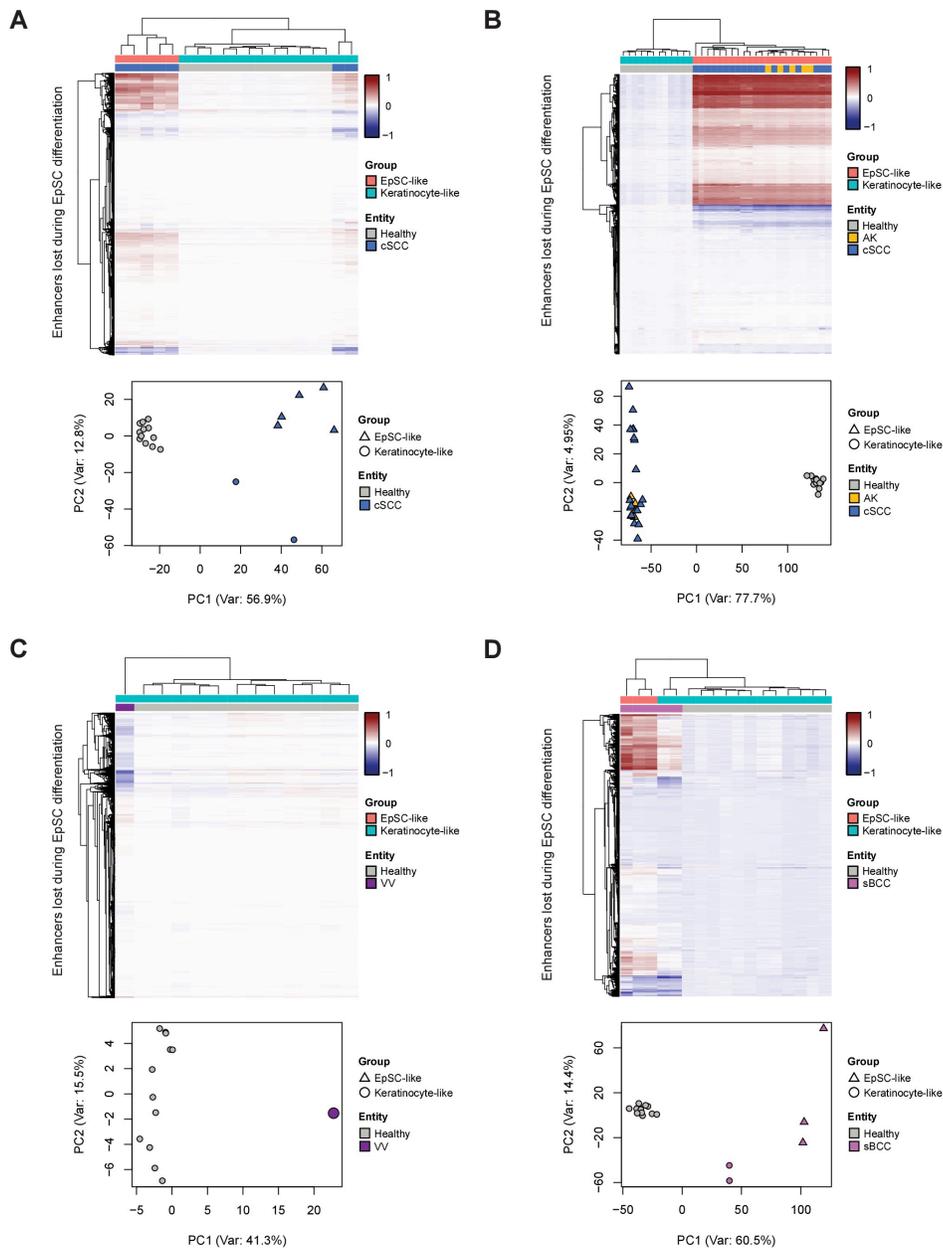
**Figure 8. EpSC-specific enhancer methylation patterns stratify KC in two cell-of-origin subclasses. (A)** Heatmap shows unsupervised hierarchical clustering of the KC cohort based on the methylation patterns of enhancers lost upon epidermal differentiation ( $n=9,932$  enhancers). **(B)** PCA of 12 Healthy, 10 SK, 14 BCC, 20 AK, 11 BD, and 35 cSCC epidermal samples using all CpGs after filtering ( $n= 632,778$ ). Coloring is according to the cell-of-origin subclass, except for healthy samples. **(C)** Scatter Plots displaying pairwise comparisons between AK, BD, and cSCC samples from the Keratinocyte-like (upper) or EpSC-like (lower) subclass. **(D)** Scatter plot displaying the comparison between AK/BD/cSCC samples from each cell-of-origin subclass as a unique entity. **(E)** Top five (by p-value) enriched GO terms using genes with differentially methylated promoters between AK/BD/cSCC samples from distinct cell-of-origin subclasses. In scatter plots, differentially methylated probes ( $P < 0.05$ , F-test) are depicted in blue. SK: seborrheic keratosis, BCC: basal cell carcinoma, AK: actinic keratosis, BD: Bowen’s disease, cSCC: cutaneous squamous cell carcinoma, DMP: differentially methylated probes.

In contrast, comparison between AK/BD/cSCC samples from the distinct cell-of-origin subclasses as a unique group yielded a total of 437,156 DMPs (Figure 8D). Gene Ontology (GO) analysis performed with the genes whose promoters were differentially methylated showed a strong enrichment for terms related to epidermal differentiation (i.e., Keratinization, Keratinocyte Differentiation, or Epidermal Cell Differentiation) (Figure 8E). Thus, further confirming that the primary source of distinction between tumor subclasses was the differentiation stage of their cell-of-origin. Overall, these findings indicate almost indistinguishable methylomes between pre-invasive, *in situ* carcinomas and cSCCs arising from the same tumor-initiating cell type and highlight significant epigenetic differences between cell-of-origin subclasses.

### 3.1.3. Validation of the cell-of-origin-based stratification of epidermal tumors in independent datasets

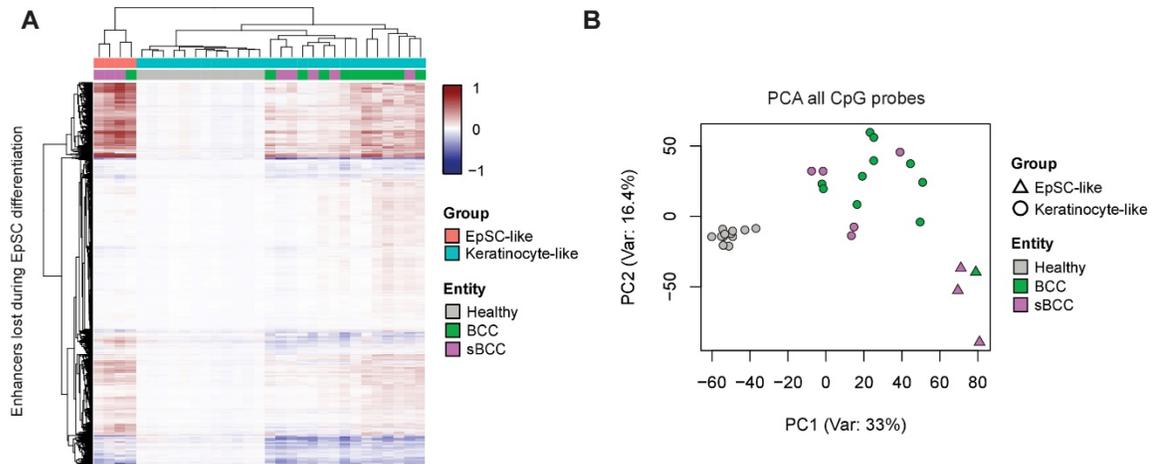
Publicly available methylation datasets were used to further validate the cell-of-origin-based stratification strategy of KC. Each dataset was combined with the 12 healthy epidermis samples from the KC cohort as the average methylation in healthy epidermis was subtracted from the tumors data to highlight differences in enhancer methylation patterns better (see Methods, section 6.2.4). The first dataset to be analyzed consisted of seven primary cSCC samples profiled with HumanMethylation450 BeadChips (Vandiver *et al.*, 2015). Unsupervised clustering according to EpSC-specific enhancer methylation patterns, stratified the tumors into both cell-of-origin subclasses (Figure 9A). Notably, healthy epidermis does not cluster separately from tumors but together with the Keratinocyte-like subclass, as described above (Figure 9A). The separation between cell-of-origin subclasses could also be observed in a PCA performed with all CpG probes (Figure 9A). Importantly, stratification of another EPIC dataset, consisting of five AK and 18 invasive cSCC samples (Hervás-Marín *et al.*, 2019), uncovered only one group since all samples were classified as EpSC-like (Figure 9B). PCA analysis performed on all probes also showed high homogeneity among Hervás-Marín *et al.* samples (Figure 9B). These findings further validate the proposed stratification strategy for cSCC samples while suggesting a more invasive phenotype in EpSC-like cSCCs.

An EPIC methylation dataset consisting of 12 verruca vulgaris (VV, also known as common warts) was also analyzed (Al-Eitan *et al.*, 2020). These VV are benign epidermal tumors associated with HPV infection (Iida *et al.*, 2020). Raw methylation data was not available for this dataset; thus, we used the processed mean beta values per sample group as provided in the original publication (Al-Eitan *et al.*, 2020).



**Figure 9. Cell-of-origin-based stratification of independent KC methylation datasets.** Upper: Heatmaps showing the unsupervised hierarchical clustering based on the methylation patterns of EpSC-specific enhancers of **(A)** seven cSCC samples from Vandiver *et al.*, 2015 **(B)** five AK and 18 cSCC samples from Hervás-Marín *et al.* 2019 **(C)** 12 averaged VV samples from AL Eitan *et al.* 2020 and **(D)** five sBCC samples from Sand *et al.* 2019. Lower: PCA using all CpG probes after filtering in datasets comprised of **(A)** seven cSCC samples from Vandiver *et al.* 2015; **(B)** five AK and 18 cSCC samples from Hervás-Marín *et al.* 2019; **(C)** 12 averaged VV samples from AL Eitan *et al.* 2020; and **(D)** five sBCC samples from Sand *et al.* 2019. All datasets were combined with 12 healthy epidermis samples. In PCA, coloring is according to sample type and shape is according to cell-of-origin subclass. SK: seborrheic keratosis, BCC: basal cell carcinoma, AK: actinic keratosis, BD: Bowen’s disease, cSCC: cutaneous squamous cell carcinoma, VV: verruca vulgaris, sBCC: sclerodermiform basal cell carcinoma.

Enhancer methylation patterns of VV samples resembled those of a more differentiated cell-of-origin and were highly similar to those present in healthy epidermis (Figure 9C). These results exhibit a strong similarity to those of the analysis of SK samples and, thus, indicate a more differentiated cell-of-origin for both senile (SK) and common (VV) warts.



**Figure 10. Combined analysis of BCC and sBCC samples. (A)** Heatmaps showing the unsupervised hierarchical clustering of the 14 in-house BCC (11 BCC and 3 sBCC) samples and five sBCC tumors from Sand *et al.*, based on the methylation patterns of enhancers lost upon epidermal differentiation. **(B)** Principal Component analysis of 12 Healthy, 11 BCC, and 8 sBCC samples using all CpGs after filtering. Coloring is according to sample type, and shape is according to cell-of-origin subclass. BCC: basal cell carcinoma, sBCC: sclerodermiform basal cell carcinoma.

Lastly, an EPIC methylation dataset consisting of five sclerodermiform BCC (sBCC) was also analyzed (Sand *et al.*, 2019). sBCC is considered one of the most aggressive subtypes of BCC as it presents high recurrence rates and higher local invasiveness (Conforti *et al.*, 2021). Enhancer methylation patterns classified three out of five samples as EpSC-like (Figure 9D). Again, the separation between both cell-of-origin subclasses could also be observed by PCA using all probes (Figure 9D). Three samples from the 14 BCC analyzed in section 3.1.2 were also diagnosed as sBCC and were stratified as Keratinocyte-like in the initial analysis. A combined stratification of these 14 BCC and the five samples from Sand *et al.* 2019 confirmed these results (Figure 10). Hence, only four out of the 19 BCC samples studied were classified as EpSC-like, and of those, three were from the sBCC subtype (Figure 10). These results indicate an enrichment for the more aggressive sBCC subtype in the EpSC-like BCC subclass.

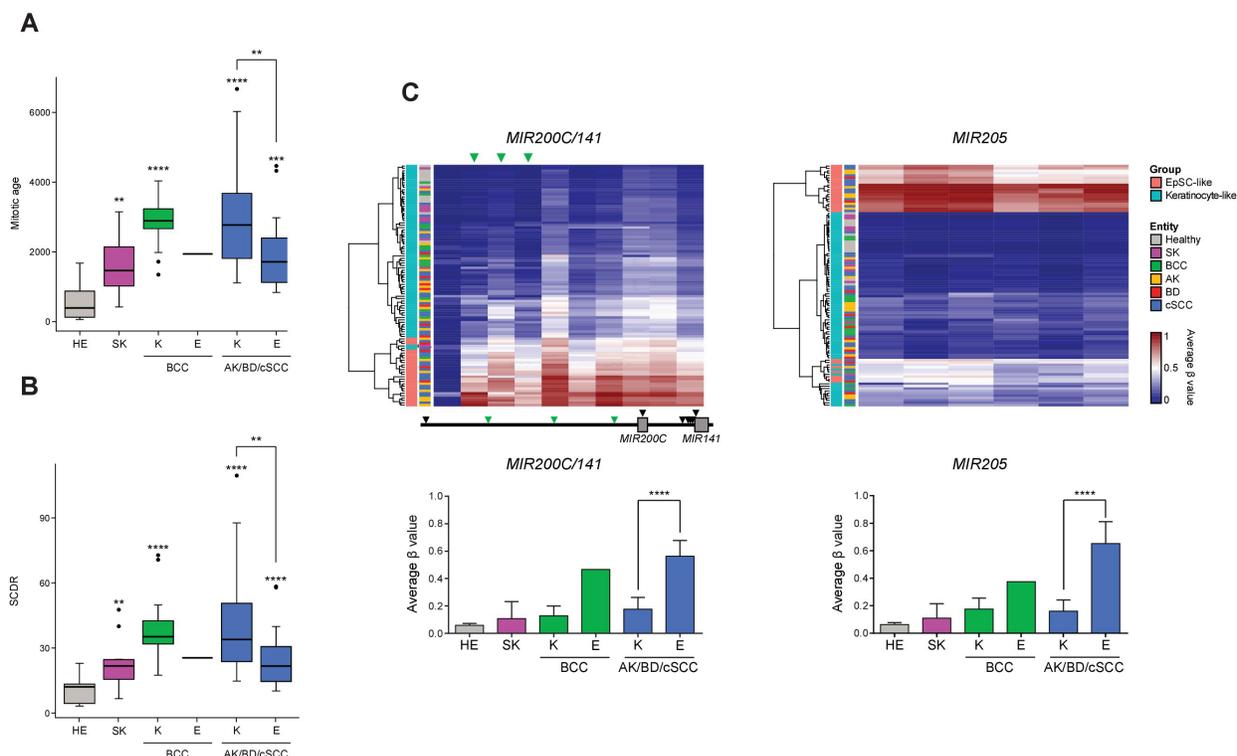
Taken together, the analysis of these independent datasets provided important validation for the observations obtained by the DNA methylation-based stratification of the KC cohort.

### 3.1.4. KC subclasses display distinct division rates and invasive features

DNA methylation can be used to estimate the mitotic age (cumulative stem cell division rates) of a given tissue (Duran-Ferrer *et al.*, 2020; Teschendorff, 2020). An increase in mitotic age has been observed in, for instance, human inflamed tissues or pre-malignant and malignant lesions (Duran-Ferrer *et al.*, 2020; Teschendorff, 2020). In that regard, the mitotic clock epiTOC2 can estimate the mitotic age based on a set of CpGs located at promoters of the polycomb repressive complex 2 (PRC2) target genes, known to be unmethylated in fetal tissues but to progressively become methylated upon aging (Teschendorff, 2020). Thus, epiTOC2 was used to assess whether the distinct KC entities and cell-of-origin subclasses display distinct mitotic ages and proliferative rates. Increased mitotic age was observed in SK samples in comparison to healthy epidermis, reflecting their hyperproliferative behavior (Figure 11A). Of note, a more substantial increase was detected in BCC and AK/BD/cSCC tumors belonging to the Keratinocyte-like subclass, while the EpSC-like tumors showed a more moderate increase (Figure 11A).

In subsequent analyses with epiTOC2, the intrinsic stem cell division rate (SCDR) of the epidermal tumors was calculated. An estimated SCDR of 10.35 divisions per stem cell per year was estimated in healthy epidermis (Figure 11B), in agreement with experimentally-assessed division rates of human EpSCs (Maeda, 2017). Consistent with the increased mitotic age, an increase in SCDR was observed in SK samples, compared to healthy epidermis (SCDR = 22.9) (Figure 11B). Division rates were also globally increased in KC tumors, with a more pronounced effect observed in BCC (SCDR = 39.2) and AK/BD/cSCC (SCDR=39.7) from the Keratinocyte-like group (Figure 11B). A more moderate effect was again observed in BCC (SCDR=25.5) and AK/BD/cSCC (SCDR = 25) from the EpSC-like subclass (Figure 11B). Overall, these results are consistent with the archetypical hyperproliferative phenotype present in tumors and suggest different proliferation rates for the two cell-of-origin subclasses.

Cellular proliferation and invasion are two independent processes with a high degree of anticorrelation in cancer cells (Gao *et al.*, 2005; Hoek *et al.*, 2008; Hecht *et al.*, 2015). Cancer cells displaying epithelial-to-mesenchymal transition (EMT) features show characteristic DNA methylation patterns such as promoter hypermethylation of specific miRNAs, including the *MIR200* family or *MIR205* (Wiklund *et al.*, 2011; Davalos *et al.*, 2012). Thus, to examine whether the less proliferative EpSC-like subclass corresponds to a more invasive phenotype, the methylation status of these miRNAs was analyzed. Consistent with the abovementioned hypothesis, the promoter regions for *MIR200C/141* cluster and *MIR205* were hypermethylated in

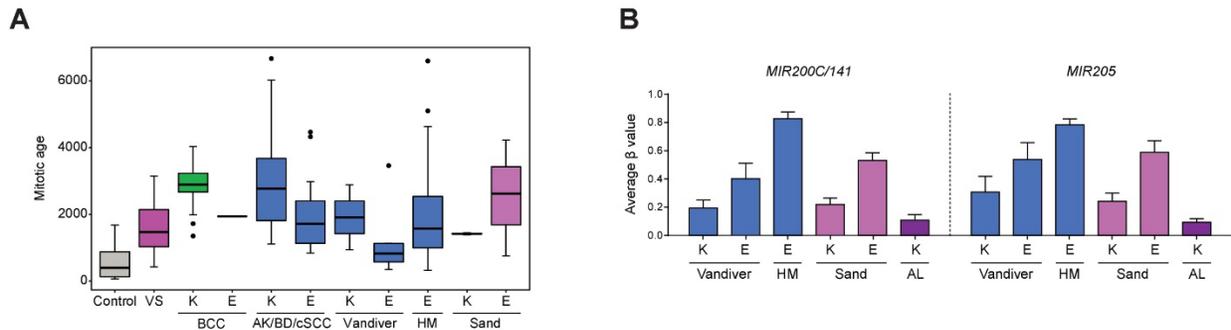


**Figure 11. KC subclasses show distinct proliferative and invasive features. (A)** Boxplots displaying the estimated mitotic age calculated with epiTOC2 in each tumor entity and cell-of-origin subclass. **(B)** Boxplots displaying the estimated stem cell division rate per stem cell and year in each tumor entity and cell-of-origin subclass. **(C)** Upper: Heatmaps showing the unsupervised clustering of the KC cohort, based on the methylation patterns of the *MIR200C/141* cluster and *MIR205* gene. Lower: Bar plots showing the average methylation in the promoter regions of *MIR200C/141* cluster (green arrows) and *MIR205* gene (all probes depicted). Statistical analyses were performed using a Wilcoxon Rank Sum test. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . In **(A)** and **(B)** comparisons between healthy epidermis and each tumor entity are shown as an asterisk, while comparison between tumor subclasses are depicted with a line. HE: healthy epidermis, SK: seborrheic keratosis, BCC: basal cell carcinoma, AK: actinic keratosis, BD: Bowen's disease, cSCC: cutaneous squamous cell carcinoma, K: Keratinocyte-like, E: EpSC-like.

EpSC-like tumors (Figure 11C). Taken together, these findings suggest distinct proliferative rates and invasive potential for both KC cell-of-origin subclasses.

These notions were also examined in the datasets previously used for validation. Hence, despite observing lower mitotic age values in all samples from Vandiver *et al.* 2015 in comparison to the KC cohort, differences between cell-of-origin subclasses were still evident, and EpSC-like tumors displayed less mitotic age (Figure 12A). Also, the samples from Hervás-Marín *et al.* 2019 showed a similar mitotic age to the EpSC-like AK/BD/cSCC tumors from the KC cohort (Figure 12A). In contrast, in sBCC samples, the contrary effect could be observed as the EpSC-like subclass displayed a higher mitotic age, comparable to the Keratinocyte-like BCC and

AK/BD/cSCC samples (Figure 12A). On the other hand, *MIR200C/141* and *MIR205* promoter hypermethylation were found in the EpSC-like tumors of all datasets (Figure 12B). These findings further suggest a less proliferative but more invasive phenotype for EpSC-like tumors except for sBCC.



**Figure 12. Mitotic age and EMT-associated methylation in validation datasets. (A)** Boxplots displaying the estimated mitotic age calculated with epiTOC2 in the KC cohort and validation datasets. **(B)** Bar plots showing the average methylation in the promoter regions of *MIR200C/141* cluster and *MIR205* gene (as described in Figure 11). HE: healthy epidermis, SK: seborrheic keratosis, BCC: basal cell carcinoma, AK: actinic keratosis, BD: Bowen's disease, cSCC: cutaneous squamous cell carcinoma, VV: verruca vulgaris, sBCC: sclerodermiform basal cell carcinoma, K: Keratinocyte-like, E: EpSC-like, HM: Hervás-Marín *et al.* 2019, AL: AL Eitan *et al.* 2020.

### 3.1.5. EpSC-like cSCC tumors display a higher metastatic potential.

Two approaches were followed to assess whether the predicted more invasive phenotype in the EpSC-like cSCC tumors would correlate with a higher metastatic potential. Both methods were based on the analysis of tumor material isolated from formalin-fixed paraffin-embedded (FFPE) samples. Thus, the correlation between DNA methylation data obtained from fresh-frozen (FF) and FFPE material was first assessed. A correlation of 90-98% was observed when comparing the methylomes obtained from isolated epidermal tumor regions from the same tumor in matched FF and FFPE samples (Figure S2), thus confirming the feasibility of using FFPE material in the hereunder described methylation analyses.

#### 3.1.5.1. Stratification of KC using amplicon-based targeted sequencing

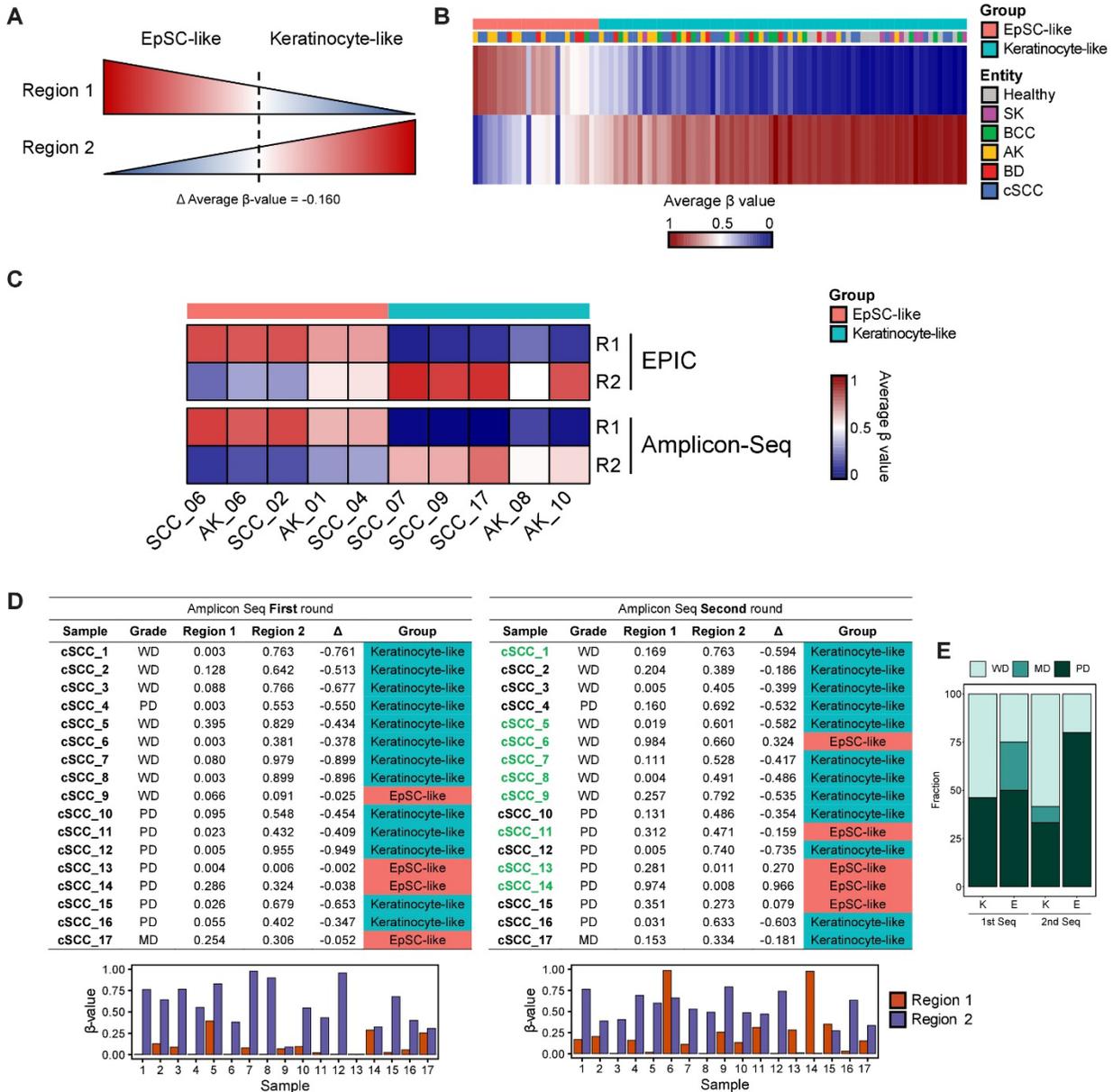
The first approach was based on the targeted amplicon-based bisulfite sequencing of two selected regions with opposed methylation patterns in the two distinct cSCC cell-of-origin subclasses (Figure 13A and 13B). These regions were identified using the complete KC dataset

and a numeric threshold for tumor stratification was calculated (analyses performed by Dr. Günter Raddatz, Division of Epigenetics, DKFZ). Using a targeted amplicon strategy to stratify tumors would allow for the systematic analysis of a higher number of tumors (up to 48 per sequencing round) and would be more cost-effective than using methylation arrays. Thus, this method was developed with the intention to analyze a large cohort of FFPE samples with available clinical histories to assess, for instance, survival, metastasis-free time, or recurrence rates for each cell-of-origin subclass.

To do so, the two regions were first tested on ten FF samples that were previously analyzed by Infinium MethylationEPIC BeadChips. Both regions displayed comparable  $\beta$ -values in both techniques, thus validating the targeted amplicon-based stratification strategy on FF samples (Figure 13C and Figure S3). As a first trial using FFPE samples, sections from a set of 17 primary cSCCs were obtained, and their epidermal regions were manually isolated (carried out by Dr. Daniel Hasche, Division of Viral Transformation Mechanisms, DKFZ). These were selected based on their degree of differentiation determined by histopathological features, thus including nine well or moderately differentiated and eight undifferentiated tumors. Tumor stratification resulted in 13 cSCC classified as Keratinocyte-like and four as EpSC-like (Figure 13D). Tumor grade and cell-of-origin stratification did not show a strong correlation as both cell-of-origin subclasses comprised approximately 50% well and moderately differentiated tumors and 50% poorly differentiated ones (Figure 13E). Importantly, while region 2 displayed the expected dynamic range of  $\beta$ -values, the methylation values detected for region 1 were found to be low in all samples (Figure 13E). Thus, the experiment was repeated by generating new amplicons from the same tumors. In some cases, the same DNA as in the first trial was used, while for some other samples, DNA was extracted from additional sections. The second sequencing round for these samples yielded a better dynamic range of  $\beta$ -values in region 1 and classified 12 out of 17 tumors in the same manner as the first sequencing (Figure 13E).

Also, an enrichment for well and moderately differentiated tumors was observed in the Keratinocyte-like subclass while an enrichment for poorly differentiated cSCC was found in EpSC-like tumors (Figure 13E). Nonetheless, five samples were stratified in a different cell-of-origin group, including tumors for which the same initial DNA was used to generate the amplicons (Figure 13D). In addition, absolute  $\beta$ -values were substantially different in some samples between sequencing rounds (i.e., cSCC\_2, cSCC\_3, or cSCC\_7) (Figure 13D). These results indicated a significant deviation between experiments using FFPE samples, thus impairing proper tumor stratification.

### 3. Results

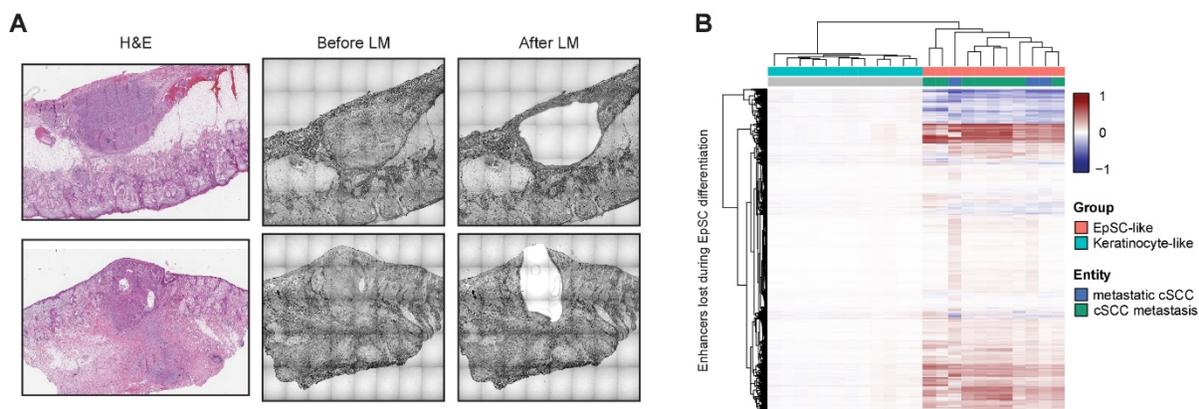


**Figure 13. Amplicon-sequencing approach for cSCC stratification.** (A) Schematic model showing the identified regions with opposed methylation values in cell-of-origin subclasses and the calculated threshold for stratification. (B) Heatmap displaying the methylation values for both regions in the KC cohort. (C) Heatmap comparing the average methylation values obtained by EPIC arrays or amplicon-sequencing for the two regions in 10 FF samples. (D) Upper: tables depicting the average  $\beta$ -values in 17 FFPE cSCC tumors in the two rounds of amplicon-sequencing. Tumor stratification based on these values is also depicted. In green are marked tumors for which DNA was extracted from additional sections for the second experiment. Lower: Bar plots depicting the average  $\beta$ - values for both regions in the 17 FFPE cSCC samples in the two sequencing rounds. (E) Bar plots showing the fraction of well, moderately and poorly differentiated cSCCs corresponding to each tumor subclass in both sequencing rounds. SK: seborrheic keratosis, BCC: basal cell carcinoma, AK: actinic keratosis, BD: Bowen's disease, cSCC: cutaneous squamous cell carcinoma, WD: well-differentiated, MD: moderately differentiated, PD: poorly differentiated.

Overall, the targeted amplicon-sequencing approach was robust and applicable to FF samples, but further optimization would be required to apply it to FFPE samples. However, this approach was not further investigated in the scope of this work.

### 3.1.5.2. Methylation profiling of metastatic cSCC

As an alternative approach to assess the potential higher metastatic potential of EpSC-like cSCC tumors, the methylomes of four primary metastasizing cSCCs and eight cSCC metastases were profiled using Infinium MethylationEPIC arrays. These samples were obtained as FFPE sections, and tumor tissue was isolated using laser microdissection, thus avoiding normal epidermis or dermal tissue contamination in the samples (Figure 14A). Tumor stratification was performed using the set of enhancers lost upon epidermal differentiation as abovementioned. All the 11 samples included in the analysis were stratified as EpSC-like tumors (Figure 14B), thus suggesting a higher metastatic potential of this particular cell-of-origin subclass and highlighting a prognostic potential for the DNA methylation-based stratification strategy.

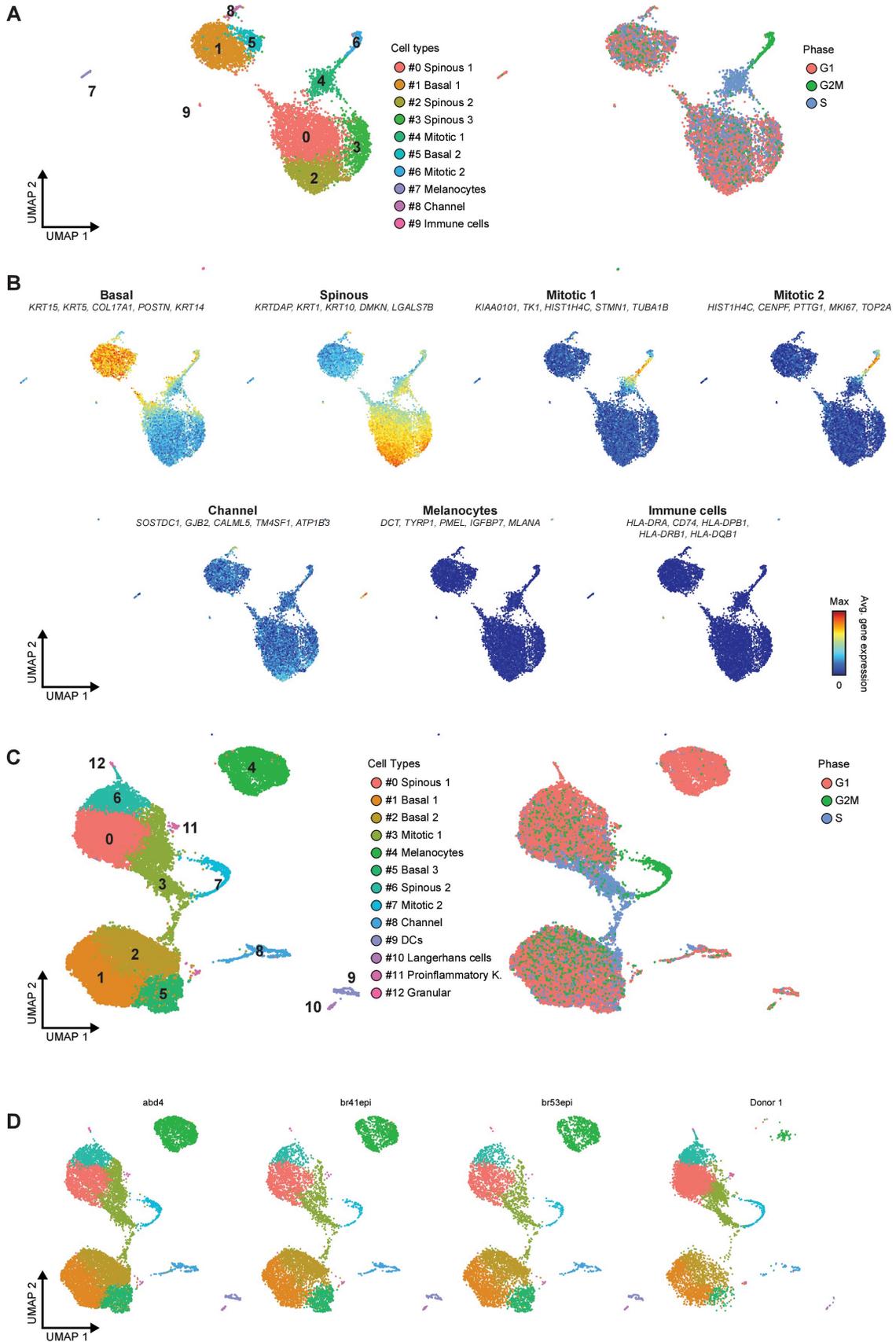


**Figure 14. Stratification of primary metastasizing cSCC and cSCC metastases. (A)** Illustrative examples of the laser microdissection approach used. For each example, a hematoxylin and eosin staining is shown (left), as well as an image of the skin section before (middle) and after (right) laser microdissection. **(B)** Heatmap displaying the unsupervised hierarchical clustering of eight cSCC metastases, three primary metastatic cSCC, and 12 healthy controls based on the methylation patterns of enhancers lost upon epidermal differentiation (n= 10,697 enhancers). H&E: hematoxylin and eosin, LM: laser microdissection, cSCC: cutaneous squamous cell carcinoma.

### 3.1.6. Single-cell transcriptomes specify KC cell-of-origin subclasses

To further investigate keratinocyte differentiation and link it to the potential cells-of-origin of KC, scRNA-seq was performed on sun-protected healthy epidermis from the ilioinguinal region of a 30 y/o male donor. Using the 10X Genomics platform, a dataset containing 7,143 cells after quality control was generated and visualized in a uniform manifold approximation and projection (UMAP) plot (Figure 15A). Unsupervised clustering using the Seurat R package (Stuart *et al.*, 2019) identified ten cell populations, including mostly keratinocytes at different stages of differentiation but also melanocytes, and immune cells (Figure 15A and 15B). To gain statistical power, the in-house dataset was subsequently combined with a previously published dataset consisting of sun-protected healthy epidermis from the abdominal region of three donors (Cheng *et al.*, 2018). After integration and removal of low-quality cells (see Methods, section 6.2.6), the dataset contained a total of 32,272 cells. Unsupervised clustering defined 13 cell clusters, which included cells from all donors, thus highlighting the lack of batch effects after data integration (Figure 15C and 15D). Nevertheless, due to the lack of a significant number of differentially expressed genes, some clusters were merged as they represented well-known epidermal cell populations (Figure 16A).

Cell identity was established by comparing known markers with the most representative expressed genes of each cluster, identifying keratinocyte populations at distinct differentiation states. Thus, two clusters represented basal undifferentiated keratinocytes and were characterized by the expression of well-known markers such as *KRT15*, *KRT5*, *COL17A1*, *POSTN*, and *KRT14* (Moll, Divo and Langbein, 2008) (Figure 16B). However, the expression of *KRT19*, *CYR61*, *CCL2*, *GBP1*, and *KRT17* clearly distinguished them, indicating that they indeed represent distinct undifferentiated populations, which were labeled as Basal 1 and Basal 2 (Figure 16A and 16B). Based on their cell cycle phase, two distinct types of mitotic keratinocytes were identified (Figure 15C and 16B). Differentiated suprabasal keratinocytes were represented by two populations: one characterized by the expression of typical markers from the spinous layer including *KRTDAP*, *KRT1*, *KRT10*, *DMKN*, and *LGALS7B* (Edqvist *et al.*, 2015); and another expressing genes associated with terminally-differentiated keratinocytes from the granular layer such as *LOR*, *LCE1C*, *CST6*, *FLG* and *LCE1B* (Edqvist *et al.*, 2015) (Figure 15B). Lastly, the pro-inflammatory *S100*-expressing keratinocytes and *channel* keratinocytes, a population of cells highly expressing genes related to ion and mitochondrial channels, were also identified as described in Cheng *et al.* 2018 analysis (Figure 15B).

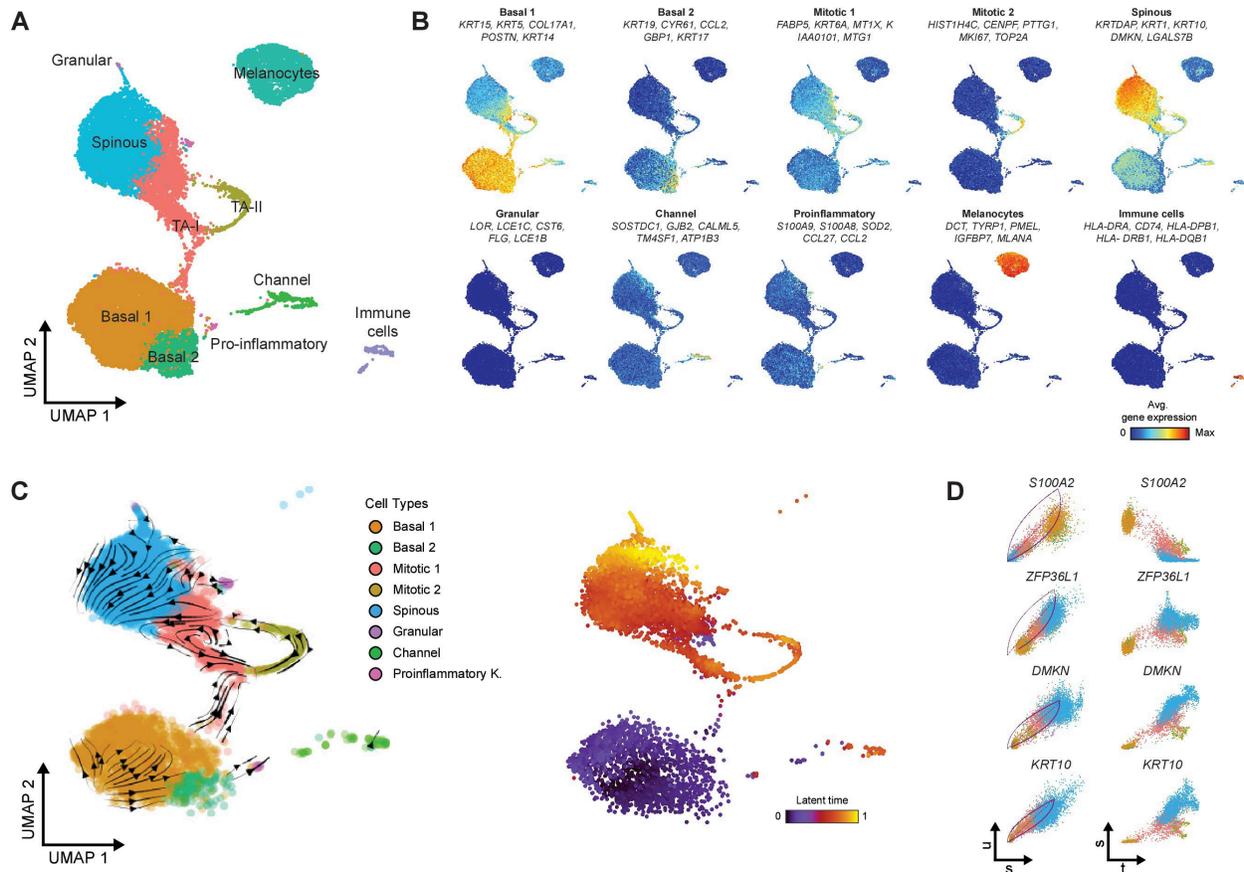


**Figure 15. Single-cell RNA sequencing analysis of sun-protected healthy human epidermis.** (A) UMAP plot displaying 7,143 single-cell transcriptomes from the epidermis of a 30 y/o male donor. Each dot represents a cell and coloring is according to unsupervised clustering (left) or cell cycle phase (right). (B) UMAP plots displaying the average expression of particular sets of cell type-specific marker genes. Red indicates maximum average gene expression while blue indicates low or no expression in log-normalized UMI counts. (C) UMAP plot displaying 32,272 single-cell transcriptomes from four human epidermal samples, after data integration. Each dot represents a cell and coloring is according to unsupervised clustering (left) or cell cycle phase (right). (D) UMAP plots displaying the contribution of each sample to the integrated dataset. Samples abd4, br41epi and br53epi correspond to the dataset published in Cheng *et al.*, 2018 while Donor 1 corresponds to the in-house dataset.

---

Lineage trajectory analysis on single-cell data can be performed using the dynamics between spliced and unspliced mRNAs, known as RNA velocity (Bergen *et al.*, 2020). Thus, RNA velocity analysis performed on the in-house dataset placed the Basal 1 population at the start of the epidermal differentiation, while no contribution to the lineage was found for Basal 2 keratinocytes, thus suggesting that Basal 1 contains the EpSC population (Figure 16C). The lineage then progressed towards the mitotic keratinocytes and finally to the spinous cluster (Figure 16C). Surprisingly, granular keratinocytes did not represent the endpoint of the RNA velocity analysis, which could be caused by the limited numbers of granular cells present in our study or their altered transcriptional program (Figure 16C). Genes with known roles in epidermal differentiation represented the top drivers, thus highlighting the robustness of the analysis (Figure 16D). (RNA analysis performed by Oliver Gilliam, Division of Epigenetics, DKFZ).

In homeostatic conditions, active gene expression is associated with unmethylated promoters, while methylated promoters are usually associated with gene silencing. Based on this assumption, the recently developed EpiSCORE algorithm uses scRNA-seq data to generate a reference matrix of cell type-specific gene markers with coupled expression and unmethylated promoters, which is then used to deconvolute cell type fractions in bulk DNA methylation datasets (Teschendorff *et al.*, 2020). Hence, this tool was used to further explore the keratinocyte composition in the defined cell-of-origin subclasses. Importantly, only keratinocyte populations involved in the lineage trajectory according to the RNA velocity analysis were included. The gene expression and DNA methylation reference matrices were generated as previously described (see Methods, section 6.2.7), resulting in a final matrix containing 74 genes (Figure 17A). This matrix was subsequently used for cell fraction estimation in the bulk methylomes of epidermal tumors and healthy controls, which revealed a highly similar cell composition in Keratinocyte-like tumors and healthy epidermis (Figure 17B). Both showed higher proportions of well-differentiated spinous and granular keratinocytes (Figure 17B).



**Figure 16. Keratinocyte differentiation trajectory inferred by RNA velocity. (A)** UMAP plot displaying 32,272 single-cell transcriptomes from four human epidermis sample, after merging similar clusters. **(B)** UMAP plots displaying the average expression of particular sets of cell type-specific marker genes. Red indicates maximum average gene expression while blue indicates low or no expression in log-normalized UMI counts. **(C)** Left: UMAP plot displaying the 7,068 keratinocytes from the in-house generated dataset used to estimate RNA velocities, which are projected on top (arrows). Right: UMAP plot displaying the latent time, a time approximation of the differentiation process. **(D)** Top four putative driver genes of the epidermal differentiation, as calculated by RNA velocity analysis. Phase portraits (left) and expression along latent time (right) are shown for each gene. In phase portraits, bold line indicates the estimated splicing dynamics while the dotted line depicts the transcriptional steady state. u: unspliced, s: spliced, t: latent time.

On the contrary, EpSC-like tumors showed an enrichment for the EpSC-containing Basal 1 population (Figure 17B). These findings suggest a distinct keratinocyte composition in both methylation-based KC subclasses, with differential enrichment for populations at the start and end of the keratinocyte differentiation, reflecting their EpSC-like and Keratinocyte-like origin.



### 3.2. Single-cell whole-genome bisulfite sequencing to identify the cells-of-origin of KC.

Although several scWGBS approaches have been published, most are still limited by the recovery of a low number of cells and/or low CpG coverage (Karemaker and Vermeulen, 2018). Nonetheless, in 2018, Mulqueen *et al.* proposed the single-cell combinatorial indexing for methylation analysis (sci-MET) as a highly scalable solution for high-throughput single-cell methylomics analysis (Mulqueen *et al.*, 2018). To further investigate DNA methylation upon keratinocyte differentiation and to identify the specific cells-of-origin that give rise to KC, it was hypothesized that a considerably high amount of cells would be optimal for the analysis. Thus, a working sci-MET protocol for human epidermis was established in collaboration with the single-cell open lab (scOpenLab) at the German Cancer Research Centre (DKFZ; Heidelberg, Germany).

sci-MET is based on a combinatorial indexing strategy, which consists of the cleavage and tagging of the double-stranded DNA by a transposome, a process called tagmentation. Thus, nuclei are tagmented using uniquely barcoded Tn5-based transposomes in combination with two rounds of FACS sorting (Adey *et al.*, 2014; Amini *et al.*, 2014). In brief, nuclei are isolated from fixed cells, depleted of nucleosomes by incubation with detergents, and subsequently sorted into a 96-well plate (up to 2,500 nuclei per well). Next, nuclei are tagmented using 96 unique transposomes containing cytosine-depleted barcodes, so that bisulfite conversion does not affect their sequence. Nuclei are then re-pooled and FACS-sorted into a new 96-well plate. To avoid nuclei tagmented with the same barcode being sorted in the same well, only 22 nuclei are sorted per well. Nuclei are then digested and DNA is bisulfite-converted before several rounds of linear amplification. This is followed by an index PCR, which introduces two additional barcodes to the library molecules together with the i5 and i7 Illumina adaptors. Libraries generated in individual wells are then pooled together, cleaned with solid-phase reversible immobilization (SPRI) beads, and quantified. Single-cell resolution is achieved after sequencing through the demultiplexing of the raw sequencing data based on the triple barcoding of the library molecules.

This chapter summarizes the set of experiments performed in collaboration with the scOpenLab (DKFZ) in order to optimize the sci-MET methodology, starting with an *in vitro* system and following with primary human epidermis. Furthermore, additional experiments performed in order to apply sci-MET to FF human epidermal samples are summarized in the Additional results, section 7.3.

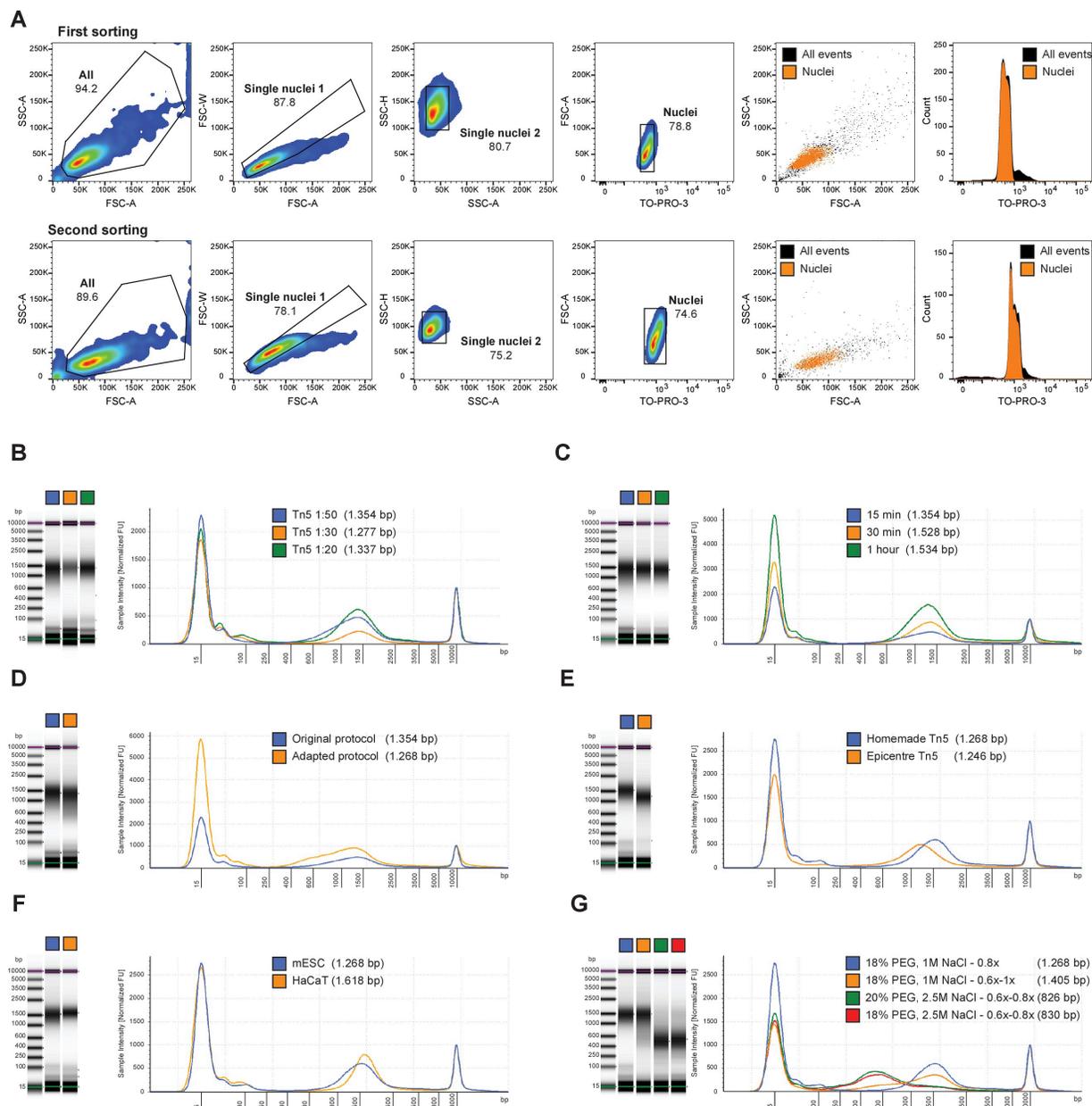
### 3.2.1. Establishment of a working sci-MET protocol

#### 3.2.1.1. Library preparation optimization

To first establish and optimize the sci-MET protocol, *in vitro* mouse embryonic stem cells (mESCs) were used as the experimental model. To make the protocol more cost-effective, a homemade Tn5 transposase was obtained from the Protein Expression and Purification Core Facility at the European Molecular Biology Laboratory (EMBL; Heidelberg, Germany). First, the concentration of Tn5 transposase at which its activity was optimal was assessed (see Methods, section 6.2.8.3). The optimal working dilution was found to be 1:50, which was the one used in all the optimization experiments, unless stated otherwise. All the optimization experiments described here were performed with 10 wells (2,500 nuclei/well) in the first sorting and 16 wells (22 nuclei/well) in the second sorting, after transposition with 10 uniquely barcoded transposomes. To ensure proper nuclei sorting the gating strategy depicted in Figure 18A was followed.

In a first attempt, the sci-MET library preparation protocol was followed as described in the original publication (Mulqueen *et al.*, 2018), and successfully generated a library. Nevertheless, the average fragment size was over 1,000 bp, impairing its proper sequencing (Figure 18B). In tagmentation-based library preparation approaches, fragment size can be influenced by Tn5 activity, quantity, and/or time of transposition (Picelli *et al.*, 2014). Thus, new libraries were generated using transposomes at a higher concentration (dilution 1:30 and 1:20). However, no differences could be observed in the library fragment distribution (Figure 18B). Increasing transposition time from 15 minutes (as in the original publication) to 30 minutes or 1 hour did also not affect fragment size (Figure 18C).

In the original protocol, cells were fixed by incubation with 1.5% formaldehyde (with methanol) for 10 minutes and nuclei were isolated by incubating cells in nuclear isolation buffer (10 mM TrisHCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% IGEPAL® CA-630, 1X protease inhibitors cocktail) for 20 minutes. Since the formaldehyde used contained methanol, another fixative solution, nuclei could be over-fixed, preventing an efficient transposition. Following this hypothesis, a different cell fixation and nuclear isolation approach was tested, adapted from a scOpenLab routinely used protocol. Hence, mESCs were fixed with 0.7% formaldehyde (without methanol) for 10 minutes and nuclei were isolated using Nuclei EZ Lysis Buffer (Sigma-Aldrich) (see Methods, section 6.2.8.1). Library generation using the adapted sample preparation protocol did not substantially reduce the fragment size (Figure 18D). Nonetheless, a broader size range could be observed and, therefore this sample preparation protocol was used for the rest of the experiments.



**Figure 18. sci-MET library preparation protocol optimization using mESCs.** (A) Representative FACS plots displaying the gating strategy for sorting single nuclei from mESCs. (B-G) Bioanalyzer D5000 ScreenTape gel images and electropherograms of sci-MET libraries obtained with different Tn5 concentrations (B), transposition times (C), sample preparation protocols (D), transposases (E), cell types (F), or SPRI beads buffer for size selection (G).

To assess whether the higher fragment size was due to a lower transposition efficiency of the homemade Tn5, a sci-MET library with the commercial transposase used in the original

protocol (Epicentre-Illumina) was produced. The resulting library did not show a significantly different profile from the one generated with the homemade Tn5 (Figure 18E). Since using a commercial Tn5 did not provide more satisfactory results, the following experiments were still carried out using the homemade Tn5.

Different cell types might generate sequencing libraries with different profiles and fragment sizes. Hence, to determine whether the fragment size observed was due to the use of mESCs as an experimental model, a sci-MET library using HaCaT cells, a spontaneously immortalized keratinocyte cell line, was generated. Libraries produced with either cell type did not show substantial differences in fragment size, indicating that the library preparation protocol is robust and that the fragment size was not influenced by the cell type used as input material (Figure 18F).

As with most next-generation sequencing library preparation protocols, sci-MET uses SPRI beads for sample clean-up after linear amplification (1.1X the volume per well) and index PCR (0.8X the volume of the pooled library). SPRI beads can be used in two sequential steps to perform a double-sided size selection, generating libraries with the desired size distribution (Stortchevoi, Kamelamela and Levine, 2020). In addition, changes in the concentration of polyethylene glycol (PEG) and NaCl present in the beads buffer can vary their affinity to distinct sizes of DNA (Stortchevoi, Kamelamela and Levine, 2020). In these experiments, AMPureXP beads (Beckman Coulter) were used for SPRI-based sample clean-up. However, following the published sci-MET protocol, the original beads buffer containing 20% PEG 8000 and 2.5 M NaCl was substituted for one containing 18% PEG 8000 and 1 M NaCl (Mulqueen *et al.*, 2018). Since none of the previously tested conditions yielded a significant change in fragment distribution, it was hypothesized that the changes in PEG and NaCl concentration could impair proper size selection.

Thus, a sci-MET experiment was performed in which the clean-up and size selection steps were carried out using AMPureXP beads (Beckman Coulter) with the buffer from the published sci-MET protocol (18% PEG and 1 M NaCl), the original AMPureXP beads buffer (20% PEG and 2.5 M NaCl) or a buffer containing 18% PEG and 2.5 M NaCl. Furthermore, a double-sided size selection was tested for the last library clean-up step, after the library material for all wells is combined. No changes in the size distribution were found when using a double-sided size selection step with the original sci-MET buffer containing 18% PEG and 1 M NaCl (Figure 18G). Nonetheless, double-sided size selection using the buffers containing 2.5 M NaCl displayed a correct size distribution for sequencing, indicating that the concentration of NaCl is crucial for proper size selection (Figure 18G).

### 3.2.1.2. Sequencing protocol optimization

A series of important inconsistencies found in the original publication and the omission of crucial methodological information regarding the sequencing protocol precluded the reproduction of the technique thus far. For instance, the sequencing primers contain locked nucleic acid (LNA) bases, which are modified RNA monomers whose conformation increases the affinity for complementary sequences (Braasch and Corey, 2001). Thus, such modifications are essential for proper sequencing. However, their location in the primer sequences was omitted in the publication. Also, the Index1 (i7) primer sequence is never mentioned in their article while Index2 (i5) is constantly referred to as Index1 in the materials and methods section. Due to the transposition step, all DNA molecules in the library contain a common sequence that needs to be omitted during sequencing to avoid failure. This can be achieved by the introduction of dark cycles, in which the Illumina sequencer does not record the fluorescent signal. In the original publication, they mistakenly state that they use 27 dark cycles. Instead, the actual number of dark cycles needs to be only 16. All inconsistencies and mistakes regarding the sequencing protocol have been summarized in Table 2 and the primer sequences are described in the materials and methods section (see Table M9).

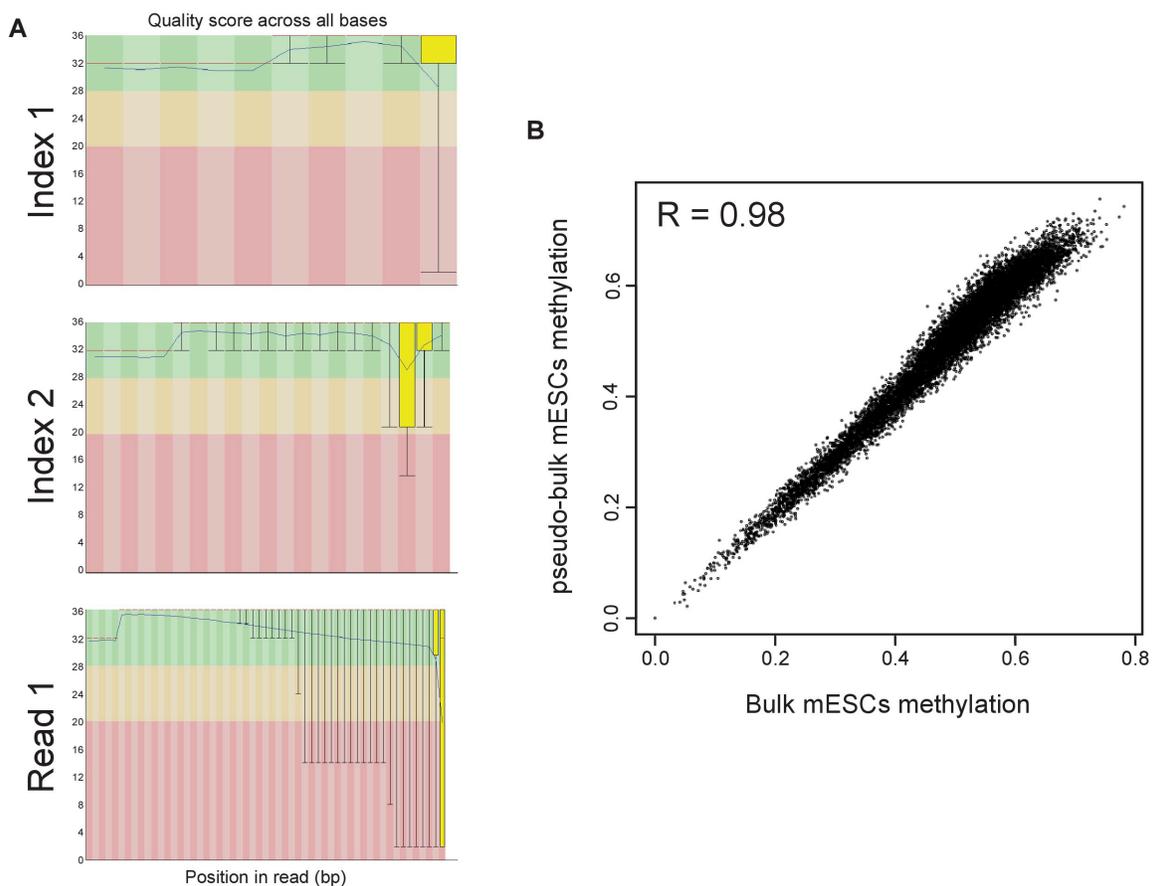
	Original protocol	Working protocol
sci-MET_Read1	100	100
sci-MET_Index1	10 + 27* + 11	10
sci-MET_Index2	10	11 + 16* + 10

\* *Dark Cycles*

**Table 2. Summary of the sequencing parameters for sci-MET libraries.** The table depicts the number of bases sequenced with each primer in the original publication and the corrected protocol.

A more complex library containing 704 mESCs (32 wells x 22 nuclei/well) was generated and sequenced using the described parameters, resulting in high-quality sequencing of the Read1 and both Index1 and Index2 (Figure 19A), and an alignment rate to the mouse genome of 80%. A total of 384 cells were detected, from which 286 displayed more than 30,000 reads and were kept for downstream analyses. To assess whether sci-MET recapitulates DNA methylation patterns found in mESCs, data from all individual cells were pooled to create a pseudo-bulk dataset, which was then compared to a bulk whole-genome bisulfite sequencing dataset generated from the same batch of cells (Figure 19B). For this comparison, only those CpGs with a coverage  $\geq 3$  in both datasets ( $n=10,867,771$ ) were used. The average methylation ratio of

these CpGs was found to be highly similar as they displayed  $\beta$ -values of 0.486 and 0.491 for pseudo-bulk and bulk experiments, respectively. Of those CpGs, only 38,613 (0.36%) were found to be differentially methylated using a Fisher's exact test with Benjamini-Hochberg correction for multiple testing. Lastly, the methylation data of both datasets was smoothed by averaging over 1,000 neighboring CpGs. By doing so, Pearson's correlation coefficient of 0.98 was achieved (Figure 19B). These findings indicated that DNA methylation patterns from bulk experiments could be recapitulated using the sci-MET protocol, thus validating the efficacy of this approach. These analyses were performed by Dr. Günter Raddatz (Division of Epigenetics, DKFZ).



**Figure 19. sci-MET recapitulates methylation patterns obtained in bulk experiments. (A)** Boxplots displaying the per-base sequence quality of Index1, Index2, and Read1 after sequencing with corrected parameters, obtained with FastQC. Green indicates good quality, orange indicates medium quality and red indicates bad quality. **(B)** Scatter plot comparing the smoothed CpG methylation values of CpGs with coverage  $\geq 3$  ( $n=10,867,771$ ) in the sci-MET and bulk WGBS datasets obtained from mESCs.

### 3.2.2. sci-MET protocol optimization for primary human epidermis

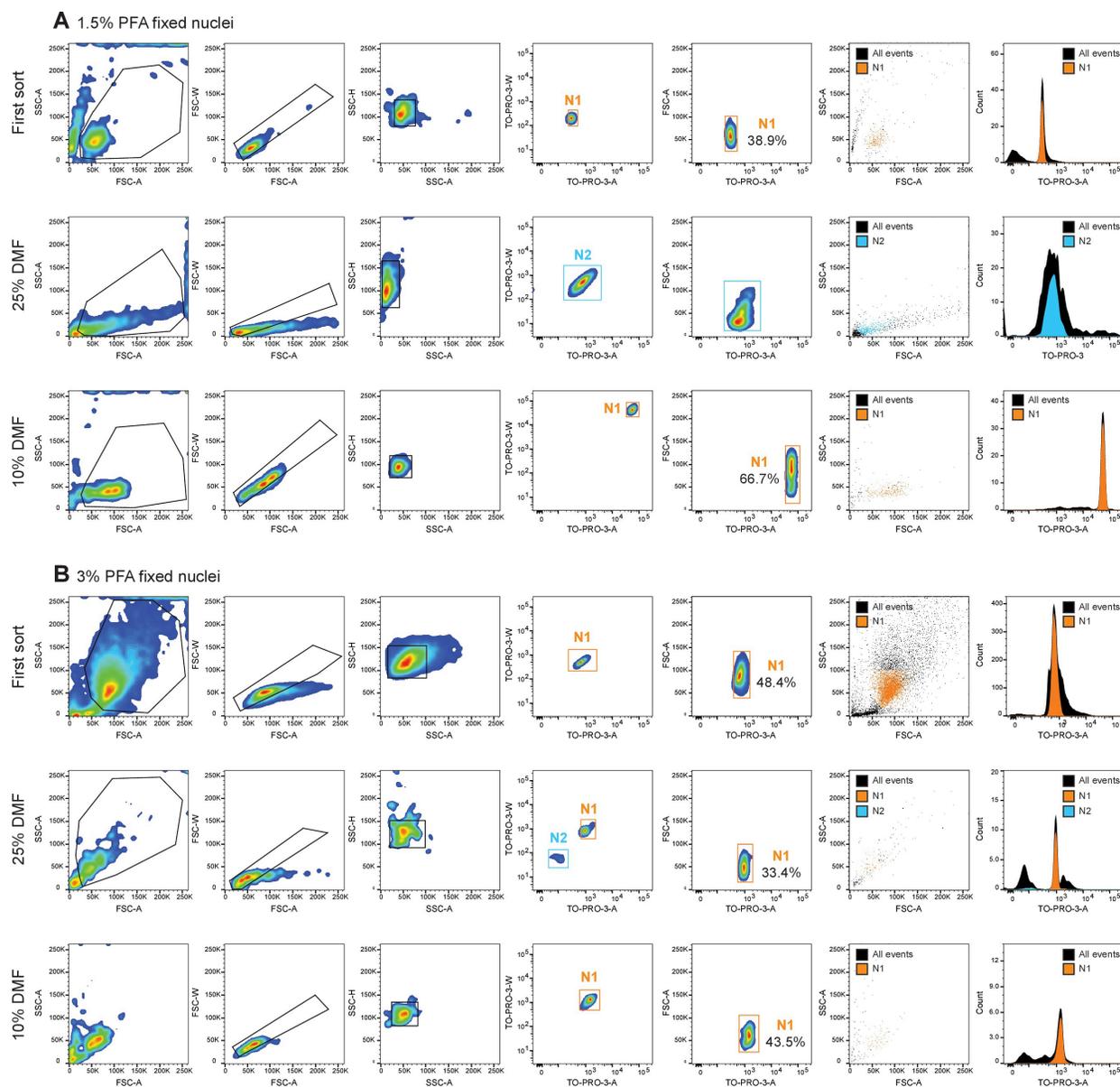
To optimize the sci-MET protocol for human epidermis, fresh whole human skin samples from the axillary or ilioinguinal region were used. In all the experiments described hereunder, the epidermis was isolated from the dermis and a single-cell suspension was subsequently obtained using the Miltenyi Biotec dissociation kit for human epidermis, following the manufacturer's instructions.

In a first attempt, epidermal cells were fixed and nuclei isolated as previously described for mESCs. However, cell sorting was strongly impaired as there was a reduced amount of nuclei to perform the first sorting step as well as almost no nuclei could be sorted when re-pooled after transposition for the second sorting step (data not shown).

N,N-Dimethylformamide (DMF) is an essential component for Tn5 transposition, acting as a crowding agent (Picelli *et al.*, 2014; Sos *et al.*, 2016). The tagmentation buffer used with the homemade Tn5 in the mESCs experiments consisted of 10 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, and 25% DMF. This DMF concentration was much higher than the one used in other tagmentation protocols for bulk or single-cell approaches (Picelli *et al.*, 2014; Sos *et al.*, 2016; Corces *et al.*, 2017; Chen *et al.*, 2018), and thus it was hypothesized that it could be detrimental to nuclei integrity. Also, increasing the concentration of formaldehyde during cell fixation could increase nuclei resistance to the sample preparation.

To test these hypotheses, human epidermal cells were isolated and fixed with 1.5%, 3%, and 5% formaldehyde. Nonetheless, the nuclei fixed at 5% could not be sorted due to the high aggregation of nuclei, therefore this condition was disregarded. Nuclei fixed with 1.5% or 3% formaldehyde were then sorted into few wells of a 96-well plate and subsequently tagmented using tagmentation buffer containing either 10% or 25% DMF. Nuclei fixed with 1.5% formaldehyde and transposed with 25% DMF could not be properly separated from cellular debris in the second sorting step and the only population that could be gated was negative for the nuclear marker TO-PRO™-3 Iodide (Figure 20A). Library preparation using this population yielded no results, thus confirming that this population does not contain nuclei (data not shown). In contrast, nuclear tagmentation with 10% DMF increased the number of available nuclei in the second sorting, comprising up to 66.7% of the sample (Figure 20A). In the case of nuclei fixed with 3% formaldehyde, both tagmentation conditions displayed an increased amount of nuclei after transposition, with the 10% DMF condition also showing a higher number (Figure 20B). Taken together, these results suggested that the best conditions for human epidermis sample preparation were 1.5% formaldehyde for cell fixation and 10% DMF for tagmentation.

### 3. Results



**Figure 20. Optimization of human epidermis sample preparation for sci-MET. (A)** Representative FACS plots displaying the gating strategy followed for sorting single nuclei from human epidermis fixed with 1.5% formaldehyde. Plots for the first sorting step (upper) or the second sorting step after tagmentation with a buffer containing 25% DMF (middle) or 10% DMF (lower) are shown. **(B)** Representative FACS plots displaying the gating strategy followed for sorting single nuclei from human epidermis fixed with 3% formaldehyde. Plots for the first sorting step (upper) or the second sorting step after tagmentation with a buffer containing 25% DMF (middle) or 10% DMF (lower) are shown.

To further improve the nuclear isolation from human epidermis cell suspensions and thus the first sorting step of the sci-MET protocol, three alternative protocols which mostly differed in the composition of the lysis buffer were tested (Wysocka, Reilly and Herr, 2001; Buro, Shah and Henriksen, 2010; Neely and Bao, 2019) (see Supplementary Methods, section 7.4.1).

The Wysocka *et al.* protocol showed the higher recovery of nuclei and thus was used for subsequent experiments (data not shown).

The working sci-MET protocols for cell lines and primary human epidermis are described in detail in the Methods chapter, section 6.2.8.

### 3.2.3. Single-cell methylomics identifies the KC cells-of-origin

Once the sample preparation conditions were optimized for fresh human epidermis samples, a sci-MET experiment was performed with a sample from the sun-protected ilioinguinal region of a male donor of 62 y/o. Thus, a single-cell suspension was obtained using Miltenyi Biotec kit for human epidermis, cells were fixed in 1.5% formaldehyde, and nuclei were subsequently isolated. Next, a full 96-well plate was sorted (1,000 nuclei/well) followed by tagmentation with the buffer containing 10% DMF. In addition, PitStop2 (Sigma-Aldrich), a molecule that disrupts the nuclear pore complex without disrupting the nuclear membrane, was shown to increase Tn5 entry to the nuclei and produce libraries with higher complexity in sci-ATAC-seq (Mulqueen *et al.*, 2019). Thus, PitStop2 was added to the tagmentation buffer at a final concentration of 70  $\mu$ M, as suggested for the sci-ATAC-seq protocol. Following tagmentation, a full 96-well plate was sorted and sci-MET library preparation was performed as previously described.

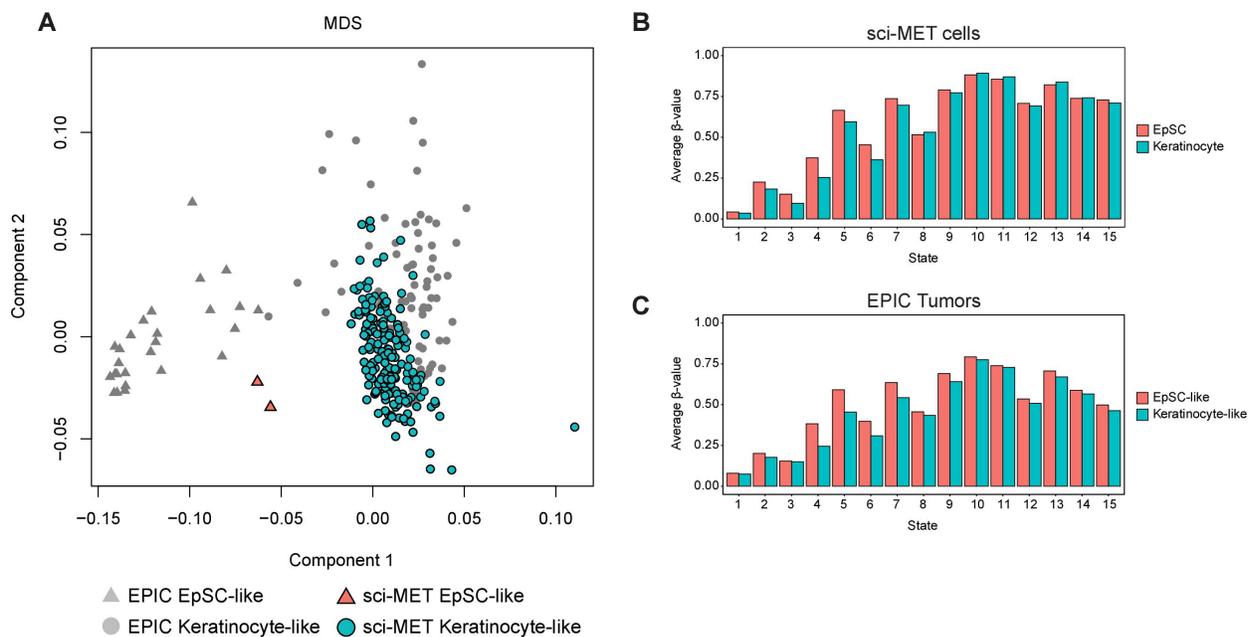
A library containing 186 cells was generated and successfully sequenced. From those 186 cells, the barcodes for 185 cells were detected, all of which displayed a CpG coverage of more than 100,000 CpGs (0.79% in average) and were kept for downstream analyses. The CpG coverage per cell obtained was lower than other approaches but consistent with the original sci-MET protocol (Karemaker and Vermeulen, 2018; Mulqueen *et al.*, 2018). These cells were combined with the 102 bulk methylomes from epidermal tumors and healthy samples (analyzed in the section 3.1) in a Multidimensional Scaling (MDS). From the 185 single cells, 183 clustered with to the Keratinocyte-like tumors and healthy samples, whereas the other two clustered with the EpSC-like tumors (Figure 21A). These findings suggested that these two groups of cells represented differentiated keratinocytes and EpSCs, respectively.

To address this notion, methylation levels were compared between both single-cell clusters at different genomic regions that were previously annotated using comprehensive epigenomic datasets from normal human epidermal keratinocytes (NHEK)(Ernst *et al.*, 2011). Low methylation levels were found at promoters that were designated as active in NHEK in both cell clusters with concomitant high methylation levels at genomic regions assigned as actively transcribed, thus in agreement with a keratinocyte identity (Figure 21B). In addition, high methylation levels were found in repressed, heterochromatic and repetitive regions in both cell clusters as previously

### 3. Results

described for such genomic regions (Deplus *et al.*, 2014) (Figure 21B). Enhancer regions, especially strong enhancers, displayed reduced methylation levels in the cells clustering with the Keratinocyte-like tumors and healthy epidermis (Figure 21C). A loss of DNA methylation has been previously described in lineage-specific regulatory elements upon epidermal differentiation in both mice (Bock *et al.*, 2012) and humans (Sen *et al.*, 2010). Thus, these results support the notion that the two single-cell clusters represent EpSCs and differentiated keratinocytes, which is also illustrated by representative differentially methylated regions (Figure S4).

KC entities displayed highly similar methylation patterns at the different genomic regions compared to the single-cell profiles, with Keratinocyte-like tumors showing substantially reduced methylation levels at enhancer regions (Figure 5C). Collectively, these findings provide an important confirmation for the cell-of-origin interpretation of the methylation-based KC subclasses described in the section 3.1 of this thesis. These analyses were performed by Dr. Günter Raddatz (Division of Epigenetics, DKFZ).



**Figure 21. sci-MET analysis from human epidermis validates the cell-of-origin-based stratification of KC. (A)**

MDS analysis including the 102 bulk DNA methylomes from epidermal tumors and healthy epidermal samples analyzed in section 3.1, and the 185 single-cell epidermal methylomes obtained with sci-MET. **(B, C)** Average  $\beta$ -value at the 15 genomic states defined by ChromHMM in NHEK for **(B)** both EpSC and differentiated keratinocytes from the sci-MET dataset or **(C)** EpSC-like and Keratinocyte-like tumors from the EPIC dataset. ChromHMM state 1: active promoter; state 2: weak promoter; state 3: inactive/poised promoter; state 4: strong enhancer; state 5: strong enhancer; state 6: weak/poised enhancer; state 7: weak/poised enhancer; state 8: insulator; state 9: transcriptional transition; state 10: transcriptional elongation; state 11: weak transcription; state 12: polycomb-repressed; state 13: heterochromatin/low signal; state 14: repetitive/copy number variation; state 15: repetitive/copy number variation.

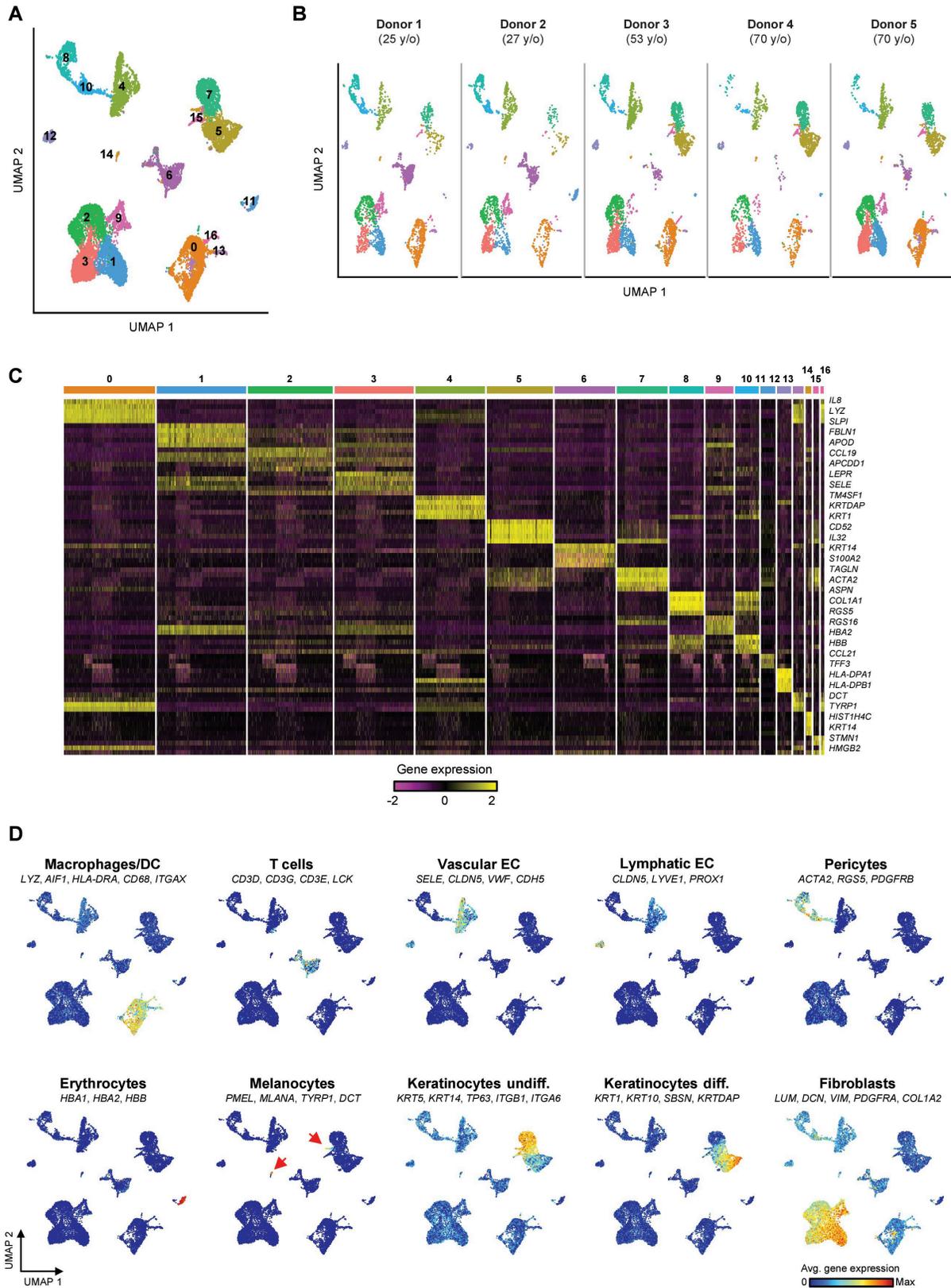
### 3.3. Single-cell transcriptomes of the aging human skin reveal loss of fibroblast priming

Single-cell analyses proved to be of high importance for defining the cell-of-origin of KC and were also applied in this thesis to study dermal fibroblasts at such detailed resolution. Thus, to analyze cellular heterogeneity in human dermal fibroblasts, scRNA-seq of healthy whole skin was performed from five male donors using the 10X Genomics platform. To also investigate the effects of aging in this particular cell type, the samples were obtained from two young (25 and 27 y/o) and three older (53, 69, and 70 y/o) donors. Skin aging is not only driven by intrinsic chronological factors but is also highly influenced by the effects of extrinsic factors, mainly chronic exposure to UVR (Rittié and Fisher, 2015; Krutmann *et al.*, 2017). Hence, to specifically address the effects of intrinsic aging and avoid the confounding effects of photoaging, all samples were obtained from the sun-protected ilioinguinal region. This resulted in a dataset containing 15,457 cells after quality control, performed with the Seurat R package (Stuart *et al.*, 2019).

#### 3.3.1. Identification of the main cell populations present in the human skin

To obtain an overview of the distinct cell populations present in the human skin, the cells from all five donors were combined using the standard integration protocol from the Seurat R package (see Methods, section 6.2.9) (Stuart *et al.*, 2019). Unsupervised clustering identified 17 cell clusters, which were visualized in a UMAP plot (Figure 22A). All clusters contained cells from all donors, thus, indicating the absence of batch effects in the analysis (Figure 22B). To assess the identity of each cell cluster, their most representative genes were compared with known markers of human skin cell types, which revealed that the 17 cell clusters represented nine main cell types. Epidermal keratinocytes and dermal fibroblasts comprise the two main cell types of the skin and accounted for seven cell clusters (Figure 22C and 22D). Keratinocytes represented three clusters (#5, #7, and #15) and reflected distinct stages of differentiation. EpSCs and other undifferentiated progenitors (clusters #7 and #15) expressed marker genes such as *KRT5*, *KRT14*, *TP63*, *ITGA6*, and *ITGB1* (Cheng *et al.*, 2018). On the other hand, differentiated keratinocytes (cluster #7) expressed *KRT1*, *KRT10*, *SBSN*, and *KRTDAP* (Cheng *et al.*, 2018) (Figure 22D). Dermal fibroblasts represented not only the most abundant cell type in the dataset (5,948 cells) but also the most heterogeneous as the expression of their well-known markers *LUM*, *DCN*, *VIM*, *PDGFRA*, and *COL1A2* (Philippeos *et al.*, 2018) was observed in four clusters (#1, #2, #3, and #9; Figure 22D).

### 3. Results

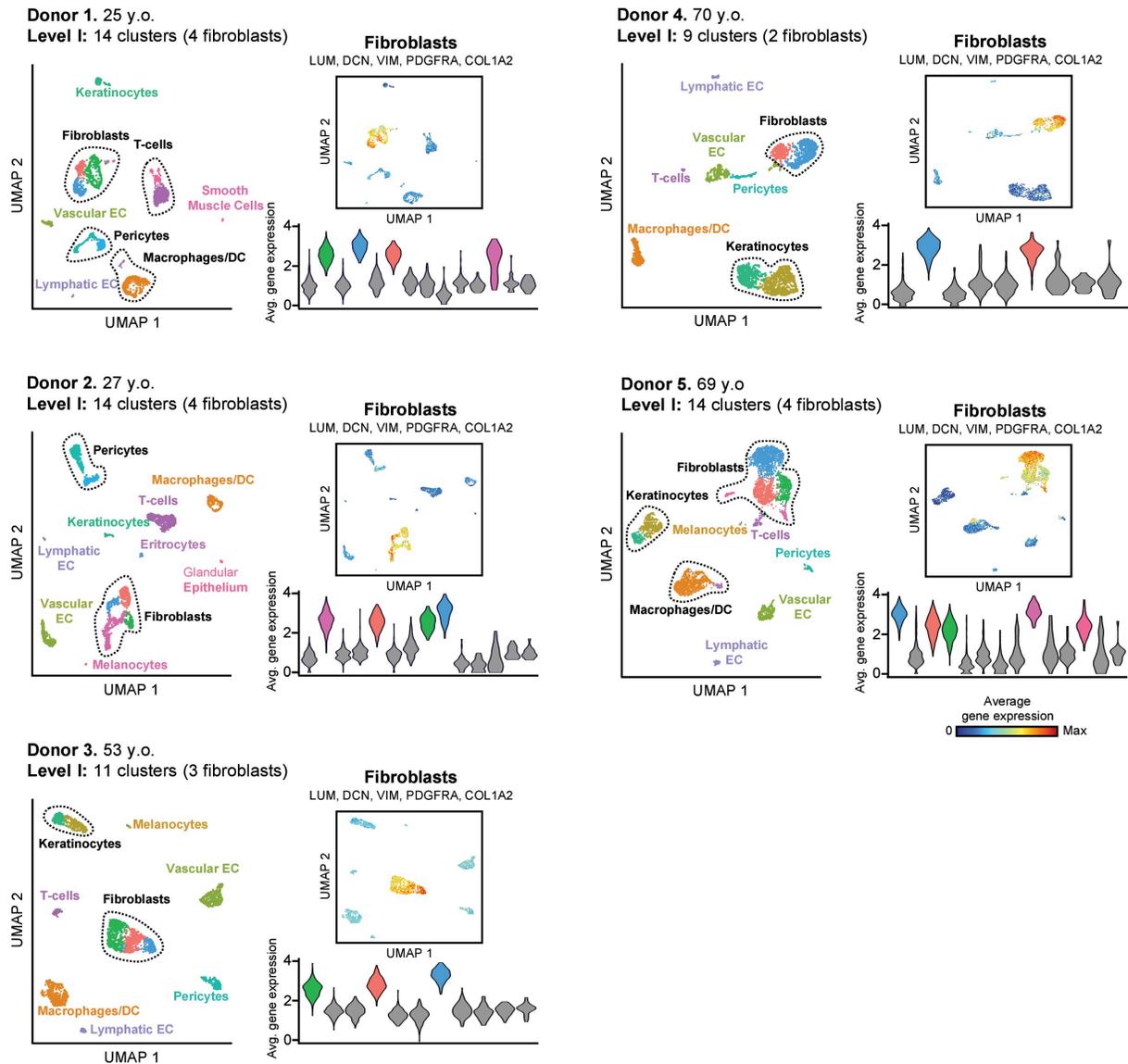


**Figure 22. Single-cell transcriptomics of sun-protected human skin identifies nine main cell types. (A)** UMAP plot displaying 15,457 single-cell transcriptomes from five human skin samples. Each dot represents a cell and coloring is according to unsupervised clustering. **(B)** UMAP plots displaying the contribution of each sample to the integrated dataset. **(C)** Heatmap showing the expression of the top five marker genes in each cluster. Top two marker genes per cluster are listed on the right. Yellow indicates maximum gene expression while purple indicates no expression in scaled log-normalized UMI counts. **(D)** UMAP plots displaying the average expression of particular sets of cell-type marker genes. Red indicates maximum average gene expression while blue indicates low or no expression in log-normalized UMI counts. DC: dendritic cells, EC: endothelial cells.

The other main cell types included immune cells such as macrophages and dendritic cells (DC), characterized by the expression of *LYZ*, *AIF1*, *HLA-DRA*, *CD68*, and *ITGAX* (Elizondo *et al.*, 2017) (clusters #0, #13, and #16), and T-cells expressing *CD3D*, *CD3G*, *CD3E* and *LCK* (Chtanova *et al.*, 2005)(cluster #6; Figure 22D). Two clusters of endothelial cells with high expression of the pan-endothelial marker *CLDN5* were also detected and could be divided into vascular endothelial cells defined by *SELE*, *VWF*, and *CDH5* expression (Leeuwenberg *et al.*, 1992) (cluster #4), and lymphatic endothelial cells expressing typical markers such as *LYVE1* and *PROX1* (Lee *et al.*, 2015) (cluster #12; Figure 22D). Lastly, two clusters of pericytes expressing *ACTA2*, *RGS5*, and *PDGFRB* (Paquet-Fifield *et al.*, 2009) (clusters #8 and #10), another one of erythrocytes displaying high *HBA1*, *HBA2*, and *HBB* expression (Cohen *et al.*, 2008) (cluster #11), and one of melanocytes expressing *PMEL*, *MLANA*, *TYRP1*, and *DCT* (Cichorek *et al.*, 2013) (cluster #14) were identified (Figure 22D).

Importantly, the individual analysis of each donor revealed a similar number of clusters and identified the same main cell populations (Figure 23). Fibroblast heterogeneity was detected in all donors, with both young donors and the older donor 5 displaying four fibroblast clusters as in the combined analysis (Figure 23). Older donor 5 also presented an additional small fifth cluster. Contrarily, older donors 3 and 4 showed a reduced number of fibroblast populations as they displayed three and two clusters, respectively (Figure 23). Nonetheless, differences in cluster numbers could be explained by the lower complexity of the data in the individual analyses and, thus, should only be interpreted as an additional control indicating that all samples display a similar cellular composition and present dermal fibroblast heterogeneity.

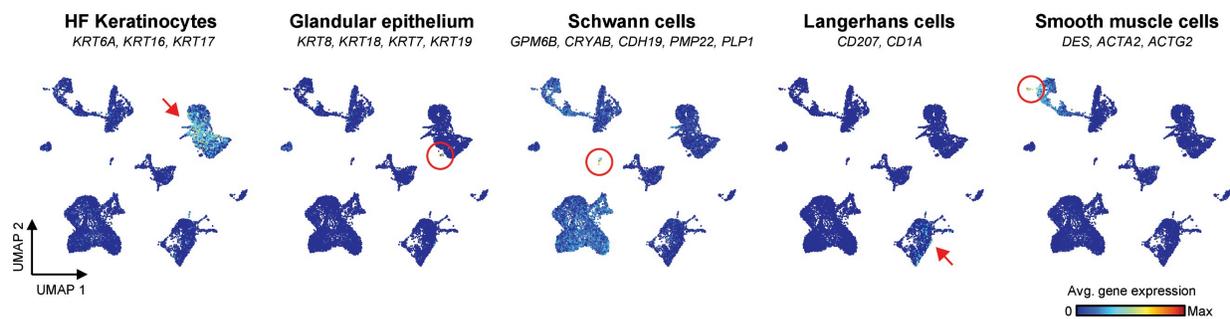
### 3. Results



**Figure 23. Individual analyses display similar cell clustering as the integrated analysis.** Left: UMAP plots displaying single-cell transcriptomes from each of the five donors included in the study. Coloring is according to the unsupervised clustering performed in each sample separately. Right: UMAP and violin plots showing the average gene expression for the set of marker genes used to determine fibroblast identity. In UMAP gene expression projections, red indicates maximum average gene expression while blue indicates low or no expression in log-normalized UMI counts. In violin plots, X-axes depict cell clusters and Y-axes show average gene expression of fibroblast identity genes in log-normalized UMI counts. DC: dendritic cells, EC: endothelial cells.

Some minor skin cell populations clustering with the main cell types were uncovered in an additional analysis (Figure 24). Thus, hair follicle-associated keratinocytes expressing *KRT6A*, *KRT16*, and *KRT17* were grouped with the rest of keratinocytes and a few cells with high

expression of keratin genes typically detected in the glandular epithelium (*KRT8*, *KRT18*, *KRT7*, and *KRT19*) (Moll, Divo and Langbein, 2008) were also identified in cluster #5 (differentiated keratinocytes). Moreover, a small subset of cells in cluster #14 (melanocytes) was found to highly express Schwann cells markers, a type of glial cell present in the dermis (*GPM6B*, *CRYAB*, *CDH19*, *PMP22*, and *PLP1*) (Tabib *et al.*, 2018). Langerhans cells, a type of epidermis-resident macrophages defined by *CD207* and *CD1A* expression, were detected in cluster #4 (macrophages and dendritic cells) (Mizumoto and Takashima, 2004). Lastly, a small subset of smooth muscle cells could be distinguished from cluster #8 (pericytes) by the expression of skeletal muscle gene markers such as *DES*, *ACTA2*, and *ACTG2* (Latif *et al.*, 2015).



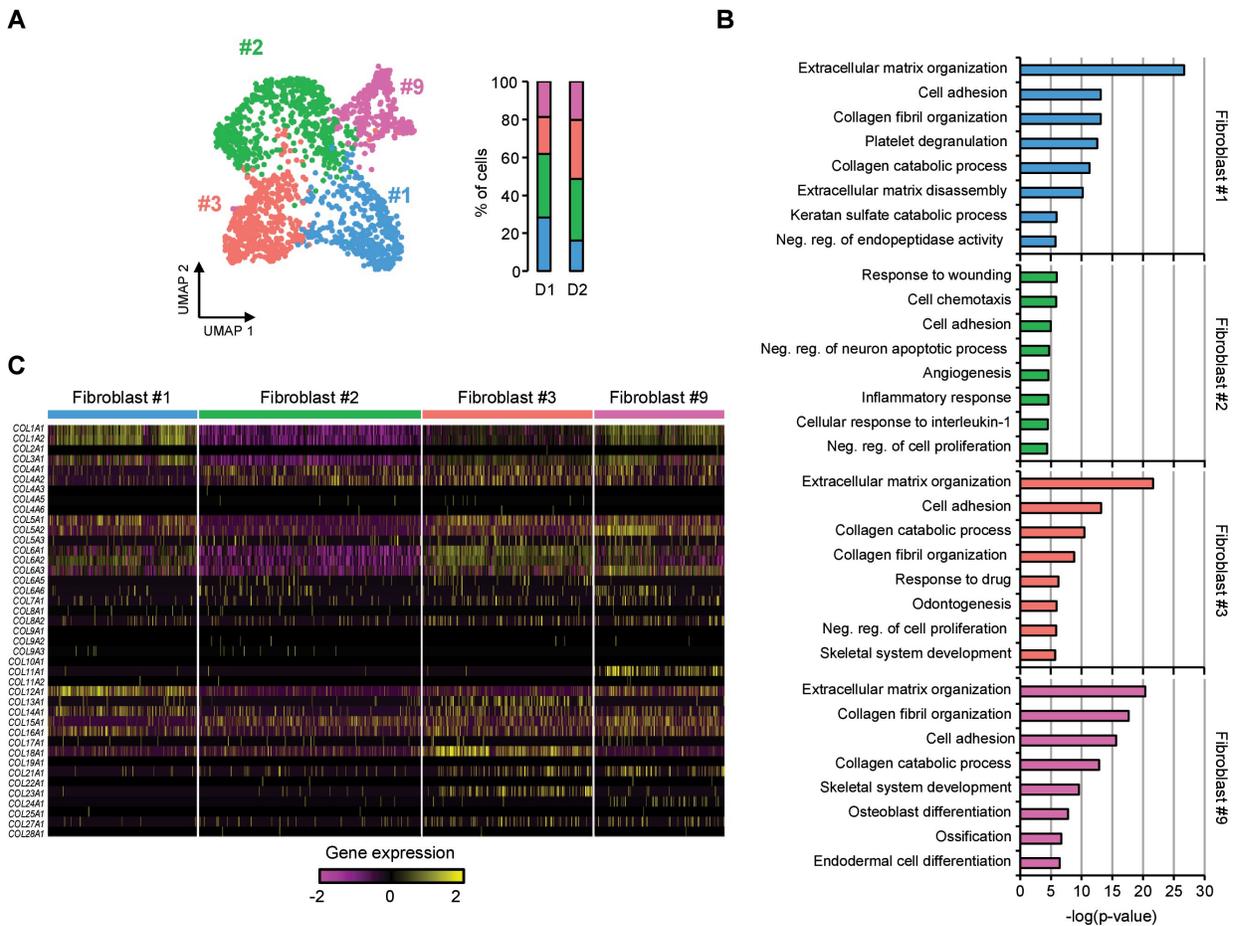
**Figure 24. Minor skin cell populations cluster with main cell types.** UMAP plots displaying the average expression of particular sets of well-known marker genes for minor skin cell populations (highlighted with a red arrow or circle). Red indicates maximum average gene expression while blue indicates low or no expression in log-normalized UMI counts. HF: hair follicle.

### 3.3.2. Functional and spatial annotation of fibroblast subpopulations.

To assess whether fibroblast clusters represented distinct functional subpopulations, gene ontology (GO) analyses were performed using the most representative genes expressed by each cluster. However, it is known that aging induces major changes in human skin and its fibroblasts, affecting their transcriptional profiles (Tigges *et al.*, 2014; Rittié and Fisher, 2015). Thus, this analysis was particularly focused on the expression of the 1,792 fibroblasts from young donors (Figure 25A). Archetypical functions of fibroblasts in homeostatic conditions, related to collagen and ECM production and organization were substantially enriched in three clusters (#1, #3, and #9; Figure 25B). Mesenchymal-related functional annotations such as *skeletal system development*, *ossification*, or *osteoblast differentiation* were enriched in clusters #3 and #9 (Figure 25B). Of note, cluster #2 did not display classical fibroblast functions in the top GO terms but showed a strong enrichment for inflammation-related terms including *inflammatory response* or

### 3. Results

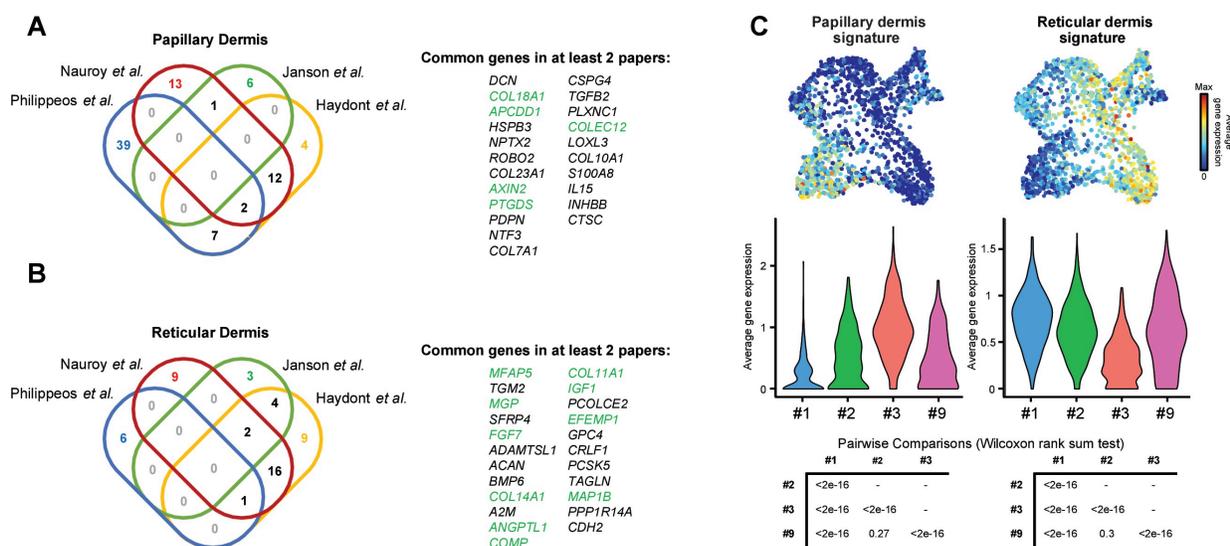
*cell chemotaxis* (Figure 25B). These results provided the first indication of fibroblast functional heterogeneity in the dataset analyzed in this thesis.



**Figure 25. Dermal fibroblast subpopulations display distinct functional transcriptomic signatures. (A)** Left: UMAP plot displaying the 1,792 fibroblasts from the two young donors. Each dot represents a cell and coloring is according to unsupervised clustering. Right: Bar plots depicting the proportions of the four fibroblast subpopulations in each of the two young donors. **(B)** Bar plots depicting the top eight enriched GO terms in each fibroblast subpopulation, sorted by p-value. **(C)** Heatmap displaying the expression of all collagen genes in the four fibroblast subpopulations. Yellow indicates maximum gene expression while purple indicates no expression in scaled log-normalized UMI counts.

Collagens are the main structural protein in connective tissues and consist of 28 subtypes with distinct expression in different organs (Ricard-Blum, 2011). Also, some of these subtypes are associated with particular fibroblast functions (Bielajew, Hu and Athanasiou, 2020). Hence, the collagen expression patterns of the four fibroblast clusters were analyzed (Figure 25C).

Consistent with the absence of ECM-related terms in the GO analysis, cluster #2 showed a generally decreased collagen expression compared to the other fibroblast clusters (Figure 25C). *COL11A1* and *COL24A1*, associated with cartilage and bone development, respectively (Wang *et al.*, 2012; Li *et al.*, 2018), were found to be exclusively expressed by cluster #9, suggesting a stronger mesenchymal component for this fibroblast subpopulation (Figure 25C). *COL13A1*, *COL23A1*, and *COL18A1*, associated with the papillary dermis and the dermal-epidermal junction, were expressed by fibroblasts in cluster #3 (Figure 25C) (Peltonen *et al.*, 1999; Veit *et al.*, 2011; Philippeos *et al.*, 2018; Haydont *et al.*, 2019). Hence, the expression of these three collagen genes suggests that fibroblasts from this subpopulation are located within the papillary layer of the dermis.



**Figure 26. Dermal fibroblast subpopulations display distinct spatial signatures.** (A, B) Left: Venn diagram comparing genes found to be expressed in the papillary dermis (A) or reticular dermis (B) in Philippeos *et al.*, 2018, Nauroy *et al.* 2017, Janson *et al.* 2012, and Haydont *et al.* 2019. Right: List of genes detected in at least two of the previous studies. In green, detected genes in our single-cell transcriptomics dataset and used as papillary (A) or reticular (B) dermis signatures. (C) UMAP and violin plots showing the average gene expression for the papillary and reticular dermis signatures. In UMAP gene expression projections, red indicates maximum average gene expression while blue indicates low or no expression in log-normalized UMI counts. In violin plots, X-axes depict cell clusters and Y-axes show average gene expression in log-normalized UMI counts. Statistical significance of the papillary and reticular signature expression differences between fibroblast subpopulations was calculated using Wilcoxon rank-sum tests and p-values for each comparison are shown in the corresponding table.

To better predict the spatial localization of the four fibroblast subpopulations, papillary and reticular dermis gene signatures were generated by comparing sets of genes associated with papillary or reticular fibroblasts in previous publications (Figure 26A and 26B) (Janson *et al.*, 2012; Nauroy *et al.*, 2017; Philippeos *et al.*, 2018; Haydont *et al.*, 2019). Thus, the papillary dermis signature comprised five gene markers (*APCDD1*, *AXIN2*, *COLEC12*, *PTGDS*, and *COL18A1*), whereas the reticular dermis signature consisted of ten genes including *MGP*, *MFAP5*, or *MAP1B* (Figure 26A and 26B). Cluster #3 showed the highest papillary dermis signature expression and lowest reticular dermis signature, which is consistent with the collagen expression results (Figure 26C). In contrast, reticular dermis signature was enriched in fibroblasts from clusters #1, #2, and #9 with the highest reticular and lowest papillary signature expression found in cluster #1 (Figure 26C).

Altogether, these findings are consistent with a priming for distinct functional roles of the four identified fibroblast subpopulations. The analysis described two subpopulations with prominent roles in the generation of structural collagen and ECM organization and displaying distinct spatial localization: one located in the reticular dermis (cluster #1, “secretory-reticular fibroblasts”) and another one in the papillary dermis (cluster #3, “secretory-papillary fibroblasts”). Two additional subpopulations with predicted mixed localization were also identified, one characterized by pro-inflammatory functions (cluster #2, “pro-inflammatory fibroblasts”) and the other with a greater mesenchymal potential (cluster #9, “mesenchymal fibroblasts”).

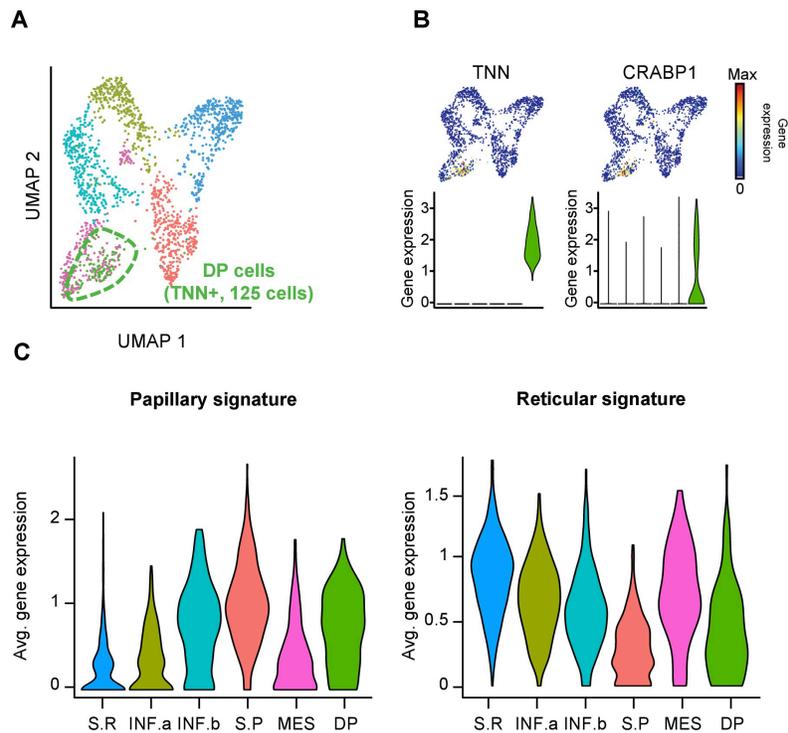
A second-level clustering performed with the 1,792 fibroblasts from young donors further validated the subdivision of fibroblasts into four main subpopulations (Figure 27A). Papillary and reticular secretory as well as mesenchymal fibroblast subpopulations were again identified in this analysis, whereas the pro-inflammatory subpopulation was further divided into two clusters (Figure 27A). These two clusters were closely related and were defined by the differential expression of a small subset of cytokine and chemokine genes such as *CXCL1* or *CCL19* (Figure 27A). Importantly, the subdivision of the pro-inflammatory subpopulation was also often found in second-level clustering analyses of individual samples from young and old donors (Figure 27B).

In addition, well-established dermal papilla (DP)-associated fibroblasts were separated from the mesenchymal subpopulation in second-level clustering of fibroblasts for some donors (Figure 27B). These highly specialized fibroblasts are located at the tip of hair follicles and regulate their growth cycle. They also play important roles in wound healing processes. Indeed, the expression of DP fibroblasts typical markers *CRABP1* and *TNN* (Driskell and Watt, 2015; Tabib *et al.*, 2018) was also detected in a subset of cells present in the mesenchymal subpopulation in the integrated dataset (Figure 28A and 28B). Analysis of the spatial signatures in dermal papilla fibroblasts



**Figure 27. Second-level analyses reveal further human dermal fibroblast heterogeneity.** (A) Left: UMAP plot displaying the 1,792 fibroblasts from the two young donors. Each dot represents a cell and coloring is according to unsupervised second-level clustering. Center: UMAP gene expression projections for some of the most representative genes in each fibroblast subpopulation. Red indicates maximum gene expression while blue indicates low or no expression in log-normalized UMI counts. Right: Violin plots depicting the expression of the most representative genes in each fibroblast subpopulation. (B) Left: UMAP plots depicting second-level clustering analyses in individual analyses in both young and old donors. Right: Violin plots showing the expression of selected marker genes for main and minor fibroblast subpopulations, including *TNN*, a marker for DP fibroblasts. In violin plots, X-axes depict cell clusters and Y-axes show average gene expression in log-normalized UMI counts.

Fibroblast populations identified in the second-level clustering analysis were clearly related to the four main fibroblast subpopulations and were therefore not considered separately for the rest of the study.



**Figure 28. Dermal papilla fibroblasts are contained within the mesenchymal fibroblast subpopulation.** (A) UMAP plot displaying the 1,792 fibroblasts from the two young donors. Each dot represents a cell. Coloring is according to unsupervised second-level clustering and highlighting the DP fibroblasts. (B) UMAP and violin plots displaying gene expression of *TNN* and *CRABP1*, known DP markers. Red indicates maximum gene expression while blue indicates low or no expression in log-normalized UMI counts. (C) Violin plots showing the expression for the papillary and reticular dermis signatures in the second-level fibroblast subpopulations and the DP fibroblasts. In violin plots, X-axes depict cell clusters and Y-axes show average gene expression in log-normalized UMI counts. S.R: secretory-reticular; INF: pro-inflammatory; S.P: secretory-papillary; MES: mesenchymal; DP: dermal papilla.

### 3.3.3. Validation of fibroblast subpopulations in skin sections

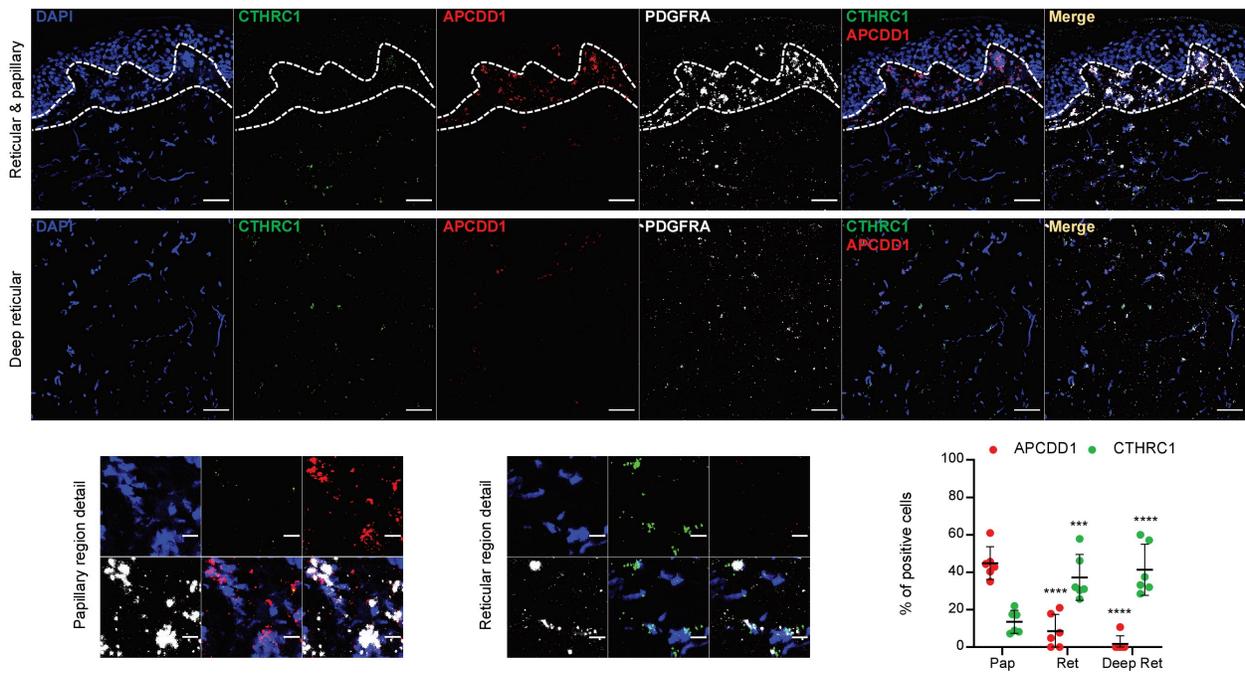
To further validate the four fibroblast subpopulations and assess their dermal location, RNA fluorescence *in situ* hybridization (FISH) assays were performed on independent FFPE skin sections from young (28-37 y/o) and older (54-86 y/o) donors. First, the most representative markers for each subpopulation were identified and are summarized in Table 3. As a control for fibroblast identity, a probe against the well-known fibroblast marker *Platelet-derived growth factor receptor alpha (PDGFRA)* was included in all hybridizations.

	Gene	FC	% cells in cluster	% cells not in cluster	Adj. p-value
Secretory-reticular	<i>WISP2</i>	15.06	0.83	0.074	0
	<i>SLPI</i>	11.64	0.729	0.065	0
	<i>CTHRC1</i>	7.69	0.759	0.092	0
	<i>MFAP5</i>	6.94	0.84	0.102	0
	<i>TSPAN8</i>	4.28	0.569	0.056	1.29E-259
Pro-inflammatory	<i>CCL19</i>	12.51	0.343	0.096	3.79E-75
	<i>APOE</i>	8.48	0.868	0.281	3.59E-275
	<i>CXCL2</i>	4.61	0.698	0.39	2.13E-80
	<i>CXCL3</i>	4.35	0.525	0.238	3.77E-63
	<i>EFEMP1</i>	3.12	0.564	0.126	6.36E-167
Secretory-papillary	<i>APCDD1</i>	6.03	0.78	0.11	0
	<i>ID1</i>	3.81	0.604	0.187	1.80E-109
	<i>WIF1</i>	3.74	0.438	0.035	3.01E-232
	<i>COL18A1</i>	2.96	0.581	0.168	1.68E-113
	<i>PTGDS</i>	2.94	0.559	0.196	1.90E-75
Mesenchymal	<i>ASPN</i>	8.75	0.666	0.067	7.31E-291
	<i>POSTN</i>	5.44	0.62	0.104	2.46E-170
	<i>GPC3</i>	3.58	0.513	0.063	2.83E-177
	<i>TNN</i>	3.42	0.337	0.007	2.10E-286
	<i>SFRP1</i>	3.26	0.406	0.04	5.61E-165

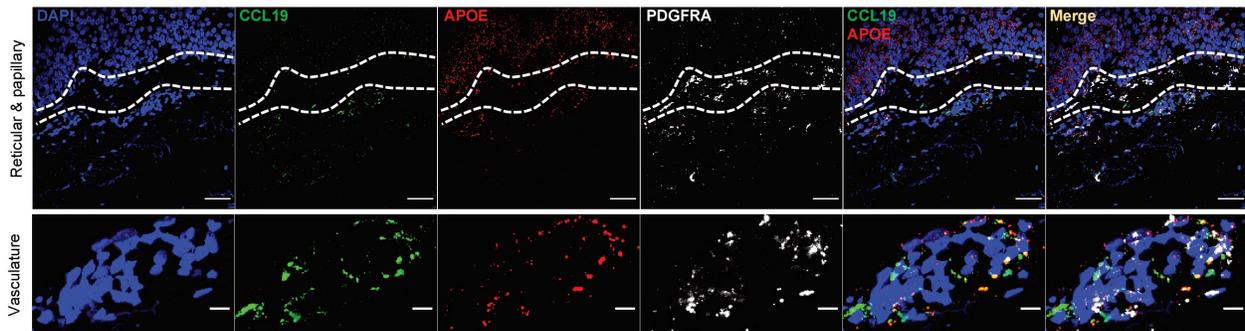
**Table 3. Most representative gene markers of human dermal fibroblast subpopulations.** Summary of the five genes selected as marker genes for each fibroblast subpopulation. Genes were selected according to their fold-change and enriched expression in the cluster (with a concomitant low percentage of fibroblasts from other subpopulations expressing the gene), using the young dataset. FC: fold-change.

### 3. Results

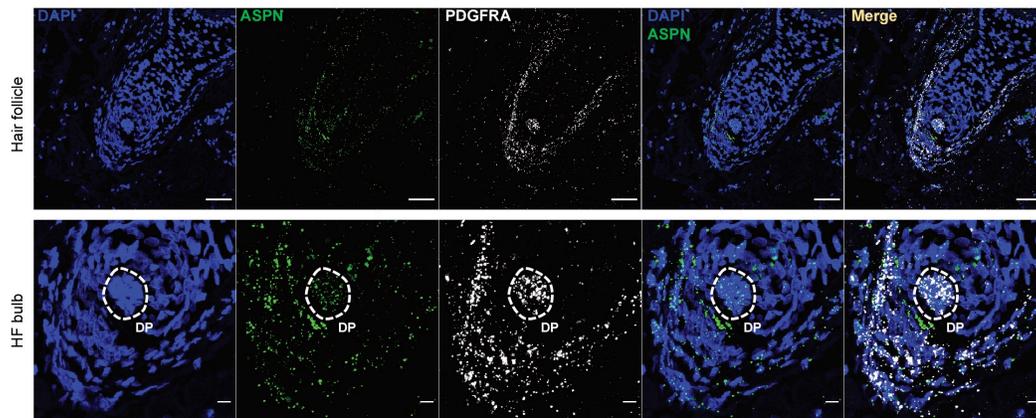
#### A Secretory-reticular and secretory- papillary



#### B Pro-inflammatory



#### C Mesenchymal



---

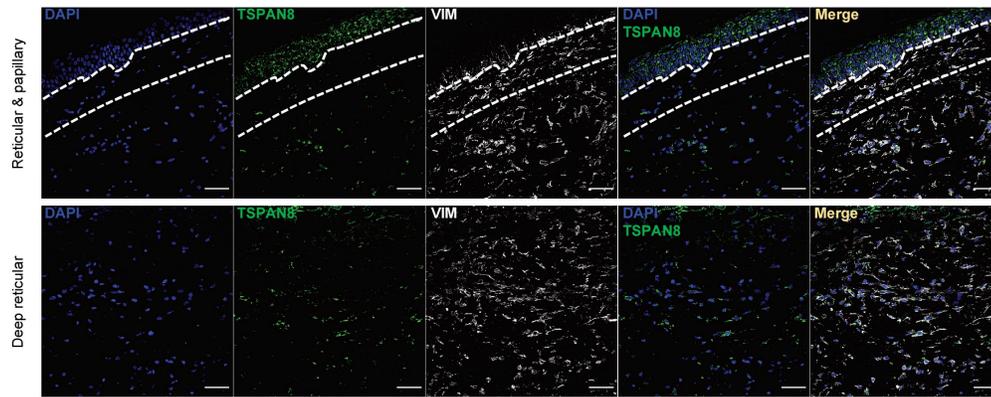
**Figure 29. Fibroblast subpopulations validation in young skin using mRNA-FISH. (A)** Representative confocal images for mRNA detection of *CTHRC1* (green) and *APCDD1* (red), markers for secretory-reticular and secretory-papillary subpopulations, respectively. Details from the papillary and reticular dermis are shown in the lower panel (left and center). Quantification of positive cells for each gene and dermal region is shown in the lower right panel. **(B)** Representative confocal images for mRNA detection of *CCL19* (green) and *APOE* (red), markers for pro-inflammatory fibroblasts. A detail of a vessel is shown in the lower panel. **(C)** Representative confocal images for mRNA detection of *ASPN* (green), a marker gene for mesenchymal fibroblasts. A detail of the HF bulb is shown in the lower panel. In **(A)** and **(B)** dashed line denotes the papillary dermis while in **(C)**, it denotes the DP. Nuclei were counterstained with DAPI and fibroblast identity was assessed by mRNA detection of *PDGFRA* (gray) in all assays. Each experiment was performed in three young FFPE skin sections (28-37 y/o). Images are shown at 40x original magnification. Scale bar, 50  $\mu\text{m}$  (main images), and 10  $\mu\text{m}$  (details). Statistical analyses in **(A)** were performed using a two-way ANOVA test (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .); error bars represent the standard deviation. Pap: papillary dermis, Ret: reticular dermis, Deep ret: deep reticular dermis, HF: hair follicle, DP: dermal papilla.

---

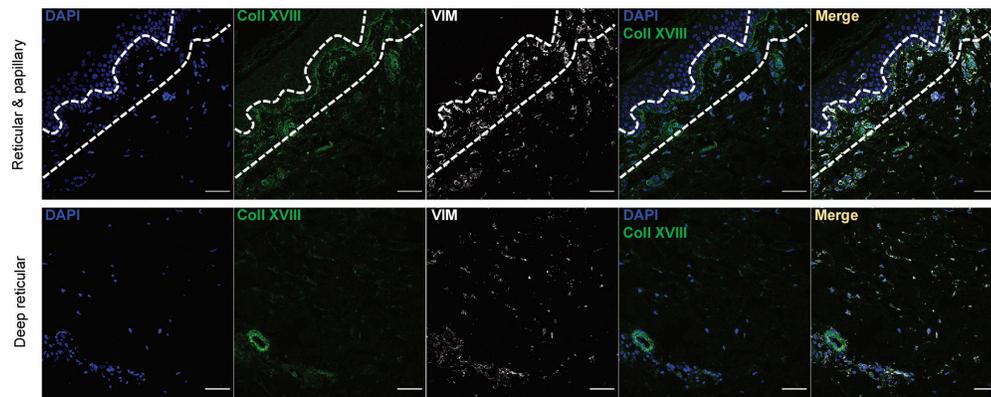
Thus, *Collagen Triple Helix Repeat-Containing 1 (CTHRC1)* and *Adenomatosis polyposis coli down-regulated 1 (APCDD1)* were selected as marker genes to detect secretory-reticular and secretory-papillary subpopulations, respectively. mRNA-FISH experiments validated the predicted spatial localization of the two secretory subpopulations as *CTHRC1* positive cells were mostly found in reticular and deep reticular dermal regions while *APCDD1* expression was exclusively found in the papillary dermis, in both young and old skin (Figure 29A and Figure S5). *C-C motif chemokine ligand 19 (CCL19)* and *Apolipoprotein E (APOE)* were used as marker genes to detect the pro-inflammatory subpopulation. Pro-inflammatory fibroblasts were mostly found in a perivascular localization and with a more widespread distribution as they could be detected in the papillary and reticular dermis (Figure 29B). Lastly, *Asporin (ASPN)* was used to detect mesenchymal fibroblasts, which were mostly found to be surrounding hair follicles (Figure 29C). Importantly, pro-inflammatory and mesenchymal fibroblasts subpopulations were also validated in aged skin (Figure S5).

In addition to the mRNA-FISH experiments, the dermal location of three of the four fibroblast subpopulations could also be confirmed at the protein level using immunofluorescence (IF) staining for additional markers. For this set of experiments, Vimentin was included as a control for fibroblast identity in all assays. Hence, Tetraspanin 8 and Collagen alpha-1(XVIII) chain were used to detect secretory-reticular and papillary fibroblasts, respectively. Consistent with the previous observations, these additional markers further validated their location at their respective dermal layers (Figure 30A and 30B). IF staining of an additional mesenchymal fibroblast marker, Periostin, was also detected in fibroblasts associated with hair follicles, although some signal was also detected in the papillary dermis region (Figure 30C).

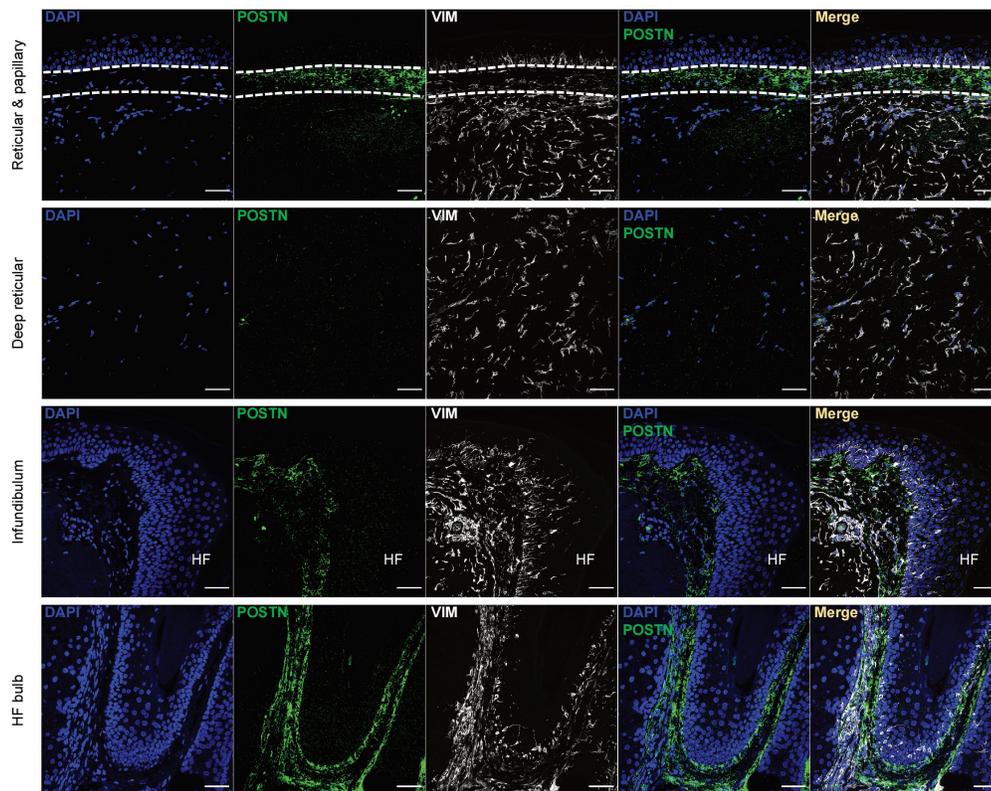
**A Secretory-reticular**



**B Secretory-papillary**



**C Mesenchymal**



---

**Figure 30. Fibroblast subpopulations validation using immunofluorescence assays.** (A) Representative confocal images of Tetraspanin 8 (TSPAN8) protein detection (green), a marker for secretory-reticular fibroblasts (n=6). (B) Representative confocal images of Collagen alpha-1(XVIII) chain (Coll XVIII) protein detection (green), a marker for secretory-papillary fibroblasts (n=1). (C) Representative confocal images of Periostin (POSTN) protein detection (green), a marker for mesenchymal fibroblasts (n=3). All assays were performed in FFPE skin sections (37-86 y/o). Nuclei were counterstained with DAPI and fibroblast identity was assessed by Vimentin (VIM) protein detection (gray). Dashed lines denote the papillary dermis region. Images are shown at 40x magnification. Scale bar, 50  $\mu$ m. HF: hair follicle.

---

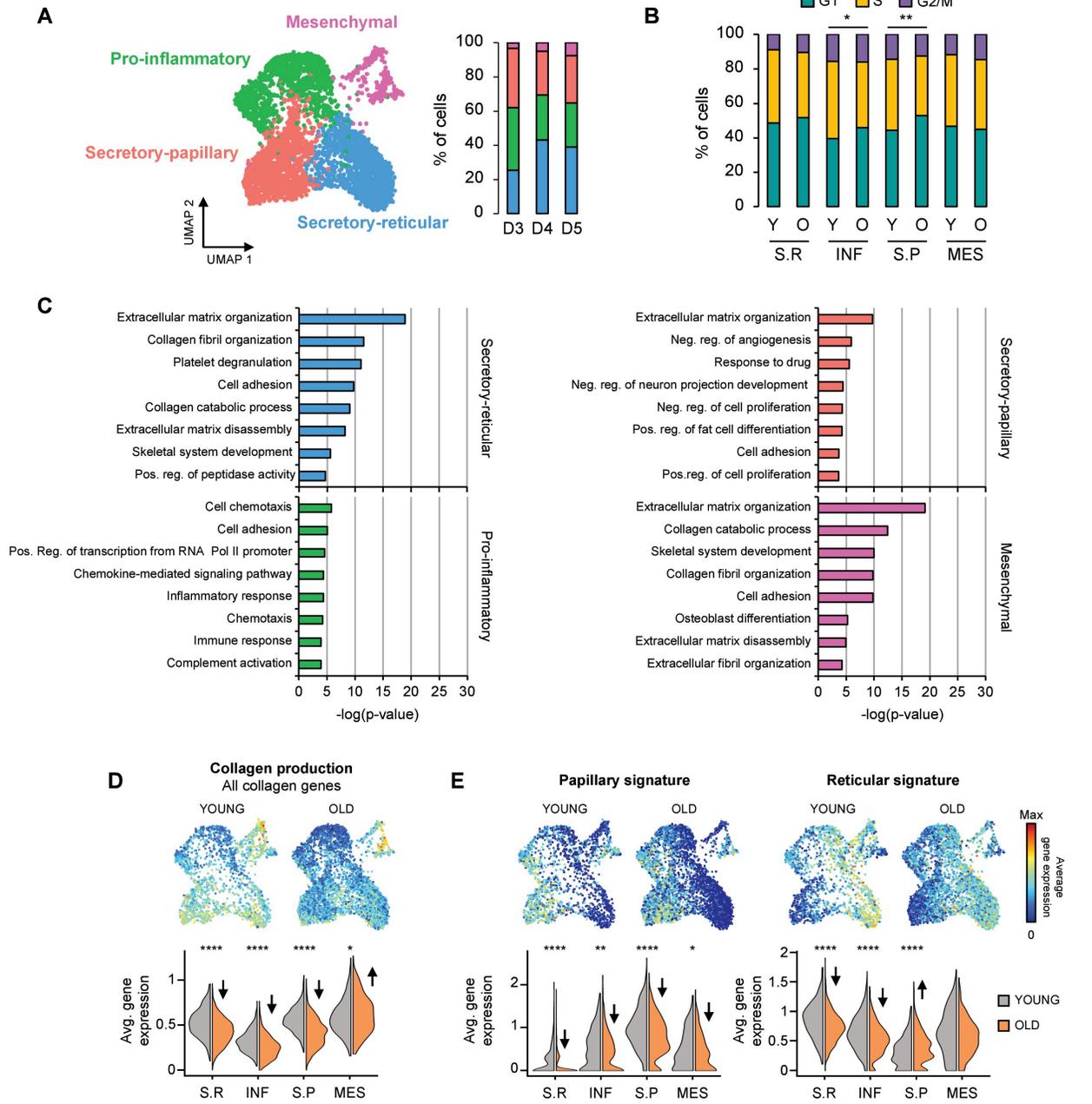
Altogether, these results provided important validation of the four fibroblast subpopulations identified in the scRNA-seq analysis in both young and aged skin, and highlighted their particular spatial distribution within the dermis.

### 3.3.4. Aging leads to loss of dermal fibroblast functional priming

Specific changes induced by intrinsic aging in dermal fibroblast subpopulations were subsequently characterized in detail. A substantial reduction in the number of mesenchymal fibroblasts in older donors compared to the younger counterparts was observed (Figure 31A). Nonetheless, validation of these findings by quantification of the mesenchymal fibroblasts in mRNA-FISH or IF assays was not possible in the scope of this thesis, as the number of hair follicles present in the available skin sections was very limited. The cell cycle prediction tool implemented in the Seurat R package (Stuart *et al.*, 2019) was used to examine whether young and old fibroblasts display distinct proliferative capacities. Indeed, a significant delay at the G1/S transition of the cell cycle was detected in the pro-inflammatory and secretory-papillary subpopulations (Figure 31B), which is consistent with the previously described less proliferative phenotype of aged cells (Wlaschek *et al.*, 2021). Although not significant, a similar trend was observed in the secretory-reticular cells while aged mesenchymal fibroblasts seemed to not present proliferation defects (Figure 31B).

Mimicking the analysis in young samples, fibroblast subpopulations from older donors were functionally annotated by performing GO analyses of the most representative genes of each subpopulation. These indicated a general loss of functional annotations upon aging, illustrated by a reduced number of GO-terms related to the previously defined functions as well as considerably less significant p-values, in agreement with a reduced number of genes associated with the terms (Figure 31C). Similarly, a global reduction in collagen production in aged fibroblasts was found, especially affecting both secretory subpopulations (Figure 31D).

### 3. Results

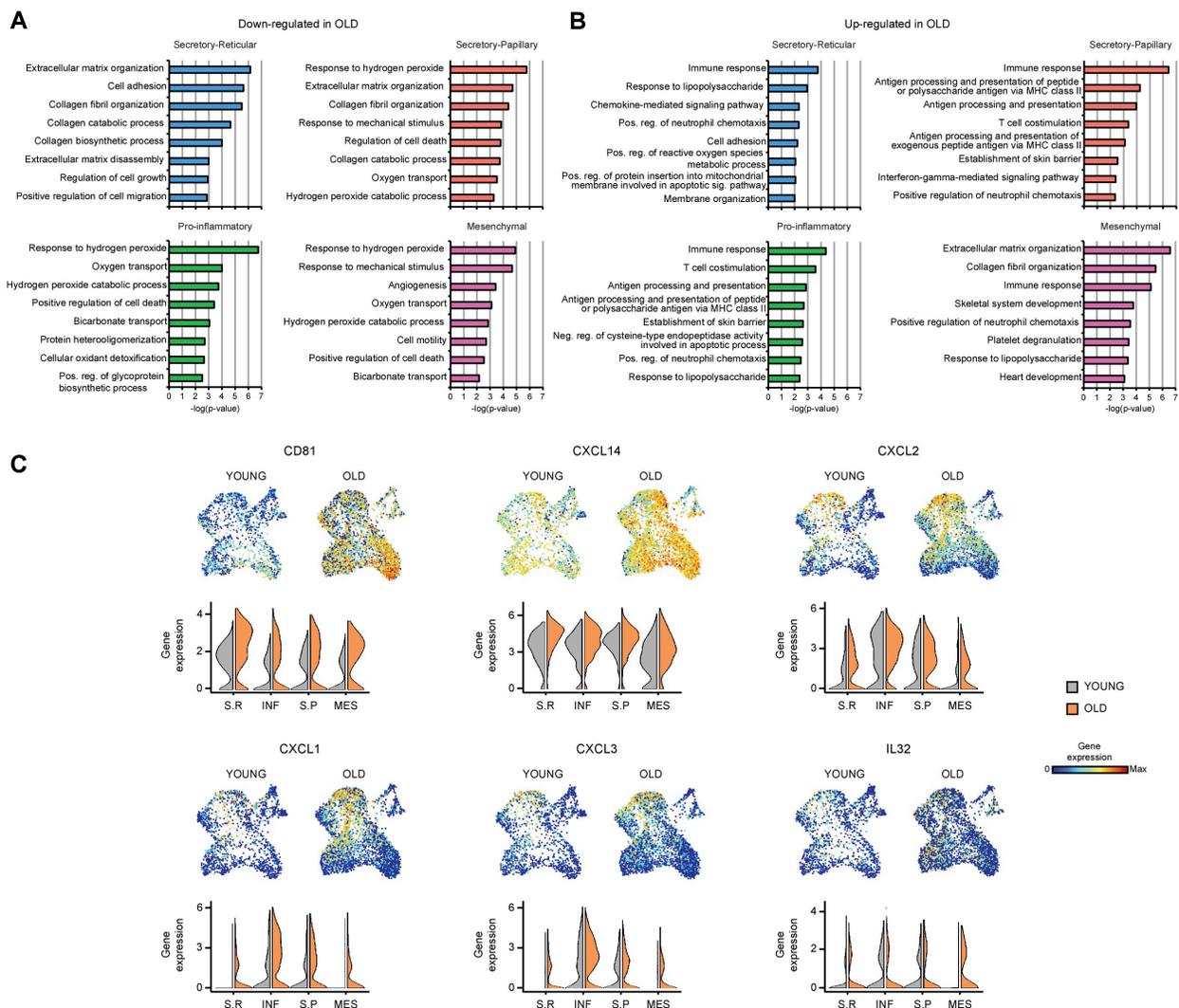


**Figure 31. Aging leads to loss of dermal fibroblast functional priming. (A)** Left: UMAP plot displaying the 4,156 fibroblasts from the three old donors. Right: Bar plots depicting the proportions of fibroblast subpopulations in each old donor. **(B)** Proportion of fibroblast at each cell cycle phase in the four subpopulations in young and old skin samples. **(C)** Bar plots depicting the top eight enriched GO terms in each old fibroblast subpopulation, sorted by p-value. **(D, E)** UMAP and violin plots showing the average expression of all collagen genes **(D)** or papillary and reticular signatures **(E)** in fibroblast subpopulations of young and old skin. In UMAP plots, red indicates maximum average gene expression while blue indicates low or no expression. In violin plots, X-axes depict cell clusters and Y-axes show the average. Gene expression is in log-normalized UMI counts. A two-sided two-proportion z-test was used in **(B)** and a Wilcoxon Rank Sum test was used in **(D)** and **(E)** for statistical analyses. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . Y: young, O: old, S.R: secretory-reticular, INF: pro-inflammatory, S.P: secretory-papillary, MES: mesenchymal.

Importantly, aging also induced important changes in the expression of the spatial signatures. Thus, aged papillary fibroblasts displayed a reduced expression of the papillary signature but higher expression of the reticular signature (Figure 31E). In contrast, reticular dermis signature expression was decreased in aged reticular fibroblasts (Figure 31E). Taken together, these results suggest an important age-dependent loss of fibroblast spatial signatures and functional priming.

### 3.3.5. Aged fibroblast transcriptomes recapitulate known old skin features.

To further examine the aging-related effects in the transcriptional profiles of dermal fibroblasts, GO analyses were performed using the up-regulated and down-regulated genes in each subpopulation upon aging. Among the enriched GO terms found with the down-regulated genes,



**Figure 32. Old fibroblasts display a global increase in immune response and decreased oxidative metabolism.**

(A) Bar plots depicting the top eight enriched GO terms obtained with down-regulated genes in old fibroblast subpopulations, sorted by p-value. (B) Bar plots depicting the top eight enriched GO terms obtained with up-regulated genes in old fibroblast subpopulations, sorted by p-value. (C) UMAP and violin plots showing the expression of significantly differentially expressed genes related to pro-inflammatory functions in fibroblast subpopulations of young and old skin. In UMAP gene expression projections, red indicates maximum gene expression while blue indicates low or no expression in log-normalized UMI counts. In violin plots, X-axes depict cell clusters and Y-axes show gene expression in log-normalized UMI counts. S.R: secretory-reticular, INF: pro-inflammatory, S.P: secretory-papillary, MES: mesenchymal.

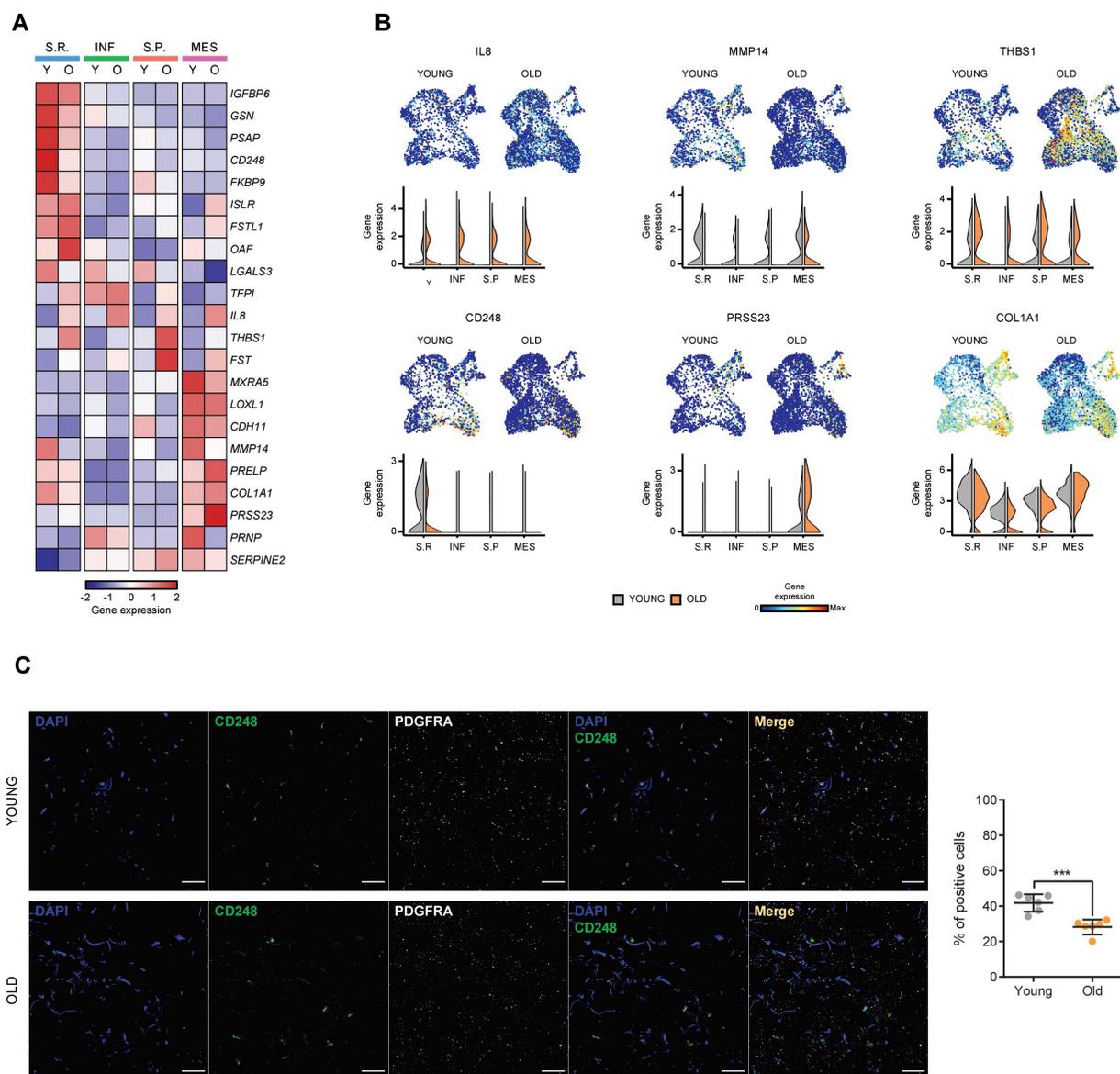
---

an enrichment for terms related to hydrogen peroxide metabolism was observed in three of the four subpopulations (Figure 32A). The reduced expression of genes related to this metabolic pathway is consistent with the previously described accumulation of ROS in aged fibroblasts (Fisher *et al.*, 2009; Tigges *et al.*, 2014). On the other hand, GO analysis performed using the up-regulated genes revealed the enrichment for immune response-related terms in all subpopulations, consistent with the known chronic, low-grade inflammatory phenotype present in old skin (Figure 32B) (Tigges *et al.*, 2014; Lee *et al.*, 2021).

#### **3.3.6. Fibroblast subpopulations display distinct skin aging-associated secreted proteins (SAASP) expression patterns.**

Aging is associated with an accumulation of senescent cells in tissues, which leads to impaired organ functionality and promotion of several age-dependent conditions (Campisi, 2005). Senescent cells acquire particular secretory features, which is known as senescence-associated secretory phenotype (SASP) (Ghosh and Capell, 2016). Similarly, a set of skin aging-associated secreted proteins (SAASP) was previously characterized in intrinsically aged human dermal fibroblasts (Waldera Lupa *et al.*, 2015). Some of the biological functions associated with the SAASP included ECM organization, immune response, and cell metabolism.

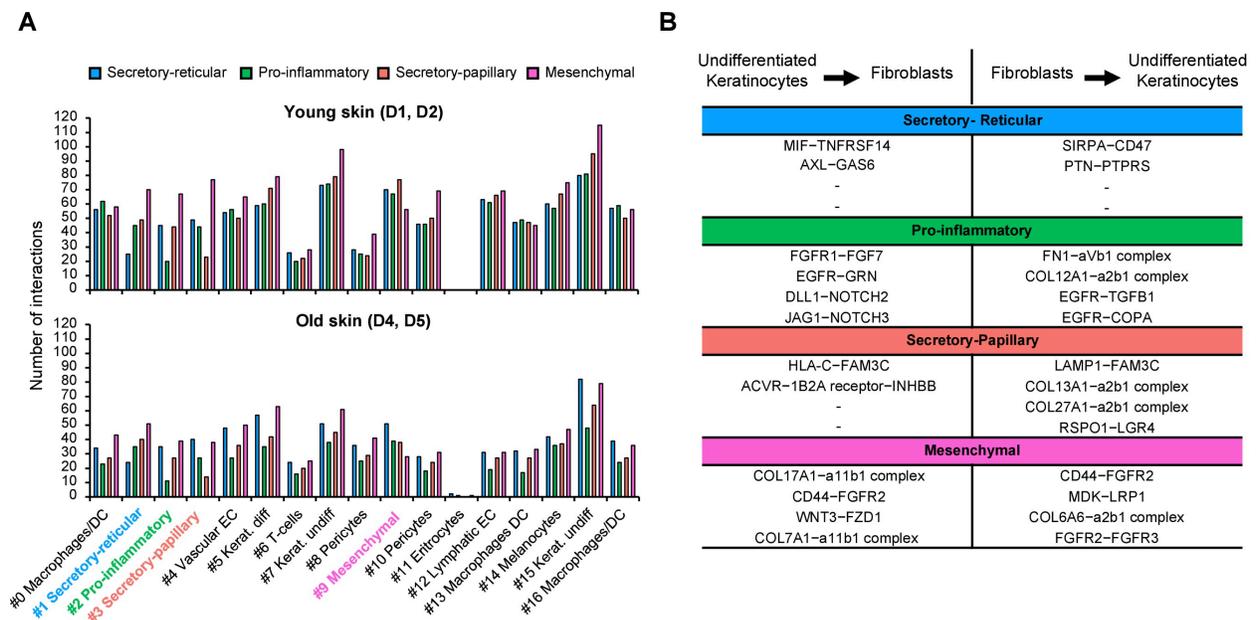
The expression of SAASP-related genes was examined in the fibroblast from both young and old donors (Figure 33A). This analysis uncovered distinct expression patterns of SAASP-related genes in the four old subpopulations (Figure 33A and 33B). To experimentally demonstrate these findings, mRNA-FISH of *CD248*, which was predicted to be significantly less expressed in the old secretory-reticular fibroblasts, was performed. In agreement with the scRNA-seq data, a reduction in the percentage of *CD248*-positive cells was observed in the reticular dermis of aged skin (Figure 33C). Taken together, these findings suggest a distinct contribution of each fibroblast subpopulation to the skin aging-associated secretory phenotype.



**Figure 33. SAASP-related gene expression in young and old dermal fibroblast subpopulations. (A)** Heatmap displaying the average expression of SAASP-related genes differentially expressed between young and old fibroblast (fold-change > 1.25) in at least one subpopulation. **(B)** UMAP and violin plots showing the expression of SAASP-related genes differentially expressed between young and old fibroblasts in at least one subpopulation. In UMAP gene expression projections, red indicates maximum average gene expression while blue indicates low or no expression. In violin plots, X-axes depict cell clusters and Y-axes show average gene expression of fibroblast identity genes. Gene expression is in log-normalized UMI counts. **(C)** Right: Representative confocal images of mRNA detection of *CD248* (green) and *PDGFRA* (grey) in deep reticular dermal regions of young and old skin. Left: Quantification of *CD248*-positive cells in young and old skin sections. Assay was performed in three young (28-37 y/o) and three old (79-89 y/o) skin FFPE sections. Images are shown at 40x magnification. Scale bar, 50  $\mu$ m. A two-sided t-test was performed for statistical analysis. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . Y: young, O: old, S.R: secretory-reticular, INF: pro-inflammatory, S.P: secretory-papillary, MES: mesenchymal.

### 3.3.7. Aging leads to a loss of cell-cell interactions in dermal fibroblasts.

Dermal fibroblasts are known to communicate with surrounding cells through multiple paracrine cell-cell interactions in homeostatic conditions (Sriram, Bigliardi and Bigliardi-Qi, 2015). Several approaches have been proposed to assess cellular communication using single-cell transcriptomics. In that regard, the publicly available repository CellPhoneDB predicts paracrine and autocrine interactions by analyzing the expression of cell-surface receptors and their known interacting ligands in the distinct cell populations contained in scRNA-seq data (Vento-Tormo *et al.*, 2018). This tool was used to identify the potential interactions between the four fibroblast subpopulations and the rest of the cell types present in both young and old skin. Interestingly, a general reduction in the number of interactions could be observed in older fibroblasts, with a particularly striking effect for interactions with undifferentiated keratinocytes (clusters #5 and #15: Figure 34A).



**Figure 34. Age-dependent loss of cell-cell interactions in dermal fibroblasts.** (A) Bar plots depicting the total number of predicted interactions between each fibroblast subpopulation and the rest of the cell types present in young and old skin. Donor 3 (53 y/o) displayed an intermediate phenotype, thus it was omitted in this analysis. (B) Table summarizing the top four exclusive interactions lost upon intrinsic aging between each fibroblast subpopulation and undifferentiated keratinocytes, sorted by p-value. For each cluster pair, interactions in both directions are shown.

The loss of cell-cell interactions was more pronounced in the two oldest samples ( $\geq 69$  y/o) while donor 3 (53 y/o) displayed an intermediate phenotype (Figure S6). Among the lost interactions between fibroblasts and undifferentiated keratinocytes, enrichment was found for those involving collagens and integrins from the epidermal-dermal junction such as COL13A1- $\alpha 2\beta 1$  complex, COL17A1- $\alpha 11\beta 1$  complex, or COL7A1- $\alpha 11\beta 1$ , consistent with the thinning of this particular junction upon aging (Peltonen *et al.*, 1999; Langton *et al.*, 2016) (Figure 34B). These findings indicate a loss of cellular communication between dermal fibroblasts and other skin cell types as an unrecognized phenomenon in human skin aging.



## 4. Discussion

Cancer cells display major epigenetic alterations, including aberrant DNA methylation profiles. In addition, cancer methylomes also reflect the epigenetic programs of the tumor-initiating cell, which can be used as a sort of fingerprint to unravel their origin (Kulis *et al.*, 2013; Moran, Arribas and Esteller, 2016). Accordingly, tumor stratification based on the epigenetic patterns of their cell-of-origin has been described in many human malignancies, often with clinical implications (Bormann *et al.*, 2018; Hoadley *et al.*, 2018; Io Riso *et al.*, 2020). A previous study from Prof. Dr. Lyko's research group identified two subclasses of cSCC and its precursor lesion AK, which were proposed to arise from keratinocytes at two distinct differentiation stages (Rodríguez-Paredes *et al.*, 2018).

In this thesis, this concept was further explored and expanded to other epidermal entities, including *in situ* carcinoma (BD), rarely metastasizing BCC, and non-cancerous senile warts (SK). This work also aimed to investigate the potential clinical implications of the cell-of-origin subclasses identified in KC, and made use of single-cell transcriptomics and methylomics to further characterize their cells-of-origin. In addition, single-cell analyses were also used in this work to explore fibroblast diversity in human skin and to characterize the changes they undergo upon intrinsic aging.

The findings presented in this thesis represent a comprehensive epigenomic overview of KC, provide substantial insight into the origin of these malignancies, and offer new possibilities to develop robust risk stratification biomarkers. Furthermore, this work also makes an important contribution to the understanding of the roles of fibroblasts in human skin homeostasis and aging.

### 4.1. Cell-of-origin-based stratification can be expanded to other epidermal entities and has potential clinical implications

A cohort comprised of 102 methylomes from the epidermis of 12 healthy, 20 AK, 35 cSCC, 11 BD, 14 BCC, and 10 SK samples was characterized in this work. The classical cancer-specific alterations, including a global hypomethylation and hypermethylation of CpG islands, could be observed in both malignant keratinocyte carcinomas, cSCC and BCC. In addition, these aberrations were also present in pre-invasive lesions (AK and BD), in line with the altered DNA methylation patterns reported in the early stages of KC development and in healthy sun-exposed skin (Vandiver *et al.*, 2015; Rodríguez-Paredes *et al.*, 2018). Non-cancerous SK also displayed

such cancer-specific methylation features, as previously observed in hyperproliferative healthy tissues (Beerman *et al.*, 2013).

DNA methylation at enhancer regions can be used to define cellular identity (Wiench *et al.*, 2011). Here, methylation analyses based on the profiles at EpSC-specific enhancers stratified all the 102 samples into two main subclasses. These subclasses displayed either EpSC-like or Keratinocyte-like methylation profiles, thus providing important confirmation for the previously proposed bimodal cell-of-origin model (Rodríguez-Paredes *et al.*, 2018). Furthermore, the same subclasses were observed in publicly available datasets comprising diverse epidermal entities. A central finding of this work is the substantial bias towards a more differentiated cell-of-origin for epidermal entities bearing a lower metastatic potential, including BCC as well as senile (SK) and common (VV) warts. In contrast, most EpSC-like tumors consisted of samples with a higher metastatic potential (cSCC) and precursor lesions that may progress to metastatic cSCC (AK and BD). Thus, suggesting a potential prognostic value for the cell-of-origin-based stratification of epidermal tumors.

To assess the DNA methylation changes that occur in the progression of cSCC, the methylomes of AK, BD, and cSCC samples were compared. Almost no significant differences were found between tumors from the same cell-of-origin subclass, as reported previously in unstratified AK and cSCC (Rodríguez-Paredes *et al.*, 2018). Therefore, DNA methylation does not seem to play an essential role in the progression of this disease continuum. However, AK/BD/cSCC from different subclasses displayed major DNA methylation variation. Important histological differences have been reported between these three entities (Ishida *et al.*, 2002; Schmitz, 2019). AK and BD have even been proposed to originate from distinct keratinocyte populations based on their very distinctive histopathological features (Saglam *et al.*, 2008). Nonetheless, the findings reported in this work indicate a homogeneous aberrant DNA methylation between AK, BD, and cSCC, as long as they arise from the same cell-of-origin. This is coherent with the limited differences described between these three entities in terms of gene and protein expression (Padilla *et al.*, 2010; Lambert *et al.*, 2014; Azimi *et al.*, 2018, 2020).

DNA methylation clocks are molecular biomarkers that use age-related methylation changes in particular sets of CpGs to accurately estimate the chronological age of a given tissue (Bell *et al.*, 2019). Although their use in cancer tissues is still controversial, some studies correlated changes in methylation age to higher risk tumors (Zheng, Widschwendter and Teschendorff, 2016; T. Wang *et al.*, 2020; Chung *et al.*, 2021). Nonetheless, other similar DNA methylation-based approaches such as mitotic clocks seem to be more reliable in cancer risk assessment (Yang *et al.*, 2016; Duran-Ferrer *et al.*, 2020; Teschendorff, 2020). Using the mitotic clock

epiTOC2 (Teschendorff, 2020), a general increase in the mitotic age and SCDR was observed in epidermal tumors compared to the healthy epidermis. This is consistent with previous observations in other human malignant and pre-malignant tissues (Yang *et al.*, 2016; Teschendorff, 2020). Interestingly, this effect was significantly higher in Keratinocyte-like tumors, whereas EpSC-like samples showed a more moderate increase. This difference was hypothesized to reflect a less proliferative but more invasive phenotype in EpSC-like tumors.

Invasive cancer cells from epithelial tumors typically undergo a molecular re-programming in which they lose their epithelial identity in favor of a mesenchymal one (Celià-Terrassa *et al.*, 2012; Lambert, Pattabiraman and Weinberg, 2017). Through this epithelial-to-mesenchymal transition (EMT), cancer cells become less differentiated and gain migratory and invasive capacities (Celià-Terrassa *et al.*, 2012; Lambert, Pattabiraman and Weinberg, 2017). This is a reversible process, which implies a dynamic regulation. Hence, several epigenetic mechanisms have been found to contribute to EMT processes, including DNA methylation (Kiesslich, Pichler and Neureiter, 2013; Carmona *et al.*, 2014). ZEB1 and ZEB2 are important transcriptional repressors of E-cadherin and the epithelial phenotype in general. These transcription factors are repressed in epithelial cells but become highly expressed during the EMT process (Celià-Terrassa *et al.*, 2012). miRNAs from the miR200 family (i.e., miR-200-c and miR-141) and miR-205 play an essential role in maintaining the epithelial phenotype by targeting ZEB1 and ZEB2 (Gregory *et al.*, 2008; Park *et al.*, 2008). Silencing of these miRNAs by promoter hypermethylation has been described in several human cancer cell lines displaying EMT features and invasive epithelial human malignancies such as muscle-invasive bladder cancer (MIBC) (Neves *et al.*, 2010; Wiklund *et al.*, 2011; Davalos *et al.*, 2012). Consistent with the interpretation of the mitotic clock results, EpSC-like tumors showed promoter hypermethylation in *MIR200C/141* and *MIR205* genes as opposed to the healthy epidermis or Keratinocyte-like tumors. Hence, EpSC-like tumors display typical DNA methylation features of the EMT phenotype, suggesting a more invasive phenotype in this tumor subclass. Furthermore, the stratification of invasive cSCC samples from the dataset published in Hervás-Marín *et al.*, 2019 as EpSC-like tumors further supported this notion.

These results are reminiscent of the emerging concept of cellular pliancy, which states that upon tumorigenesis, distinct biological pathways can be activated in a particular cancer type due to the intrinsic characteristics of their cell-of-origin (i.e., differentiation stage), leading to distinct phenotypes and/or clinical outcomes (Puisieux *et al.*, 2018). In fact, this cellular pliancy has been observed in mouse cSCC as the epigenetic landscape of the cell-of-origin was found to prime the tumors to display EMT features (Latil *et al.*, 2017), also consistent with the findings described in this thesis.

An amplicon-based targeted sequencing approach was developed to stratify tumors more systematically and cost-effectively, and with the intention to assess metastasis-free time, recurrence, and survival rates in the cell-of-origin subclasses by analyzing FFPE archived material with clinical histories. This method was based on two regions with differential methylation dynamics between cell-of-origin subclasses, and its validation using FF tumors indicated an accurate tumor stratification. Nonetheless, using FFPE-derived material yielded contradictory results, which prevented a satisfactory tumor stratification. DNA extracted from FFPE tissues is often of poor quality due to the cross-linking and fragmentation induced by the formalin treatment, impairing its proper amplification (Gilbert *et al.*, 2007). Importantly, the methylomes obtained from FFPE epidermal tumors were highly correlated to those from FF samples when analyzed using the methylation arrays, in agreement with previous observations (Kling *et al.*, 2017). In contrast to the amplicon-based approach, FFPE-derived DNA is subjected to a restoration protocol before hybridizing to the methylation arrays (Kling *et al.*, 2017). This restoration step renders DNA to a sufficient quality for amplification steps. Using such restoration protocols could substantially improve the targeted amplicon strategy for FFPE samples and should be evaluated in future research.

The methylomes of 11 cSCC metastases and primary metastatic cSCC samples were also characterized. Malignant tissue was isolated from FFPE sections using laser microdissection, and DNA methylation was profiled using methylation EPIC arrays. Thus, taking advantage of the high correlation between methylomes from FFPE and FF material mentioned above. Interestingly, all samples were stratified as EpSC-like, suggesting a higher metastatic potential for this cell-of-origin subclass.

Collectively, these findings indicated that the previously proposed cell-of-origin-based stratification strategy can be applied to a wide range of epidermal tumors, and highlighted important biases in the cellular origins of KC with different metastatic potential. In addition, further characterization of their methylomes suggested cell-of-origin-related differences in the proliferative and invasive properties of KC and, thus, establish novel opportunities for improving patient risk assessment based on robust molecular markers.

### **4.2. Single-cell genomics defined the cellular origin of KC**

To further investigate epidermal differentiation, the single-cell transcriptomes from more than 32,000 keratinocytes were analyzed in this work. Two types of basal undifferentiated keratinocytes were identified, in line with the heterogeneity previously described in this epidermal

compartment by others (Cheng *et al.*, 2018; S. Wang *et al.*, 2020). Lineage trajectory analysis based on splicing dynamics revealed that only one of the two undifferentiated populations contributes to the epidermal lineage, which then progressed towards mitotic keratinocytes and eventually to the more differentiated spinous and granular populations. Thus, this analysis suggested that the homeostatic EpSCs are located in the Basal 1 population, expressing *KRT15* but not *KRT19*. In contrast, *KRT19*-expressing Basal 2 keratinocytes could represent a stem cell pool reservoir for the interfollicular epidermis. These results agree with previous research describing a subset of *KRT15*<sup>+</sup>/*KRT19*<sup>+</sup> keratinocytes believed to be self-renewing but not involved in terminal differentiation (Michel *et al.*, 1996; Pontiggia *et al.*, 2009). Of note, the lineage trajectory analysis did not identify granular cells as the endpoint of the epidermal differentiation. This could be explained either by the low number of granular cells present in the dataset or by the transcriptional bursting these cells experience before dying (Wotherspoon, Rogerson and O'shaughnessy, 2020), and that might affect the splicing dynamics used for velocity estimation (Bergen *et al.*, 2020).

A newly developed tool, EpiSCORE, based on the anticorrelation of gene promoter methylation and gene expression, allows the estimation of cell type fractions in bulk DNA methylation data using scRNA-seq data (Teschendorff *et al.*, 2020). Deconvolution of bulk KC methylomes revealed a heterogeneous composition with enrichment of distinct keratinocyte populations in EpSC-like and Keratinocyte-like tumors. This heterogeneity is in agreement with previous scRNA-seq results that uncovered the presence of undifferentiated, mitotic, and well-differentiated keratinocyte populations in human cSCC (Ji *et al.*, 2020), even if their monoclonal origin has been reported (Reeves *et al.*, 2018). Consistent with the cell-of-origin-based stratification, this analysis indicated a strong enrichment of EpSC-containing Basal 1 population in EpSC-like tumors, while differentiated spinous and granular keratinocytes were more prominent in Keratinocyte-like samples.

Single-cell DNA methylation studies are critically important for identifying the exact cell type that provides a tumor cell-of-origin. However, such methods are still in early developmental stages, and the application of current techniques is still limited to reasonably low cell numbers and/or CpG coverage (Karemaker and Vermeulen, 2018). In 2018, a new method with the potential to facilitate high-throughput studies for single-cell methylomics due to its combinatorial indexing strategy and high scalability was published (Mulqueen *et al.*, 2018). Nevertheless, the original publication contains important inaccuracies and omits crucial information needed for successful sequencing, which has prevented the method's reproducibility thus far. In collaboration with the scOpenLab (DKFZ, Germany), these methodological shortcomings were addressed, and

a working protocol for cell lines and primary human epidermis were optimized in this thesis. Importantly, nuclei preparation was found to be highly sample-type dependent but critical for successful sci-MET experiments. The combined analysis of the single-cell methylomes obtained in 185 epidermal cells and the 102 bulk KC and healthy epidermis methylomes identified two cells as EpSC and 183 as differentiated keratinocytes. These numbers are consistent with previous data indicating that the human EpSCs represent the 1% of the epidermal cells (Rachidi *et al.*, 2007) and validate the cell-of-origin-based stratification of human epidermal tumors.

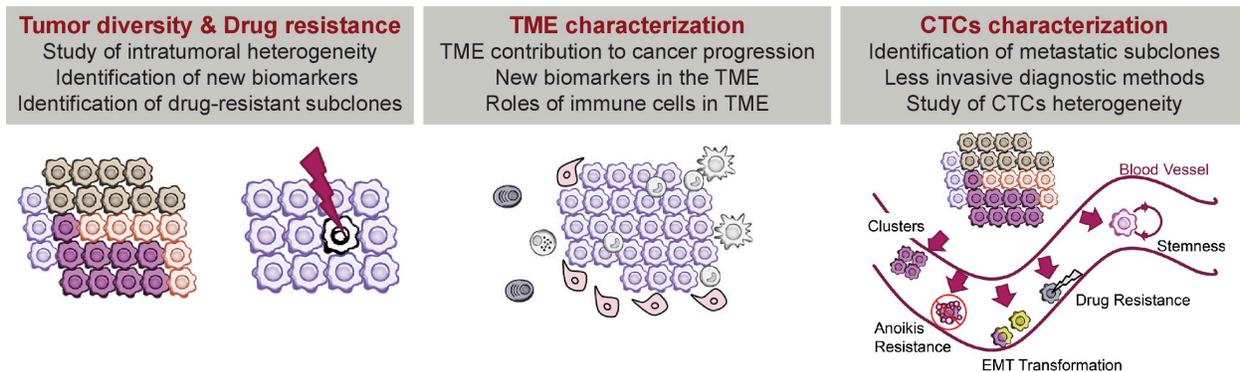
Specific analytical tools for single-cell methylation datasets are still limited, and future research in this regard is needed to bring this field forward. Most scWGBS methods provide very sparse data due to the low CpG coverages per cell, complicating downstream analyses (Karemaker and Vermeulen, 2018). Some tools have been developed to impute the missing methylation states through, for instance, probabilistic clustering (de Souza *et al.*, 2020) or deep neuronal networks (Angermueller *et al.*, 2017). The 185 single-cell datasets analyzed in this work displayed an average CpG coverage per cell of 0.79%. Although similar to the one reported originally, this coverage is lower than other methods, and the analysis could potentially benefit from such tools (Karemaker and Vermeulen, 2018; Mulqueen *et al.*, 2018).

Altogether, single-cell genomics provided important confirmation for the proposed cell-of-origin stratification of human KC and other epidermal tumors, as well as provided novel insight into the cellular origin of such malignancies.

### **4.3. Single-cell (epi)genomics contribution to precision oncology**

Cancer is a highly heterogeneous disease, with solid tumors consisting of multiple clones harboring distinctive genetic and epigenetic alterations (Greaves and Maley, 2012). Such alterations promote cellular plasticity, thus the ability of a cell to change its identity by, for instance, de-differentiation or re-programming processes (Flavahan, Gaskell and Bernstein, 2017). Furthermore, cancer clones and transient states change and evolve during tumor progression and are subjected to selective pressures (Flavahan, Gaskell and Bernstein, 2017; Ciriello and Magnani, 2021). The resulting intra-tumoral heterogeneity complicates patient treatment and promotes disease recurrence as not all clones might be affected by the selected therapy. Therefore, advances towards more personalized approaches are needed to improve cancer patient outcomes, and single-cell analyses have arisen as a promising tool for bringing cancer precision medicine forward.

Single-cell genomics methodologies, mostly single-cell transcriptomics, are already being applied in translational research to characterize tumor heterogeneity, clonal evolution, and drug resistance (Zhu *et al.*, 2017; Nath and Bild, 2021). For instance, scRNA-seq revealed important treatment-resistant clones in glioblastoma (Wang *et al.*, 2017), distinct cell populations with specific gene expression and influencing patient survival in breast cancer (Karaayvaz *et al.*, 2018) or identified different subpopulations of metastatic and non-metastatic circulating tumor cells (CTCs) in non-small cell lung cancer patients (Stewart *et al.*, 2020). Importantly, single-cell epigenomic approaches could improve these analyses as epigenetics can define cellular identity and transitional states more robustly than gene expression (Bond *et al.*, 2020). As an illustrative example, single-cell Assay for Transposase-Accessible Chromatin using sequencing (scATAC-seq) was used to investigate chromatin accessibility in human hematopoiesis and to study the clonal evolution in acute myeloid leukemia (Corces *et al.*, 2016). Also using scATAC-seq, important regulatory programs were identified in therapy-responsive T cell subsets from the microenvironment of BCC (Satpathy *et al.*, 2019), thus indicating that single-cell epigenomics can also be applied to study drug responses in cancer patients.



**Figure 35. Single-cell genomics in translational cancer research.** The application of the various recently developed single-cell approaches to cancer research will allow the study of the intratumoral cell diversity at the different levels of genomic regulation. This might contribute in the near future to the identification of new biomarkers for early diagnosis. In addition, such methodologies may facilitate the identification of therapy-resistant and responsive cancer subclones, thus improving the development of targeted therapies. The tumor microenvironment (TME) influences the course of the disease. Hence, characterizing the distinct cell populations that constitute the TME of particular cancer types and how they affect tumor invasiveness or therapy responsiveness will most likely also contribute to developing new biomarkers and targeted therapies. Lastly, using single-cell genomics to characterize CTCs provides novel opportunities to identify biomarkers in disseminating and metastatic cancer cells, which may help develop more robust and less invasive diagnostic approaches.

*Adapted from Zhu et al., 2017.*

While the systematic application of single-cell methylomics in the clinic will depend on overcoming the current technological limitations, these approaches hold great promise to improve precision oncology. For instance, such methods could refine the identification of cell subsets with important clinical implications such as therapy-resistant clones or metastatic CTCs, as illustrated by pioneering studies (Pixberg *et al.*, 2017; Gkoutela *et al.*, 2019). Furthermore, multi-omics approaches are also being developed, allowing the analysis of different layers of genomic and epigenomic information from the same cell simultaneously (Hou *et al.*, 2016; Clark *et al.*, 2018). Thus, the future application of single-cell (epi)genomics in clinical settings will provide a deeper understanding of clinically relevant biological aspects of human cancer, potentially translating into better clinical outcomes for cancer patients.

### **4.4. Human dermal fibroblast display functional priming**

Dermal fibroblasts are a highly heterogeneous cell type with multiple roles in skin homeostasis and regeneration (Driskell and Watt, 2015; Rognoni and Watt, 2018). Furthermore, skin aging majorly affects the dermal compartment of the skin and its resident fibroblasts (Haydont, Bernard and Fortunel, 2019). However, the diversity of this particular cell type in humans and the cellular changes they experience upon intrinsic aging are still not fully understood. With the recent technological advances, such questions can be interrogated using high-throughput single-cell transcriptomics. In this work, healthy whole human skin was systematically analyzed by characterizing more than 15,000 single cells, focusing on fibroblasts, which represented more than 5,000 cells in the dataset. Skin aging is highly influenced by several factors such as gender, pigmentation, and sun exposure (Farage *et al.*, 2008; Rittié and Fisher, 2015); therefore, it is of outermost importance to ensure proper sample selection in order to avoid confounding effects. Thus, all five samples included in this study were obtained from the sun-protected ilioinguinal region of male donors, ranging from 25 y/o to 70y/o.

The generated whole skin dataset recapitulated the cellular composition previously observed in a single-cell study comprising healthy human sun-exposed skin and using the same tissue dissociation protocol (Tabib *et al.*, 2018). Importantly, dermal fibroblasts represented the most heterogeneous population in the sun-protected skin. Four main fibroblast subpopulations with distinct functional annotations were identified and interpreted to reflect a functional priming. Importantly, these subpopulations also displayed distinct spatial identities, and their dermal localization was validated *in situ* in skin sections from young and old donors. Two subpopulations represented secretory fibroblasts with prominent roles in the archetypical fibroblast functions

related to collagen production and ECM organization and were differentially located at the papillary or reticular dermis. These secretory-papillary and secretory-reticular subpopulations displayed major differences in their gene expression, as previously described for fibroblasts located in these dermal layers (Janson *et al.*, 2012; Driskell and Watt, 2015; Sriram, Bigliardi and Bigliardi-Qi, 2015).

In addition, a pro-inflammatory subpopulation was also detected, evidencing the immunoregulatory functions exerted by this cell type (Haniffa *et al.*, 2007). Fibroblasts are able to initiate and maintain inflammation by, for instance, secreting cytokines, chemokines, and growth factors (Smith *et al.*, 1997; Haniffa *et al.*, 2007). Furthermore, fibroblasts also have roles in leukocyte recruitment and extravasation (Enzerink and Vaheri, 2011). In agreement with that notion, the pro-inflammatory subpopulation, characterized by expression of *APOE* and *CCL19*, was found exclusively in a perivascular location in both papillary and reticular dermis. Importantly, tissue fibrosis occurs in response to inflammatory injury and is promoted by activated fibroblasts known as myofibroblasts (McAnulty, 2007; Griffin *et al.*, 2020). Research in mouse models suggested the perivascular fibroblasts as a predominant origin of myofibroblasts (Dulauroy *et al.*, 2012; di Carlo and Peduto, 2018). Whether the pro-inflammatory fibroblasts described in this thesis represent a similar fibroblast subtype remains unclear, and further research should be performed to establish the potential roles of the pro-inflammatory fibroblast subpopulation in regulating inflammatory and wounding processes.

The fourth fibroblasts subpopulation was defined by a higher mesenchymal potential, in line with previous studies indicating that fibroblasts retain the capacity to differentiate into other mesenchymal cell types under certain conditions (Chang, Li and Guo, 2014; Ichim, O'Heeron and Kesari, 2018). Mesenchymal fibroblasts were characterized by *ASPN* expression, which was found exclusively at the vicinity of hair follicles in what is known as the dermal sheath (DS). Consistently, dermal papilla (DP) fibroblasts, a highly specialized subtype located at the tip of hair follicles and regulating their growth cycle (Driskell *et al.*, 2011), were also contained in the mesenchymal subpopulation. Interestingly, the DS and DP fibroblasts are thought to contain dermal stem cells with essential roles in the homeostasis and regeneration of the epidermis (Rahmani *et al.*, 2014; Tsuboi *et al.*, 2020). The DP cells have been shown to differentiate into a wide range of cell types (Driskell *et al.*, 2011). Thus, consistent with the mesenchymal functions described in the single-cell analysis for this subpopulation.

These results showed some differences from previous scRNA-seq studies characterizing human dermal fibroblasts. For instance, only two main fibroblast populations, defined by the expression of *SFRP2/DPP4* and *FMO1/LSP1*, as well as five minor closely related subtypes, were

identified in sun-exposed skin (Tabib *et al.*, 2018). Nonetheless, no functional definition was provided, and their significance was not further explored. These subpopulations were not overlapping with the ones identified in this work as the expression of *DPP4*, *FMO1*, and *LSP1* was found to be either low in all fibroblasts or not restricted to a particular population (Figure S7). On the other hand, *SFRP2* was also not specific as it was highly expressed in both secretory populations and a subset of the pro-inflammatory fibroblasts (Figure S7). Another scRNA-seq analysis performed with 184 flow-sorted fibroblasts identified five subpopulations (Philippeos *et al.*, 2018). While the two main ones represented papillary and reticular fibroblasts, the others remained unclear as one comprised only five cells, and the other two partially contained pericytes and pre-adipocytes (Philippeos *et al.*, 2018). Most of the genes used to define such subpopulations did not show high expression in the four fibroblasts populations characterized in the sun-protected skin (Figure S7). These differences could be attributed to the use of a more heterogeneous group of samples or different single-cell technologies.

A human skin cell atlas comprising more than 500,000 cells from 19 donors and including fetal, adult, and chronically inflamed tissues was recently published (Reynolds *et al.*, 2021). Interestingly, this atlas was obtained using a different dissociation protocol than the dataset presented in this thesis, which could influence the observed cellular composition and gene expression (Botting *et al.*, 2017). Nonetheless, the authors compared their dataset with the one generated in this thesis, uncovering the same cellular composition (Reynolds *et al.*, 2021). Thus, highlighting the quality of the sun-protected skin dataset and the robustness of the scRNA-seq approaches to study cellular heterogeneity in complex tissues. Although not a primary focus of their study, the authors also described some dermal fibroblast heterogeneity and identified three subsets, primarily associated with ECM remodeling and maintenance (Reynolds *et al.*, 2021). One subpopulation also displayed an increased expression of chemokines but was only found in skin affected by atopic dermatitis.

A handful of studies interrogating human dermal fibroblast heterogeneity have been published since the publication (or during the revision process) of the scRNA-seq dataset described in this thesis. For instance, three large fibroblast clusters were identified in healthy abdominal skin obtained from three female donors (Vorstandlechner *et al.*, 2020). These represented functional subpopulations with roles associated with ECM remodeling, immune functions, cartilage development, and response to growth factors. In contrast to the analysis presented here, there was no distinction between papillary and reticular fibroblasts (Vorstandlechner *et al.*, 2020). Also, in a study comparing lesional and non-lesional skin from patients with atopic dermatitis and healthy skin, three main fibroblast populations were identified (He *et al.*, 2020). Although they find

some degree of functional specialization among them, the authors focused on characterizing a fibroblast population only present in the actinic dermatitis skin and with pro-inflammatory properties (He *et al.*, 2020).

Most importantly, a recent publication analyzed the similarities between the datasets published in Vorstandlechner *et al.*, 2020, He *et al.*, 2020, Tabib *et al.*, 2018 and the sun-protected dataset of this thesis (Ascensión *et al.*, 2021). The authors analyzed each dataset independently and identified clusters with similar gene expression in each one. Such clusters were subsequently merged into three axes in a semi-supervised manner, which represented three major dermal fibroblast populations with functional specification (Ascensión *et al.*, 2021). Hence, one group was responsible for ECM organization, another showed immune surveillance and inflammation-promoting roles, and the third comprised specialized subpopulations such as DP and dermo-hypodermal junction fibroblasts (Ascensión *et al.*, 2021). Interestingly, the distinction between papillary and reticular fibroblast was not evident in their analysis. These results highlighted the importance of comparing scRNA-seq datasets as a whole and not only based on gene markers, as distinct gene markers might be identified in different analyses even if they represent similar populations. These findings provide important confirmation for the functional fibroblast priming described in this thesis and indicate the presence of the defined subpopulations in other scRNA-seq datasets containing samples from diverse ages, gender, and anatomical region.

#### **4.5. Intrinsic aging leads to loss of fibroblast identity and cellular communication**

Aging is the time-dependent decline in functional and structural properties in any given organ due to the accumulation of harmful biological changes. A common set of cellular and molecular processes has been proposed to contribute to the aging phenotype, including increased genomic instability, epigenetic alterations, mitochondrial dysfunction, and cellular senescence (López-Otín *et al.*, 2013). Skin aging is characterized by major structural and cellular changes, grossly affecting the dermal connective tissue and its associated fibroblasts (Rittié and Fisher, 2015; Shin *et al.*, 2019). To expand on this knowledge, the effects of intrinsic aging on dermal fibroblasts were characterized here at the single-cell level.

Comparison of the transcriptional profiles of fibroblasts between young and older donors uncovered substantial differences. Most importantly, an age-dependent loss of fibroblast priming was observed. Furthermore, the spatial identities of the fibroblast subpopulations were also less defined in aged skin. These results are complementary to the loss of fibroblast identity recently

described in murine dermis upon aging (Salzer *et al.*, 2018). While this loss was accompanied by an increased expression of adipogenesis genes in mice, this was not observed in the human fibroblasts.

A general upregulation of genes related to the immune response and inflammation was detected in older fibroblasts. This is consistent with the age-dependent low-grade chronic inflammation present in several human aged tissues, including the skin, and known as inflammaging (Franceschi and Campisi, 2014). Aged and senescent fibroblasts are known to promote this phenomenon by increased expression of pro-inflammatory cytokines and chemokines (Wlaschek *et al.*, 2021). The scRNA-seq results suggest that all fibroblast subpopulations contribute to the inflammaging observed in the human dermis and not exclusively the pro-inflammatory fibroblasts, as it could have been hypothesized. Aged fibroblasts accumulate reactive oxygen species (ROS), leading to the accumulation of DNA damage and ultimately promoting inflammaging (Tigges *et al.*, 2014). Consistently, a decreased expression of genes related to hydroxide peroxide metabolism was detected in aged fibroblasts. Thus, the scRNA-seq analysis recapitulated known aging phenotypes in dermal fibroblasts.

An apparent reduction of the mesenchymal fibroblast subpopulation was also observed, although it could not be adequately quantified in the scope of this thesis. However, hair loss is a common feature of skin aging and has been associated with reduced DP fibroblasts (Elliott, Stephenson and Messenger, 1999). Furthermore, a single-cell study on mice hair follicles showed that DP and dermal stem cells from the DS are reduced and present functional defects upon aging (Shin *et al.*, 2020). Thus, these studies are compatible with the loss of functional priming and reduced numbers proposed for the mesenchymal fibroblast subpopulation in aged human skin.

Lastly, dermal fibroblasts communicate with other skin cell types through paracrine and direct cell-cell interactions (Sriram, Bigliardi and Bigliardi-Qi, 2015; Lynch and Watt, 2018). This complex fibroblast interactome was observed in the skin from young donors. However, it was strongly dampened in aged skin as the number of interactions with other cell types was significantly reduced, including those with undifferentiated keratinocytes and involved in the dermal-epidermal junction. This general loss of cellular interactions might represent a previously unknown molecular feature of skin aging.

Altogether, this work provided novel insight into human dermal fibroblast diversity and their functional roles in homeostatic conditions and identified the loss of functional fibroblast priming and cellular interactions as important age-related phenotypes in the human skin.

## 4.6. A framework for the study of fibroblast roles in skin disease

How fibroblast heterogeneity is functionally relevant for skin homeostasis and disease is still a matter of intense research. It is speculated that particular subsets of fibroblasts might be involved differently in skin disease (Lynch and Watt, 2018). The different contribution of the four fibroblast subpopulations defined in this thesis to the SAASP phenotype is in line with this notion. Furthermore, recent work using scRNA-seq on human skin with fibrotic disorders (keloid and scleroderma) identified four subpopulations of fibroblast with similar functions and gene markers as the ones described in this work (Deng *et al.*, 2021). Importantly, an enrichment for mesenchymal fibroblasts was found in fibrotic skin, promoting fibrosis through collagen overexpression (Deng *et al.*, 2021). These results highlight the importance of characterizing the contributions of particular fibroblast subpopulations to disease phenotypes. This notion will be crucial for developing and improving targeted therapies for skin diseases in the following years.

Moreover, future research should also focus on deciphering how the age-related loss of fibroblast priming described in this work might influence skin disease and the molecular mechanisms that lead to such loss. Dermal fibroblasts accumulate epigenetic alterations upon skin aging (Tigges *et al.*, 2014). Thus, it is tempting to hypothesize that dysregulation of the epigenetic mechanisms might be involved in this age-related phenotype. Importantly, such loss of priming or identity could promote carcinogenic processes, as illustrated by the previously described shift from papillary to reticular phenotype in cSCC-associated fibroblasts, which enhanced tumor invasion (Hogervorst *et al.*, 2018). This shift was similarly observed upon intrinsic aging, and thus exemplifies the potential disease-promoting roles of the loss of fibroblast identity.

The accumulation of cellular damage seen in both aging and cancer is often a result of the same molecular mechanisms (Aunan, Cho and Sørreide, 2017); thus, older individuals are more likely to develop cancer than their younger counterparts. This is particularly important for KC, whose incidence dramatically increases in individuals over 60 years old (Garcovich *et al.*, 2017). There is growing evidence that the aged tumor microenvironment has higher cancer-promoting roles (Fane and Weeraratna, 2020). Consistent with this notion, skin aging and dermal fibroblast senescence have been found to contribute to skin carcinogenesis (Lee *et al.*, 2021). Furthermore, fibroblasts comprise the main component of the KC microenvironment (Amôr, Santos and Campanelli, 2021). Activated fibroblasts in the tumor microenvironment are known as cancer-associated fibroblasts (CAFs). Importantly, single-cell analyses of the tumor microenvironment of diverse human cancers have revealed significant heterogeneity of these CAFs (Kanzaki and Pietras, 2020). Nonetheless, the origin of such heterogeneous CAFs is still largely unknown.

Thus, further research is needed to clarify the roles and origin of CAFs in the context of the different types of skin cancer.

Hence, this thesis provides an important basis for future research in human skin homeostasis and aging. Furthermore, further characterization of the particular contributions of fibroblast subtypes in skin disease may contribute in developing better targeted interventions in conditions such as fibrotic disorders or skin cancer.

## 5. Conclusion

In this thesis, the cellular origin of KC and other epidermal tumors was investigated by characterizing a cohort comprised of 102 methylomes including both keratinocyte carcinomas (BCC and cSCC), the main precursor lesions of cSCC (AK and BD) as well as non-cancerous warts (SK) and healthy epidermis. A central finding of this work is the bias towards a more differentiated cell-of-origin for epidermal tumors with low metastatic potential. In contrast, tumors with higher metastatic potential can arise from both differentiated and undifferentiated keratinocytes. Furthermore, the described cell-of-origin subclasses were found to display distinct proliferative and invasive characteristics and were further validated using single-cell transcriptomics and methylomics.

Additionally, this work also made use of single-cell transcriptomics to characterize human dermal fibroblast diversity. Importantly, four distinct fibroblast populations with clearly distinctive functional annotations and spatial localization were identified. Furthermore, the loss of this functional fibroblast priming and the reduced communication of aged fibroblasts with other skin cells were identified as prominent aging-associated phenotypes in the human skin.

Altogether, this thesis thus provides novel insight into KC initiation and proposes a general stratification strategy for epidermal tumors that might improve patient risk assessment, as well as describes relevant fibroblast populations in human skin and provides novel insight into their role in skin aging.



## 6. Materials and Methods

### 6.1. Materials

#### 6.1.1. Table M1. Chemicals, reagents, and enzymes

Name	Company	Reference
Agarose	Roth	3810.3
AMPure XP beads	Beckman Coulter	A63881
BSA Type H1	Gerbu	1063.025
CaCl <sub>2</sub>	Fisher Chemical	C/1500/53
Citric acid	Fisher Chemical	5949-29-1
cOmplete™, Mini, EDTA-free Protease-Inhibitor-Cocktail	Roche	11836170001
D1000 Reagents	Agilent Technologies	5067-5583
D5000 Reagents	Agilent Technologies	5067-5590
DAPI	Thermo Fisher Scientific	D1306
Digitonin (5%)	Invitrogen	BN2006
DL-Dithiothriol (DTT) solution	Sigma-Aldrich	43816
DNA Gel Loading Dye (6X)	Thermo Fisher Scientific	R0611
dNTP mix 100 mM (25 mM each)	VWR	6100850-1000
EB Buffer	Qiagen	19086
Ethanol absolute	Sigma-Aldrich	32205-M
Formaldehyde solution (w/v) methanol-free (16%)	Thermo Fisher Scientific	28906
Glycerol 99%	Sigma-Aldrich	15523
Glycine	Gerbu	1023.1
HCl (37%)	Sigma-Aldrich	30721-M
HEPES	Gerbu	1009
Hyperladder 50 bp	Bioline	BIO-33054
IGEPAL® CA-630	Sigma-Aldrich	I3021
KAPA HiFi HotStart Ready Mix (2X)	Roche	7958935001
KCl	Roth	6781.1
Klenow(3'-5'exo-), 50000 U/ml	Biozym	280310
MACS® Tissue Storage Solution	Miltenyi Biotec	130-100-008
MgCl <sub>2</sub>	Sigma-Aldrich	M8266
N,N-Dimethylformamid (DMF)	Sigma-Aldrich	D4551
NaCl	Fischer Chemical	S/3160/60
NaOH	AppliChem	UN1824
NEBuffer™ 2.1	New England BioLabs	B7202S
Normal Goat Serum	Dako	X0907
Opal™ 520 Reagent Pack	Akoya Biosciences	FP1487001KT
Opal™ 570 Reagent Pack	Akoya Biosciences	FP1488001KT

## 6. Materials and Methods

---

Opal™ 690 Reagent Pack	Akoya Biosciences	FP1497001KT
PBS(10X), pH 7.4	Gibco	70011044
PitStop 2	Sigma-Aldrich	SML1169
Prolong® Gold Antifade Mountant	Thermo Fisher Scientific	P36930
RPMI 1640	Gibco	61870-010
Sodium citrate	Sigma-Aldrich	71405
Sodium dodecyl sulfate (SDS)	Sigma-Aldrich	71725
Sucrose	MP Biomedicals	821713
Tn5 Transposase	Obtained from EMBL	-
TO-PRO™-3 Iodide (642/661)	Thermo Fisher Scientific	T3605
Triton X-100	Gerbu	2999.005
Trizma®-Base	Sigma-Aldrich	T1503
Tween® 20	Sigma-Aldrich	P1379
UltraPure™ DNase/Rnase-Free Distilled H <sub>2</sub> O	Invitrogen	10977-035
Unmethylated Lambda DNA	Promega	D152A
Vectashield® Antifade Mounting Medium	Vector laboratories	H-1000-10
Xylene	VWR International	8080.5000

---

### 6.1.2. Table M2. Commercial Kits

Kit name	Company	Reference
Chromium Single Cell 3' Reagents (v2)	10X Genomics	PN-120237
Dead Cell Removal Kit	Miltenyi Biotec	130-090-101
Epidermis Dissociation Kit, human	Miltenyi Biotec	130-103-464
EpiTect® Bisulfite Kit	Qiagen	59104
EZ-96 DNA Methylation-direct MagPrep Kit	Zymo Research	D5044
Nextera® XT Index Kit v2 Set A	Illumina	FC-131-1001
Nuclei EZ-Prep Kit	Sigma Aldrich	NUC101-1KT
PyroMark PCR kit	Qiagen	978703
QIAamp® DNA Investigator Kit	Qiagen	56504
QIAamp® DNA Micro Kit	Qiagen	56304
QIAquick® Gel Extraction kit	Qiagen	28704
Qubit™ dsDNA HS Assay Kit	Thermo Fisher Scientific	Q32851
RNAScope® Multiplex Fluorescent Detection Kit v2	ACD Bio	323110
Whole skin Dissociation Kit, human	Miltenyi Biotec	130-101-540

---

**6.1.3. Table M3. Consumables**

<b>Name</b>	<b>Company</b>	<b>Reference</b>
8-strip PCR Microtubes 0.2 ml	4titude	4ti-0792
Adhesive PCR foil seal	VWR International	732-3218
AdhesiveCap 0.5ml tube	Zeiss	415190-9201-000
Cell Counting slides for TC10™	Bio-Rad	1450011
Cell Culture Dish 100/20mm	Greiner Bio-One	664160
Cell Strainer (40 µm)	Corning	15360801
Cell Strainer (70 µm)	Corning	15370801
Chromium Single Cell A Chip	10X Genomics	PN-1000009
Cover Slips	Thermo Fisher Scientific	004711180
Cryotube Vials	Thermo Fisher Scientific	377267
D1000 ScreenTape	Agilent Technologies	5067-5582
D5000 ScreenTape	Agilent Technologies	5067-5588
FACS Tubes	Corning	352235
Filtropur S 0.2 sterile filters (0.2 µm)	Sarstedt	83.1826.001
gentleMACS™ C Tubes	Miltenyi Biotec	130-093-237
MembraneSlide NF 1.0 PEN	Zeiss	415190-9081-000
Microcentrifuge tube 1.5 ml	Sarstedt	72.690.001
Microcentrifuge tube 2 ml	Sarstedt	72.691
Microseal 'B' PCR Plate Sealing Film	Bio-Rad	MSB1001
MS Columns	Miltenyi Biotec	130-042-201
Optical Cap 8x strip	Agilent Technologies	401425
Optical Tube 8x strip	Agilent Technologies	401428
Qubit™ assay tubes	Thermo Fisher Scientific	Q32856
Reagent reservoir	Gilson	GILSF267660
DNA LoBind® Tubes (1.5 ml)	Eppendorf	022431021
SuperFrost® Plus Slides	Thermo Fisher Scientific	J1800AMNZ
Tube 15 ml Cellstar®	Greiner Bio-One	188271
Tube 50 ml Cellstar®	Greiner Bio-One	227261
twin.tec® PCR plate 96 semi-skirted	Eppendorf	0030128575
twin.tec® PCR plate 96 skirted	Eppendorf	0030128648

**6.1.4. Table M4. Equipment and devices**

Name	Company
Agarose gel electrophoresis chamber	Bio-Rad
Agilent 2200 TapeStation System	Agilent Technologies
Agilent 4200 TapeStation System	Agilent Technologies
Automated Cell Counter TC10	Bio-Rad
BD FACSAria™ III Cell Sorter	BD Biosciences
Cell Culture hood HERA safe KS	Thermo Fisher Scientific
Centrifuge 5804/5804 R	Eppendorf
Chromium Single Cell Controller	10X Genomics
CO <sub>2</sub> incubator MC0-20AIC	Sanyo
E1-ClipTip Electronic Pipette	Thermo Fisher Scientific
EK-200i Compact Balance	A&D weighing
EV265/EV835 Electrophoresis Power Supply	Consort
FirstLight UV Illuminator	Analytik Jena
gentleMACS® Dissociator	Miltenyi Biotec
Heraus Fresco 17 microcentrifuge	Thermo Fisher Scientific
HI2211 Basic pH/Redox/°C Meter	Hanna instruments
HybEZ™ hybridization oven	ACDBio
Intelli-mixer rotator	neoLab
Leica HistoCore Arcadia H Paraffin Embedding Station	Leica Microsystems
Leica TCS SP5 microscope	Leica Microsystems
Luna™ Automated Cell Counter	Logos Biosystems
MACS Multistand	Miltenyi Biotec
MiniMACS™ Separator	Miltenyi Biotec
Qubit™ 4 Fluorometer	Thermo Fisher Scientific
Rotatory Microtome HM 355S	Thermo Fisher Scientific
Stereo Microscope Lightning KL1500 Compact	Olympus
Stereo Microscope System SZX10	Olympus
Thermocycler C1000 Touch	Bio-Rad
Thermocycler PTC-200	MJ Reasearch
Thermocycler T100	Bio-Rad
Thermocycler T3000	Biometra
Thermomixer compact	Eppendorf
Transluminator U:GENIUS3	Syngene
TW12 Water bath	Julabo
Vortex REAX top	Heidolph instruments
Zeiss Axio Vert.A1 microscope with AxioCam ICm1	Zeiss
ZeissPALM MicroBeam	Zeiss

## 6.1.5. Table M5. Buffers and solutions

Name	Application	Composition
Buffer A	sci-MET sample preparation	10 mM HEPES pH 7.9 10 mM KCl 1.5 mM MgCl <sub>2</sub> 0.34 M Sucrose 10% Glycerol 1 mM DTT 1X Protease inhibitor cocktail
Nuclear Isolation Buffer	sci-MET sample preparation	10 mM TrisHCl pH 7.5 10 mM NaCl 3mM MgCl <sub>2</sub> 0.1% IGEPAL-CA-630 1X Protease inhibitor cocktail
Tagmentation Buffer 1 (for cell lines)	sci-MET tagmentation	10mM Tris-HCl pH 7.5 10 mM MgCl <sub>2</sub> 25% DMF
Tagmentation Buffer 2 (for primary human epidermis)	sci-MET tagmentation	10 mM Tris-HCl pH 7.5 10 mM MgCl <sub>2</sub> 10% DMF 70 µM PitStop2
TAE 50X	Targeted amplicon sequencing (gel electrophoresis)	40 mM Tris 2 mM EDTA 20 mM acetic acid  <i>Adjusted to pH 8.5</i>
Citrate buffer pH 6.0	Immunofluorescence assays	10 mM Citric acid 0.05% Tween-20  <i>Adjusted to pH 6.0</i>
Blocking solution	Immunofluorescence assays	10% Normal Goat Serum  <i>in 1X PBS</i>
Permeabilization solution	Immunofluorescence assays	0.4% Triton X-100 1% Normal Goat Serum  <i>in 1X PBS</i>
Blocking solution (for Collagen XVIII)	Immunofluorescence assays	1% BSA 22.52 mg/ml glycine 0.1% Tween-20  <i>in 1X PBS</i>
5X SCC	mRNA-FISH	750 mM NaCl 75 mM sodium citrate  <i>Adjusted to pH 7.0</i>

**6.1.6. Table M6. Antibodies and mRNA probes**

<b>mRNA probes</b>	<b>Company</b>	<b>Reference</b>	
RNAScope® Probe-Hs-CTHRC1	ACDBio	413331	
RNAScope® Probe-Hs-APCDD1	ACDBio	535851-C2	
RNAScope® Probe-Hs-CCL19	ACDBio	474361	
RNAScope® Probe-Hs-APOE	ACDBio	433091-C2	
RNAScope® Probe-Hs-ASPN	ACDBio	404481	
RNAScope® Probe-Hs-CD248	ACDBio	404481	
RNAScope® Probe-Hs-PDGFRA	ACDBio	604481-C3	
<b>Antibodies</b>	<b>Company</b>	<b>Reference</b>	<b>Dilution</b>
$\alpha$ -Periostin	Santa Cruz	sc-398631	1:100
$\alpha$ -Tetraspanin 8	Abcam	Ab230448	1:200
$\alpha$ -Vimentin	Cell Signaling	D21H3	1:100
$\alpha$ -Vimentin	Abcam	Ab24525	1:2000
$\alpha$ -Collagen XVIII	Kind gift from Ritva Heljasvaara (University of Oulu, Finland)		1:50
goat anti-rabbit Alexa Fluor 488	Thermo Fisher Scientific	A11034	1:500
goat anti-rabbit Alexa Fluor 555	Thermo Fisher Scientific	A32732	1:500
goat anti-chicken Alexa Fluor 633	Thermo Fischer Scientific	A21103	1:500

**6.1.7. Table M7. Cell lines**

<b>Name</b>	<b>Cell Type</b>	<b>Source</b>
129/Ola ESCs	Mouse Embryonic Stem cells	scOpenLab
HaCaT	Spontaneously immortalized human keratinocytes	DKFZ

**6.1.8. Table M8. Cell Culture reagents**

<b>Reagent</b>	<b>Company</b>	<b>Reference</b>
Chelex® 100 Chelating Resin, analytical grade, 100–200 mesh, sodium form, 500 g	Bio-Rad	1422832
DMEM, high glucose, no glutamine, no calcium	Gibco	21068028
DMSO	Sigma-Aldrich	D2650
Fetal Bovine Serum	Gibco	10500-064
L-Glutamine (200 mM)	Gibco	A2916801
Penicillin-Streptomycin (10,000 U/mL)	Gibco	15140122
Powerstem ESPRO1	PAN Biotech	P04-77010K
Trypsin 0.25%	Gibco	25200-056

## 6.1.9. Table M9. Primers and oligos

Application	Name	Sequence (5'-->3')
Targeted Amplicon Sequencing	Region 1 F	<b>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG</b> TAGTTTATGTTTTGG TTTTAGGTTG
	Region 1 R	<b>GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG</b> AAACTACAACATA TAACTACTCCCC
	Region 2 F	<b>TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG</b> TTTTTTTTAGAAAGAG TTTTTGAGGAG
	Region 2 R	<b>GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG</b> CTCATAAAATCCT CTCCCTACCTC
Tn5 activity test	FC121-1030	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
	FC121-1031	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
	Tn5MERev	[phos]CTGTCTCTTATACACATCT
	P5	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTC
	P7	CAAGCAGAAGACGGCATAACGAGATGTCTCGTGGGCTCGG
sci-MET (Library preparation)	9-N Random Primer 96 Transposase-loaded Oligos sci-MET_i7 (12) sci-MET_i5 (8)	Obtained from Mulqueen et al. (2018)
sci-MET (Sequencing)	sci-MET_Read1	+TGGT+AGAG+AGGG+TGAGATGTGTATAAGAGATAG
	sci-MET_Index1	AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC
	sci-MET_Index2	+CTA+TCT+CTT+ATA+CAC+ATC+TCACCCTCTCTACCA
	PhiX_Read1	ACACTCTTCCCTACACGACGCTCTTCCGATCT
	PhiX_Index1	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

**In green:** adapter sequences needed for sequencing with the MiSeq platform.

[phos]: phosphorylated

+: LNA modification

## 6.1.10. Table M10. Software and computational tools

Software	Version	Purpose
Adobe Illustrator	25.2.2	Figure assembly
CellPhoneDB	2.0.0	Cell-cell interaction prediction in scRNA-seq
BD FACSDiva	9.0	Nuclei sorting
DAVID	6.8	GO analyses
Fiji/Image J	1.52i & 1.53c	Image processing
FlowJo	10.7.2	FANS graph plotting
Leica LAS AF	1.6.3	Confocal image acquisition
Microsoft Office Standard 2016	16.0.5122.1000	Data analysis and Thesis writing
Primer3	4.1.0	Primer design
R	3.5.1 & 4.0.2	Bioinformatic analyses, plot generation
Zen	Blue 2011	Image acquisition

## 6.2. Methods

### 6.2.1. Clinical Samples

Clinical samples used in this thesis were obtained from patients undergoing routine surgery at the Department of Dermatology, University Hospital of Heidelberg, after written informed consent and as approved by the Ethics Committee of Heidelberg University (S-091/2011). Data protection and privacy of participating patients were ensured by handling samples and data in a pseudonymized manner, in compliance with current legislation.

All tumor samples included were diagnosed by a trained dermatohistopathologist at the University Hospital of Heidelberg. Also, tumor histopathological features were routinely examined and reviewed before inclusion in this study to ensure proper sample selection.

For the different experiments with healthy skin samples, the absence of skin diseases and skin-affecting co-morbidities was confirmed by conducting a complete body examination by a dermatologist to all donors before resection and a thorough revision of their medical records. None of the patients included in this study showed clinical evidence or history of inflammatory or systemic skin diseases such as systemic sclerosis or lupus erythematosus. Likewise, there was neither a record of skin-affecting co-morbidities (i.e., chronic immunosuppression, chronic renal failure). In samples collected from sun-protected body areas, there was no record of UV sensitivity of the skin, acute or chronic actinic skin damage at the time of resection.

#### 6.2.1.1. For bulk methylation analyses

A total of 4 AK, 14 BCC, 11 BD, 17 cSCC, and 10 SK samples were collected as punch biopsies (4 mm) from the center of the lesion, directly fresh-frozen by immersion in liquid nitrogen and stored at -80°C until further sample processing. Furthermore, the 12 healthy epidermis, 16 AK, and 18 cSCC samples previously analyzed were also included in this study (Rodríguez-Paredes *et al.*, 2018). As described in the original publication, healthy epidermal samples were collected using suction blisters on the forearms of male donors (Rodríguez-Paredes *et al.*, 2018). In contrast, the AK and cSCC samples were obtained as punch biopsies (4 mm), comparably to the samples acquired at the Department of Dermatology, University Hospital of Heidelberg.

To ensure high tumor purity, the epidermal part of the tumor samples was isolated from the dermis by heat split. In brief, FF samples were immersed in pre-warmed PBS at 37°C for one minute and subsequently transferred to pre-warmed PBS at 56°C for 5 minutes. Then, the epidermis was carefully dissected manually under a magnifying glass. Samples in which

epidermis separation was not successful or unclear were discarded from the study. DNA was isolated using the QIAamp® DNA Investigator Kit (Qiagen) following the manufacturer's instructions.

FFPE primary metastasizing cSCC and cSCC metastases were obtained as 7µm sections placed on MembraneSlide NF 1.0 PEN (Zeiss) slides. Sections were then deparaffinized in xylene (5 minutes, two times) and rehydrated in a gradient of ethanol (100% ethanol for 1 minute, 96% ethanol for 1 minute, and 70% ethanol for 1 minute). Tumor tissue was subsequently isolated by laser microdissection using the Zeiss PALM MicroBeam system (Zeiss), and DNA was extracted using the QIAamp® DNA Micro Kit (Qiagen) following the manufacturer's instructions.

#### **6.2.1.2. For targeted amplicon bisulfite sequencing**

Extracted genomic DNA from 10 FF tumor samples previously analyzed by Infinium MethylationEPIC BeadChips (Rodríguez-Paredes *et al.*, 2018) was used for validation of the two target regions.

An independent dataset consisting of 17 primary cSCC samples was obtained as 10 µm FFPE sections (2 sections per tumor). The epidermal part of the tumor was isolated by manual dissection under a magnifying glass by Dr. Daniel Hasche (Division of Viral Transformation Mechanisms, DKFZ). Genomic DNA from isolated tissue was extracted using QIAamp® DNA Investigator Kit (Qiagen) following the manufacturer's instructions for FFPE samples.

#### **6.2.1.3. For sci-MET experiments**

Surgical remnant, clinically healthy whole human skin samples, not required for diagnostics, were used to optimize the sci-MET protocol. Two sample types were collected: fresh and FF. Fresh samples were obtained as approximately 1 cm skin specimens from sun-protected body regions (axillary or ilioinguinal) from male and female donors of diverse ages and immersed in MACS Tissue Storage Solution (Miltenyi Biotec). FF samples consisted of skin specimens from sun-exposed areas (head, neck, and face) from male and female donors of diverse ages. Donor characteristics for samples used in final sci-MET experiments are shown in Table S1.

#### **6.2.1.4. For scRNA-seq experiments**

Surgical remnant, clinically healthy whole skin samples, not required for diagnostics, were collected from the sun-protected ilioinguinal region of fair-skinned male donors. Immediately after resection, samples were immersed in MACS Tissue Storage Solution (Miltenyi Biotec) and kept

on ice for no longer than one hour. Then, fat was carefully removed, and skin tissue was cut in 4 mm punch biopsies and further processed.

For whole skin scRNA-seq experiments, two young donors (25 and 27 y/o) and three older (53, 70, and 69 y/o) were included. Samples were enzymatically and mechanically dissociated using the Whole Skin Dissociation Kit for human material (Miltenyi Biotec) and the gentleMACS® Dissociator (Miltenyi Biotec), following the manufacturer's instructions. Donor characteristics and sequencing statistics are shown in Table S2.

For only epidermis scRNA-seq experiments, a sample obtained from a 30 y/o donor was enzymatically and mechanically dissociated using the Epidermis Dissociation Kit for human material (Miltenyi Biotec), following the manufacturer's instructions.

Apoptotic and dead cells were removed using the Dead Cell Removal Kit (Miltenyi Biotec) before continuing with single-cell capture and library preparation.

### **6.2.1.5. For mRNA-FISH or IF**

FFPE skin samples from old and young donors were used for IF and mRNA-FISH experiments. Young skin samples were either purchased as 4 µm sections from Genoskin (France) or obtained as surgical leftovers from the Department of Dermatology, University Hospital of Heidelberg. All young samples were taken from sun-protected body regions of healthy (male and female) subjects with no present co-morbidities. All old skin samples were obtained as excess surgical skin from the Department of Dermatology of Heidelberg University Hospital. Old skin specimens were taken only from male donors and represented both sun-protected and sun-exposed body areas.

### **6.2.2. Cell Culture**

Mouse embryonic stem cells (129/Ola ESCs) were cultured in Powerstem ESPRO1 medium (PAN Biotech) on gelatin-coated plates. Leukaemia inhibitory factor (LIF) was produced in HEK cells and added freshly to the medium. Dr. Michelle Liberio performed the mESCs cell culture at the scOpenLab (DKFZ).

HaCaT cells were cultured in calcium-free DMEM medium (Gibco) supplemented with 10% chelexed fetal bovine serum (FBS), 4 mM L-Glutamine (Gibco), 0.03 mM CaCl<sub>2</sub> and 1% penicillin/streptomycin (Gibco)

Both cell lines were cultured under standard 37°C and 5% CO<sub>2</sub> conditions. On the day of the experiment, cells were washed once with 1X PBS and subsequently detached from plates using

Trypsin 0.25% (Gibco). Trypsinized cells were then used to optimize the sci-MET protocol under different conditions.

### 6.2.3. DNA methylation profiling and data analysis

Genomic DNA from tumor samples was extracted as described in section 6.2.1.1, and methylation profiles were obtained using Infinium MethylationEPIC BeadChips (Illumina) at the Genomics and Proteomics Core Facility (DKFZ), according to the manufacturer's instructions. FFPE samples were subjected to a modified protocol that restores degraded DNA to an appropriate state for amplification and hybridization with Infinium arrays. Data analysis was performed using the R Bioconductor package `minfi` (version 1.34.0) (Aryee *et al.*, 2014) and following the bioinformatics pipeline designed by Dr. Julian Gutkunst (Division of Epigenetics, DKFZ). In brief, low-detected ( $P > 0.01$ ) and self-hybridizing CpG probes, as well as probes located in sex chromosomes or SNP positions, were filtered out of the raw sequencing data. Next, between-array data normalization was performed using the `preprocessFunnorm()` function from the `minfi` package. Differentially methylated probes were identified by fitting a linear model and using an empirical Bayes method for statistical analysis with correction for multiple testing (Benjamini-Hochberg method). Only CpG probes with a  $P < 0.05$  (F-test) were selected as differentially methylated.

Tumor purity and leukocyte content was estimated using `MethylResolver` R package (version 0.1.0) with default parameters (Arneson, Yang and Wang, 2020).

R package `methylGSA` (version 1.6.1) was used for performing GO analyses with genes displaying differentially methylated promoter regions (Ren and Kuan, 2019). As promoter annotation, CpGs corresponding to the UCSC reference locations TSS1500, TSS200, 1stExon, and 5'UTR were used. CpG locations were obtained from the Illumina annotation contained in the `IlluminaHumanMethylationEPICanno.ilm10b4.hg19` R package (Hansen, 2017).

Mitotic age and stem cell division rates (SCDR) were calculated using the publicly available `epiTOC2` R Script with default parameters (Teschendorff, 2020).

### 6.2.4. Keratinocyte enhancer methylation analysis

A set of EpSC-specific enhancers known to be lost upon epidermal differentiation ( $n=11,825$ ) was described in a previous study (Rinaldi *et al.*, 2016). Tumor stratification based on the methylation patterns of this set of enhancers was performed as previously described (Rodríguez-Paredes *et al.*, 2018). First, CpG probes contained in the Illumina Methylation EPIC BeadChips and located within the enhancer regions were identified ( $n=96,424$ ). Then, methylation values for CpGs

located in the same enhancer region were averaged in each sample to give one methylation value per enhancer. Averaged enhancer methylation values of the 12 healthy epidermis samples were subtracted to visualize methylation differences better. Tumor samples were then clustered by complete-linkage hierarchical clustering using Euclidean distances and visualized as heatmaps.

### 6.2.5. Targeted amplicon bisulfite sequencing

Genomic DNA from FF or FFPE tumors was extracted as described in sections 6.2.1.1 and 6.2.1.2. Extracted DNA was bisulfite-converted using the Epitect® Bisulfite Kit (Qiagen), following manufacturer's instructions with a longer incubation time:

Denaturation	5 min	95°C
Incubation	25 min	60°C
Denaturation	5 min	95°C
Incubation	85 min (1h 25min)	60°C
Denaturation	5 min	95°C
Incubation	295 min (4h 55min)	60°C
Hold	∞	20°C

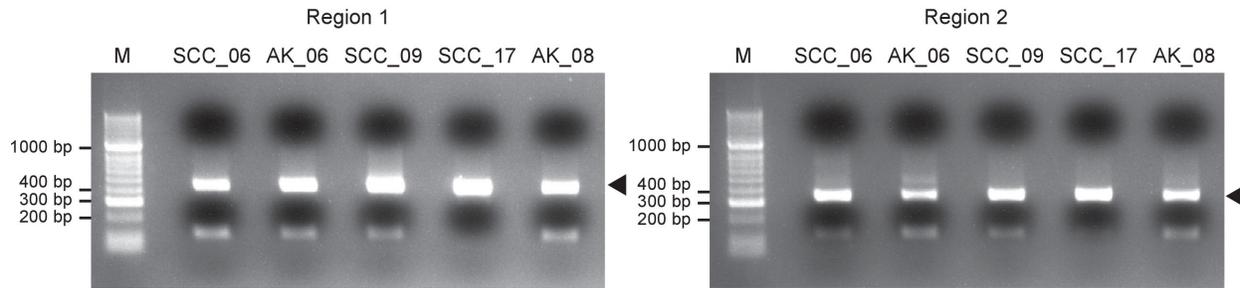
Bisulfited DNA was cleaned following the manufacturer's instructions and eluted in 20 µl ddH<sub>2</sub>O. PCR amplification of the selected regions was carried out using the PyroMark PCR kit (Qiagen) in a reaction containing 12.5 µl PyroMark PCR Master Mix (2X), 2.5 µl Coral Load Concentrate (10X), 0.5 µl PrimerF/PrimerR (mixed, 10µM), 10-20 ng bisulfited DNA and ddH<sub>2</sub>O up to 25 µl. The thermocycler program used was as follows:

Initial PCR activation step	15 min	95°C	} 45 cycles
Denaturation	30 sec	94°C	
Annealing	30 sec	56°C	
Extension	30 sec	72°C	
Final Extension	10 min	72°C	
	∞	4°C	

PCR amplicon length was checked by electrophoresis with a 1.8% agarose (in 1X TAE) gel (Figure M1). Correct bands were isolated and gel-purified using the QIAquick® Gel Extraction kit (Qiagen), following the manufacturer's instructions. Purified amplicons were eluted in 20 µl ddH<sub>2</sub>O, from which 5 µl were subsequently used for indexing with Nextera® XT Index Kit v2 Set A (Illumina) in a PCR reaction containing 25 µl KAPA HiFi HotStart ReadyMix (2X), 5 µl Nextera XT Index 1 (N7XX), 5 µl Nextera XT Index 2 (S5XX) and 10 µl ddH<sub>2</sub>O.

The thermocycler program used was as follows:

Initial denaturation	3 min	95°C	} 8 cycles
Denaturation	30 sec	95°C	
Annealing	30 sec	55°C	
Elongation	30 sec	72°C	
Final elongation	5 min	72°C	
	∞	4°C	



**Figure M1. PCR products for targeted amplicon bisulfite PCR.** Illustrative examples of amplicons for Region 1 (left; 436bp) and Region 2 (right; 328bp), generated from DNA of fresh-frozen tumor samples. M = marker (Hyperladder II)

Indexed amplicons were then purified with AMPureXP beads (Beckman Coulter) and quantified using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific). All amplicons were then pooled in an equimolar manner to generate a sequencing library at 10 nM in 130 µl, which was sequenced on a MiSeq V2 system (paired-end 150 bp Nano protocol) at the High Throughput Sequencing Core Facility (DKFZ). Sequencing raw data was analyzed with BisAMP (Bormann *et al.*, 2019), with more permissive parameters for mapping reads to the reference sequence (parameter *v* set to 1 and alpha set to 0). Results were visualized as heatmaps displaying reads that passed quality control in either Read1 or Read2 (Figure S3).

## 6.2.6. scRNA-seq of human epidermis

### 6.2.6.1. Library preparation and sequencing

A single-cell suspension of healthy human epidermis from a 30 y/o fair-skinned male donor was obtained as described in section 6.2.1.4. Approximately 20,000 cells were used for generating a sequencing library with the Chromium Single Cell Controller and Chromium Single Cell 3' Reagents (v2) from 10X Genomics, according to manufacturer's guidelines. Briefly, single-cell suspensions were mixed with retrotranscription reagents and loaded into a Chip A Single Cell

(10X Genomics) previously loaded with Single-Cell 3' Gel Bead suspension and Partitioning Oil in their respective wells. Then, the chip was inserted in a Chromium Single Cell Controller (10X Genomics) for cell capture into nanoscale droplets containing both reverse transcription reagents and a gel bead, named gel bead-in-emulsions (GEMs). Each Single Cell 3' v2 Gel Bead carries a specific 10X Genomics barcode, an Illumina R1 sequence, a Unique Molecular Identifier (UMI), and a poly-dT primer sequence. Cells were then lysed, and poly-A-containing mRNAs were attached to the beads. GEMs were subsequently transferred to a thermocycler for retrotranscription of the mRNA into full-length cDNA, which contained a unique barcode per cell and a unique UMI per transcript. Next, cDNA was amplified and further processed by fragmentation, end-repair, and A-tailing double-sided size selection using AMPure XP beads (Beckman Coulter). Lastly, Illumina sequencing adaptors and a sample index were added through a PCR step, and library fragment size was adjusted by double-sided size selection (0.6X-0.8X) using AMPure XP beads (Beckman Coulter). Libraries were quantified using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific), and cDNA integrity and fragment size were assessed using D100 ScreenTapes (Agilent Technologies). Libraries were sequenced in a HiSeq 4000 device (Illumina) as a paired-end (26+74 bp) for 100 cycles at the High Throughput Sequencing Core Facility (DKFZ).

### 6.2.6.2. Data analysis

Cell Ranger software (version 2.1.0) from 10X Genomics was used for raw sequencing data processing, including read demultiplexing and alignment, barcode assignment, and UMI quantification. Quality control performed with Cell Ranger resulted in a dataset containing 7,752 single-cell transcriptomes. These steps of the data analysis were performed by Dr. Günter Raddatz (Division of Epigenetics, DKFZ).

Seurat package version 3.1.1 was used for the downstream analysis of the data (Stuart *et al.*, 2019). A second quality control performed with this package further filtered out low-quality cells (cells expressing less than 200 or more than 2,500 genes, and cells expressing more than 5% mitochondrial genes). This resulted in a final dataset containing 7,143 single-cell transcriptomes. Standard cell clustering and visualization were performed with Seurat as follows. First, gene expression was scaled using the `ScaleData()` function with default parameters, and PCA dimensions were calculated with the `RunPCA()` function with the 2,000 most variable genes. Then, the first 20 PCA dimensions were used as an input for the `FindNeighbors()`, and a resolution of 0.5 was set in the `FindClusters()` function for clustering the cells in an unsupervised manner. Clustering was achieved by the latter function using a SNN modularity optimization-based

clustering algorithm. For dimensional reduction and visualization of the single-cell data, the RunUMAP() function was used with default parameters and 20 PCA dimensions as input. Unsupervised clustering identified 10 cell clusters, and their most representative expressed genes were determined using the FindAllMarkers() function.

In subsequent analyses, this sample was combined with another healthy sun-protected epidermis scRNA-seq dataset containing three abdominal samples (Cheng *et al.*, 2018). To do so, the standard integration protocol included in the Seurat package was used (Stuart *et al.*, 2019). In brief, preprocessing of each dataset was performed by log-normalizing its UMI counts and by identifying its 2,000 most variable genes before data integration. Then, the FindIntegrationAnchors() function was used to carry out pairwise comparisons between all samples, identifying correspondences between them, named anchors. These anchors were subsequently used to integrate all samples with the IntegrateData() function. Default parameters with 30 Canonical Correlation Analysis (CCA) dimensions were used for both FindIntegrationAnchors() and IntegrateData() functions. After integration, gene expression was scaled, PCA dimensions were calculated, and cells were clustered as mentioned above, using the 2,000 most variably expressed genes in the integrated dataset. In this case, 30 PCA dimensions and a resolution of 0.4 were used in the FindNeighbors() and FindClusters() functions, respectively. Clusters were visualized as a UMAP, calculated with the RunUMAP() function with 30 PCA dimensions. A total of 13 cell clusters were identified, and their most representative genes were determined using the FindAllMarkers() function on the uncorrected expression value.

### 6.2.6.3. RNA velocity analysis

Spliced and unspliced reads from the in-house scRNA-seq dataset were obtained by running the command line interface of velocity (version 0.17) (la Manno *et al.*, 2018). Data was then preprocessed by normalization, log-transformation and identification of highly variable genes, before calculating the RNA velocity. Additionally, melanocytes and immune cells were removed from the dataset, resulting in 7,068 keratinocytes that were used for further analysis. RNA velocity was then estimated by the generalized dynamical model of scVelo (version 0.2.3) (Bergen *et al.*, 2020), using the recover\_dynamics() function with 100 maximum iterations. Lastly, RNA velocities were projected and visualized onto the UMAP embedding calculated by Seurat. The latent time was calculated using the default values and manually assigning the root cells to cluster Basal 1, based on the RNA velocity results. This analysis was performed by Oliver Gilliam (Division of Epigenetics, DKFZ).

### **6.2.7. Bulk DNA methylation data deconvolution**

Cell type proportions in bulk DNA methylation profiles of epidermal tumors were estimated using the EpiSCORE R package (version 0.9.2) (Teschendorff *et al.*, 2020). The transcriptional profiles of 15,000 keratinocytes from the integrated scRNA-seq epidermis dataset (see Methods, section 6.2.6) were used to generate a gene expression reference matrix containing 784 cluster-specific marker genes, with a maximal specificity score of three. Only keratinocyte populations involved in the differentiation trajectory defined by RNA velocity were included in this analysis. The reference expression matrix was then validated on a downsampled dataset comprised of 10,000 keratinocytes, resulting in an overall cell identity prediction accuracy of 74%. To impute a reference DNA methylation matrix for human epidermis, EpiSCORE compared the reference expression matrix with two available datasets with paired gene expression and DNA methylation data (SCM2 and Epigenomics roadmap). The resulting DNA methylation matrix contained 74 genes and was subsequently used to deconvolute the cell fractions in the 102 bulk epidermal tumors methylation profiles.

### **6.2.8. Single-cell combinatorial indexing for methylation analysis**

#### **6.2.8.1. Sample preparation for cell lines**

mESCs or HaCaT cells were trypsinized and subsequently fixed in 0.7% formaldehyde (without methanol) (Thermo Fisher Scientific) diluted in 1X PBS for 10 minutes at room temperature with gentle shaking. Then, 80  $\mu$ l of 2.5 M Glycine were added, and cells were incubated for 5 minutes on ice to stop the reaction. Cells were centrifuged at 550 g for 10 minutes at 4°C. Nuclei were then isolated by resuspending the pellet in 200  $\mu$ l Nuclei EZ Lysis Buffer (Sigma-Aldrich) by vortexing for 15 sec, followed by addition of 800  $\mu$ l of Nuclei EZ Lysis Buffer (Sigma-Aldrich) and 5 minutes incubation on ice. Nuclei were then centrifuged at 550 g for 5 minutes at 4°C. The vortexing-incubation-centrifugation steps were repeated for a total of three cycles to ensure proper nuclear isolation.

#### **6.2.8.2. Sample preparation for human epidermis**

Freshly resected whole human skin samples were maintained in MACS® Tissue Storage solution (Miltenyi Biotec) after resection and until further processing. Samples were enzymatically and mechanically dissociated using the Epidermis Dissociation Kit for human material (Miltenyi Biotec), following the manufacturer's instructions. Cells were then fixed by incubation with 1.5%

formaldehyde (without methanol) (Thermo Fisher Scientific) in 1 ml 1X PBS for 10 minutes, with gentle shaking. The reaction was stopped by adding 80  $\mu$ l of 2.5 M glycine followed by 5 minutes incubation on ice. The sample was centrifuged at 550 g for 10 minutes at 4°C to recover fixed cells. A previously published protocol was used to isolate nuclei from epidermal cells (Wysocka, Reilly and Herr, 2001). Thus, isolated epidermal cells were resuspended at  $1 \times 10^6$  cells/25  $\mu$ l of Buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.34 M sucrose, 10% glycerol, 1 mM DTT, and 1X protease inhibitor cocktail). Triton X-100 was freshly added to Buffer A at a final concentration of 0.1%, and cells were incubated on ice for 10 minutes. Nuclei were recovered by centrifugation at 1,300 g, for 5 minutes at 4°C.

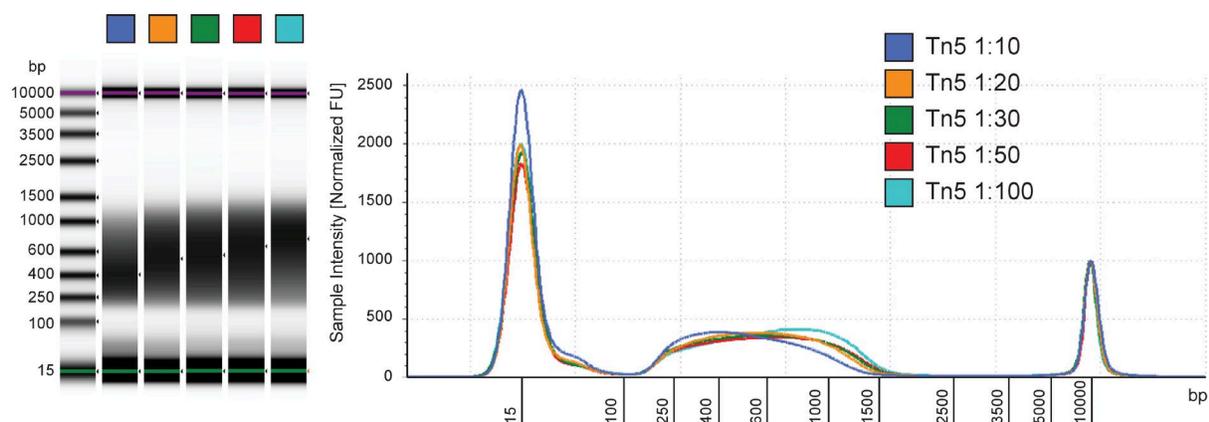
### 6.2.8.3. Assessment of transposase activity and transposome assembly

To test the activity of the homemade Tn5 used in sci-MET experiments, the protocol provided by the Protein Expression and Purification Core Facility at the EMBL (Heidelberg, Germany) was followed. First, linker sequences were resuspended at 100  $\mu$ M in EB buffer (Qiagen), and each forward primer (FC121-1030 and FC121-1031) was combined in a 1:1 ratio with the reverse primer (Tn5MERev). Primer sequences are summarized in Table M9. Then, oligonucleotides were denatured by incubation at 95°C for 5 minutes and subsequently annealed by gentle cooling to 65°C (-0.1°C/s), 5 minutes incubation at 65°C and gentle cooling to 4°C (-0.1°C/s).

Tn5 transposome assembly was performed by adding 0.5  $\mu$ l of each annealed linker to 10  $\mu$ l of undiluted Tn5. The mixture was incubated in a thermocycler at 23°C for one hour and subsequently diluted to test the appropriate working concentration. Tagmentation was performed in a 5  $\mu$ l reaction containing 2.5  $\mu$ l 2X Tagmentation Buffer 1 (20 mM Tris-HCl pH 7.5, 20 mM MgCl<sub>2</sub>, 50% DMF), 1.25  $\mu$ l of diluted Tn5 and 150 pg of cDNA as input, by incubating samples at 55°C for 3 minutes. Then, reactions were cooled down to 10°C and subsequently stopped by adding 1.25  $\mu$ l of 0.2% SDS and incubating for 5 minutes at room temperature. Directly after incubation, 10  $\mu$ l of the following PCR mixture were added to each reaction: 6.75  $\mu$ l KAPA HiFi HotStart ReadyMix (Roche), 0.75  $\mu$ l DMSO, 1.25  $\mu$ l 10  $\mu$ M Illumina i7 adapter, and 1.25  $\mu$ l 10  $\mu$ M Illumina i5 adapter. Illumina adapter sequences are summarized in Table M9. Tagmented cDNA was amplified using the following PCR program:

72°C	3 min	
95°C	30 sec	
98°C	20 sec	} 12 cycles
58°C	15 sec	
72°C	30 sec	
72°C	3 min	
10°C	$\infty$	

PCR products were cleaned up by adding 1X AMPureXP beads (Beckman Coulter) and incubating at room temperature for 5 minutes, followed by two washes with 80% ethanol. Lastly, DNA was eluted in 10  $\mu$ l ddH<sub>2</sub>O and subsequently quantified using Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific), and fragment distribution was examined using D5000 ScreenTapes (Agilent Technologies). An illustrative example of Tn5 activity assessment is shown in Figure M2.



**Figure M2. Assessment of homemade Tn5 transposase activity.** Bioanalyzer D5000 ScreenTape gel images and electropherograms of tagmented and amplified cDNA. Different Tn5 dilutions were tested with no significant differences found in fragment size distribution.

The Tn5 dilution 1:50 displayed the best results in terms of yield and fragment size and was used for the assembly of the sci-MET transposomes. To do so, 96 unique cytosine-depleted linkers (Mulqueen et al., 2018) were resuspended at 100  $\mu$ M in EB Buffer (Qiagen). Each linker was mixed in a 1:1 ratio with the reverse complement primer, also at 100  $\mu$ M (Table M9). Oligonucleotides were annealed following the abovementioned thermocycler program and were subsequently diluted 1:50 in EB Buffer (Qiagen). Stock Tn5 was diluted 1:50 in 50% glycerol. One volume of each annealed and diluted linker was added to 10 volumes of the diluted Tn5 (i.e., 1  $\mu$ l diluted linker to 10  $\mu$ l diluted Tn5), and transposomes were assembled by incubation at 23°C in a thermocycler for 1 hour. Assembled transposomes were stored at -20°C for up to one month.

#### 6.2.8.4. sci-MET library preparation

Fixed nuclei were resuspended in 800  $\mu$ l 1X NEBuffer 2.1 (New England BioLabs) containing 0.3% SDS and were incubated for 30 minutes at 42°C with intense shaking. Then, 200  $\mu$ l of 10% Triton X-100 were added, and nuclei were incubated for additional 30 minutes at 42°C and intense shaking. Nuclei were collected by centrifugation (500 g, 5 minutes, 4°C), washed once in 1X PBS,

and filtered through a 40  $\mu\text{m}$  cell strainer before proceeding to fluorescence-activated nuclei sorting (FANS) using TO-PRO™-3 Iodide (1:10.000).

In a first step, 1,000-2,500 nuclei/well were sorted into a 96-well plate containing 10  $\mu\text{l}$  of 1X Tagmentation buffer diluted in Nuclear Isolation Buffer (10mM TrisHCl pH 7.5, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% IGEPAL® CA-630, 1X protease inhibitors cocktail) in a 1:1 ratio. Next, 4  $\mu\text{l}$  of a uniquely indexed transposome were added to each well for nuclei tagmentation. Tagmentation buffer and conditions varied depending on the sample type. Thus, cell lines (i.e. 129/Ola ESCs and HaCaT) were incubated at 55°C for 15 min in 1X Tagmentation Buffer 1 (10mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 25% DMF), whereas primary human epidermal cells were incubated at 55°C for 30 min in 1X Tagmentation Buffer 2 (10 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 10% DMF, 70  $\mu\text{M}$  PitStop2 (Sigma-Aldrich)). After tagmentation, all wells were pooled, mixed, and filtered through a 40  $\mu\text{m}$  cell strainer before adding fresh TO-PRO™-3 Iodide (1:10.000) for a second sorting step. A total of 22 nuclei/well were then sorted into a 96-well plate containing 5  $\mu\text{L}$  Zymo Digestion Reagent (2.5  $\mu\text{L}$  MDigestion Buffer, 2.25  $\mu\text{L}$  nuclease-free H<sub>2</sub>O, and 0.25  $\mu\text{L}$  Proteinase K), following manufacturer's instructions for EZ-96 DNA Methylation-Direct MagPrep Kit (Zymo Research). Three wells per plate were sorted with only 10 nuclei and 35 pg of unmethylated control Lambda DNA (Promega) previously barcoded with one of the unique transposomes were added as spike-in unmethylated controls. Nuclei were then digested by 20 minutes (cell lines) or 4 hours (primary epidermal samples) incubation at 50°C. Subsequently, 32.5  $\mu\text{l}$  Zymo CT Conversion Reagent were added to each well for bisulfite conversion following the manufacturer's instructions (98°C for 8 minutes, 64°C for 3.5 hours).

After the conversion was completed, 5  $\mu\text{l}$  of Zymo M-Binding Beads diluted in 150  $\mu\text{l}$  of M-Binding Buffer were added to each well, and the plate was incubated at room temperature for 5 minutes. Next, beads were washed with 80% ethanol in a magnetic rack followed by resuspension in 50  $\mu\text{l}$  of M-Desulphonation Buffer. Once DNA was desulphonated after a 15 minutes incubation at room temperature, beads were washed once with 80% ethanol in a magnetic rack, and DNA was eluted in 25  $\mu\text{l}$  of Zymo M-Elution Buffer. Elution was directly transferred into a new plate containing the necessary reagents for the first linear amplification round (16  $\mu\text{l}$  nuclease-free H<sub>2</sub>O, 5  $\mu\text{l}$  10X NEBuffer 2.1, 2  $\mu\text{l}$  10 mM dNTP mix and 2  $\mu\text{l}$  10  $\mu\text{M}$  9-Nucleotide Random Primers). DNA denaturation was achieved by incubating the plate at 95°C for 45 seconds in a thermocycler. Then, 10U of Klenow (3'→ 5' exo-) polymerase (Biozym) were added to each well, and the plate was incubated in a thermocycler for 5 minutes at 4°C; then, the temperature was increased by 1°C/15 seconds until 37°C, where the temperature was held for 90 minutes. This step was repeated a total of 4 times, adding fresh reagents (1.25 $\mu\text{l}$  4X NEBuffer 2.1, 1 $\mu\text{l}$  10  $\mu\text{M}$  9-Nucleotide

Random Primer, 1 µl 10 mM dNTP mix) and 10U Klenow(3'→5' exo-) per well. Amplified material was then purified with 1.1X AMPure XP beads (Beckman Coulter). DNA was eluted in 21 µl EB Buffer (Qiagen) and directly transferred into the plate containing 25 µl of 2X KAPA HiFi HotStart ReadyMix (Roche), 2 µl of 10 µM i7 index PCR primer, and 2 µl of 10 µM i5 index PCR for indexing PCR reaction in a thermocycler (95°C for 2 minutes, then 18-21 cycles of 94°C for 80 seconds, 65°C for 30 seconds and 72°C for 30 seconds). Libraries were then pooled together and purified using a double-sided size selection strategy with AMPure XP Beads (0.6X+0.8X). Library quantification was carried out using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific), and fragment size was assessed using D5000 and D1000 ScreenTapes (Agilent Technologies).

Libraries were sequenced in a NextSeq550 paired-end 150bp High-Output system following a custom-made recipe (Read 1: 100 imaged cycles; Index Read 1: 10 imaged cycles; Index Read 2: 11 imaged cycles, 16 dark cycles, and 10 imaged cycles) at the High Throughput Sequencing Core Facility (DKFZ). Primer sequences for library preparation and sequencing are shown in Table M9.

### **6.2.8.5. Data analysis for human epidermis sci-MET dataset**

The analysis outlined here corresponds to the sci-MET experiment described in the Results section 3.2.3. All single-cells displaying a minimum of 100,000 sequencing reads were kept for further analysis, which resulted in 185 cells. The reads corresponding to these cells were trimmed by removing stretches of bases with a quality score <30 at the end of the reads. The trimmed reads were mapped using bsmmap (Xi and Li, 2009). As a reference sequence for the bisulfite mapping, the hg19 assembly of the human genome was used. In order to achieve comparability with the binary nature of methylation values of single-cells, the beta-values of bulk methylomes were binarized by setting every value  $\geq 0.5$  to 1 and every value  $< 0.5$  to 0. The methylation data of all 185 single-cells were then combined with the binarized data of the 102 bulk DNA methylomes from KC and healthy epidermal samples and a 2-dimensional MDS analysis was performed. For the ChromHMM analysis, the Chromatin State Segmentation by HMM from ENCODE/Broad for epidermal keratinocytes (NHEK) was used (Ernst *et al.*, 2011), as provided by the UCSC genome server (<http://genome.ucsc.edu>). Methylation values of all CpGs located within each of the 15 types of genomic segments were averaged for each sci-MET cell type and EPIC tumor subclass. These analyses were performed by Dr. Günter Raddatz (Division of Epigenetics, DKFZ).

### 6.2.9. scRNA-seq of sun-protected healthy whole human skin

Single-cell suspensions from healthy whole human skin were obtained as described in section 6.2.1.4, and libraries were prepared and sequenced as described in section 6.2.6. A first quality control was performed by Cell Ranger (version 2.1.0), obtaining a total of 16,062 single-cell transcriptomes (performed by Dr. Günter Raddatz; Division of Epigenetics, DKFZ). Using Seurat, potential cell doublets and apoptotic cells were filtered out by removing cells with more than 7,500 expressed genes or more than 5% of their expressed genes corresponding to mitochondrial genes. These quality control steps resulted in a final dataset of 15,457 single-cell transcriptomes.

To correct potential batch effects in our data, samples were combined into a unique dataset using the standard integration workflow included in the Seurat package and as described in section 6.2.6.2. Default parameters with 30 CCA dimensions were used for both `FindIntegrationAnchors()` and `IntegrateData()` functions. Unsupervised clustering performed as described in section 6.2.6.2 with 20 PCA dimensions and resolution 0.4, identified 17 cell clusters. Downstream analysis of gene expression was performed using the uncorrected gene expression values as described in the Seurat package. The `FindAllMarkers()` function was used to determine genes with an enriched expression in each cell cluster. These representative genes were used for cell-type identification, in combination with well-established gene markers for known cell populations present in the human skin.

A second-level clustering was performed to further explore the cellular heterogeneity of dermal fibroblasts by subsetting this cell type from the complete dataset. Then, the `FindNeighbors()` and `FindClusters()` functions were applied again using 20 PCA dimensions and a resolution of 0.5. Likewise, the UMAP plots were re-calculated with default parameters and using 20 PCA dimensions.

To assess age-related differences in cellular identity, the `FindAllMarkers()` function was applied to the young and old skin datasets separately. This allowed the identification of representative genes in each cell cluster by age group. Also, to determine the genes differentially expressed by each fibroblast subpopulation upon aging, the `FindMarkers()` function was used.

GO analyses were performed using the DAVID Bioinformatics Database (version 6.8). Gene lists resulting from the different comparisons described in this study were queried into DAVID's Gene Functional Annotation Tool, and the `GOTERM_BP_DIRECT` option was selected as output. The first eight GO terms with a p-value <0.05 were displayed as the most significant categories.

To assess fibroblast expression of all collagen genes or described sets of genes defining cell or spatial identities, the average gene set expression per cell was calculated in young and old

fibroblast subpopulations. Data were projected onto UMAP or violin plots for visualization. To test for significant differences in expression, a Wilcoxon Rank Sum test was used.

Potential cellular interactions between the identified cell populations in the human skin were analyzed using the publicly available repository CellPhoneDB (v2.0.0). Using their web interface, pairwise comparisons between all cell populations were performed in both young and old skin datasets, as well as in each sample individually. Default parameters were used, thus including only those receptors or ligands expressed by at least 10% of the cells in a given cell cluster. For calculations, 1,000 iterations were used for the young and old datasets, while 100 iterations were used for the individual analyses. Interactions with a p-value < 0.05 were selected as significant.

### **6.2.10. mRNA fluorescence *in situ* hybridization (mRNA-FISH)**

The expression of selected target genes was assessed by mRNA-FISH using the RNAScope® Multiplex Fluorescent Detection Kit v2 (ACDBio) on 4 µm sections of FFPE healthy skin samples from young (28-37 y/o) and old (54-89 y/o) donors. The manufacturer's instructions were followed with slight modifications in sample pretreatment. Thus, such pretreatment consisted of a 15 minutes incubation with hydrogen peroxide, followed by a mild boil (98-102°C) with target retrieval reagents for 30 minutes in a water bath, and a final incubation of 30 minutes at 40°C in the HybEZ™ oven (ACDBio) with Protease Plus. All reagents and solutions used in the pretreatment steps were included in the kit. After sample pretreatment, probes against human *CTHRC1*, *APCDD1*, *CCL19*, *APOE*, *ASPN*, *CD248*, and *PDGFRA* mRNA were hybridized for 2 hours at 40°C in the HybEZ™ oven (ACDBio). Signal amplification, fluorophore addition, and HRP development steps were performed following the manufacturer's instructions. As fluorophores, Opal™ 520, Opal™ 570, and Opal™ 690 dyes (Akoya Biosciences) were used at a 1:1000 dilution in TSA buffer. Nuclei were counterstained with DAPI, and sections were mounted using ProLong® Gold Antifade Mountant (Thermo Fisher Scientific). For each selected target gene, the experiment was carried out in three biological replicates per age group, always including *PDGFRA* as a pan-fibroblast marker. All images were taken with a TCS SP5 confocal microscope (Leica Microsystems) using a 40X oil immersion lens and were further processed using Fiji software.

The expression of *APCDD1* and *CTHRC1* in different dermal regions (papillary, reticular, and deep reticular dermis) was quantified as follows. *APCDD1*- and *CTHRC1*-positive cells were counted in two images per dermal region for each independent skin sample in each age group. Statistical analysis was carried out with a two-way ANOVA test (Dunnett's correction) to compare the percentage of positive cells in the reticular and deep reticular dermal regions to the percentage calculated in the papillary dermis for both genes independently.

The expression of *CD248* in the deep reticular dermis in young and old skin was quantified as follows. *CD248*-positive cells were counted in two deep reticular images for each independent skin sample in both age groups. Statistical significance was tested using an unpaired two-sided t-test, comparing the percentage of positive cells between both age groups.

#### **6.2.11. Immunofluorescence assays (IF)**

Protein expression of selected target genes was assessed by IF assays on young and old human skin. Surgical leftover healthy whole-skin tissue from different donors was fixed overnight in 4% formalin in PBS, paraffin-embedded, and cut into 4  $\mu$ m sections. On the day of the experiment, sections were deparaffinized in xylene and subsequently rehydrated in a decreasing gradient of ethanol. Sections were rinsed with abundant distilled water before heat-induced antigen retrieval by 30 minutes of incubation at 95°C in a water bath in 10 mM citrate buffer (pH 6.0) supplemented with 0.05% Tween-20.

After antigen retrieval, permeabilization of skin sections used for Tetraspanin 8 and Periostin staining was performed by incubation with 0.4% Triton X-100 in 1% Normal Goat Serum (nGS; Dako) for 10 minutes at room temperature, twice. Prior to antibody labeling, non-specific antibody binding was blocked by incubation with 10% nGS for 1h. Sections were incubated at 4 °C overnight with primary antibodies properly diluted in 10% nGS. The primary antibodies and the dilutions used are summarized in Table M6. Primary antibodies against Vimentin, a pan-fibroblast marker, were also included in all assays. Either rabbit anti-Vimentin or chicken anti-Vimentin were used, depending on the combination with other primary antibodies used. The next day, samples were rinsed in 1X PBS containing 0.1% Tween-20 and then incubated for 10 minutes in 10% nGS for a second blocking step before secondary antibody incubation.

In contrast, unspecific antibody binding in skin sections used for Collagen alpha-1 (XVIII) chain staining was performed by incubation with 1% BSA, 22.52 mg/ml glycine in PBS with 0.1% Tween-20 for 1h at room temperature. Then, sections were incubated at 4°C overnight with primary antibodies in blocking solution. Dr. Ritva Heljasvaara from the University of Oulu (Finland) kindly provided the mouse anti-Collagen alpha-1(XVIII) chain antibody (DB144-N2, 1:150). Rabbit anti-Vimentin (Cell Signaling, D21H3, 1:100) was also added as a pan-fibroblast marker. The next day, sections were rinsed with 1X PBS containing 0.1% Tween-20 followed by secondary antibody incubation. All sections used in IF assays were incubated during 2 hours at room temperature with corresponding Alexa Fluor-conjugated secondary antibodies (Life Technologies). Nuclei were then counterstained with DAPI (1:10,000), and sections were mounted using Vectashield® Antifade Mounting Medium (Vector Laboratories).

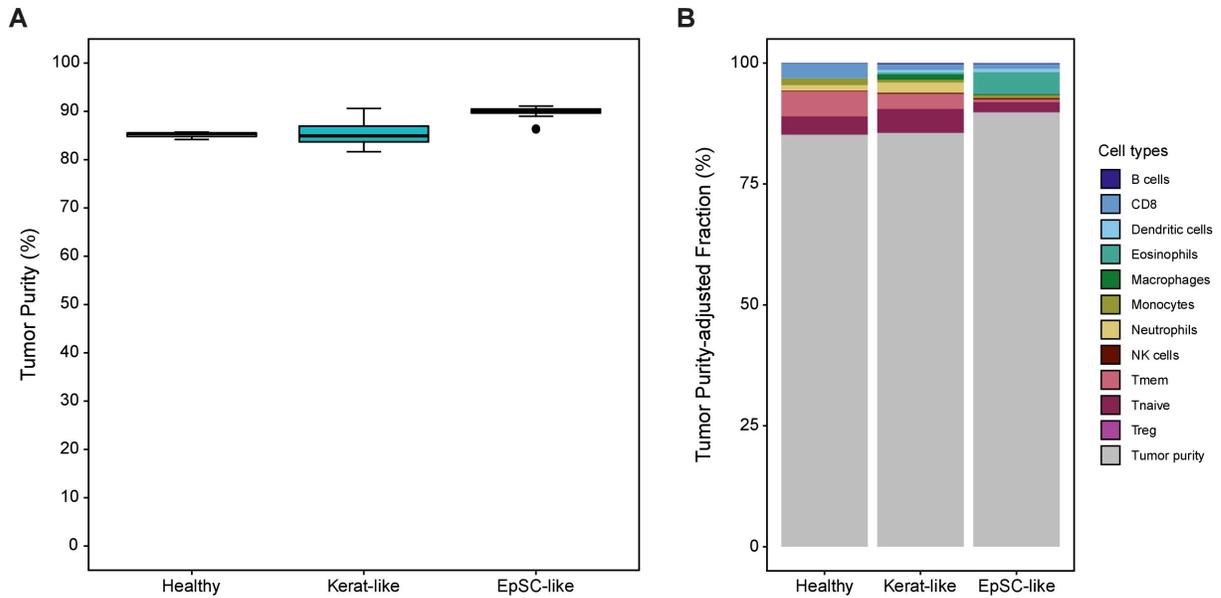
These assays were performed in six biological replicates of donors aged 37-79 y/o for Tetraspanin 8, in three skin samples of donors aged 51-66 y/o for Periostin, and in one sample of a 37 y/o donor for Collagen alpha-1(XVIII) chain. All IF assays were imaged using a TCS SP5 confocal microscope (Leica Microsystems) using a 40X oil immersion lens and were further processed using Fiji software.

### **6.2.12. Statistical analyses**

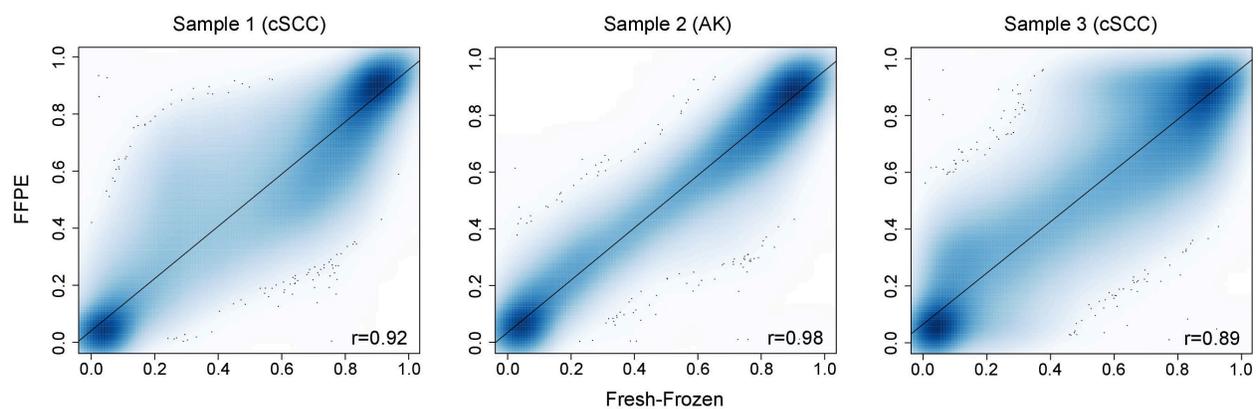
Statistical analyses including Fisher's Exact Test, Wilcoxon Rank Sum tests, Two-sided two-proportion z-test, two-way ANOVA (Dunnett's correction), and unpaired two-sided t-tests were performed in R (The R Foundation, version 3.3.1 and 4.0.2). The specific test used in particular experiments is properly described in the Results and Methods sections or in corresponding figure legends. P-values < 0.05 with a 95% confidence interval were considered as statistically significant.

## 7. Appendix

### 7.1. Supplementary Figures

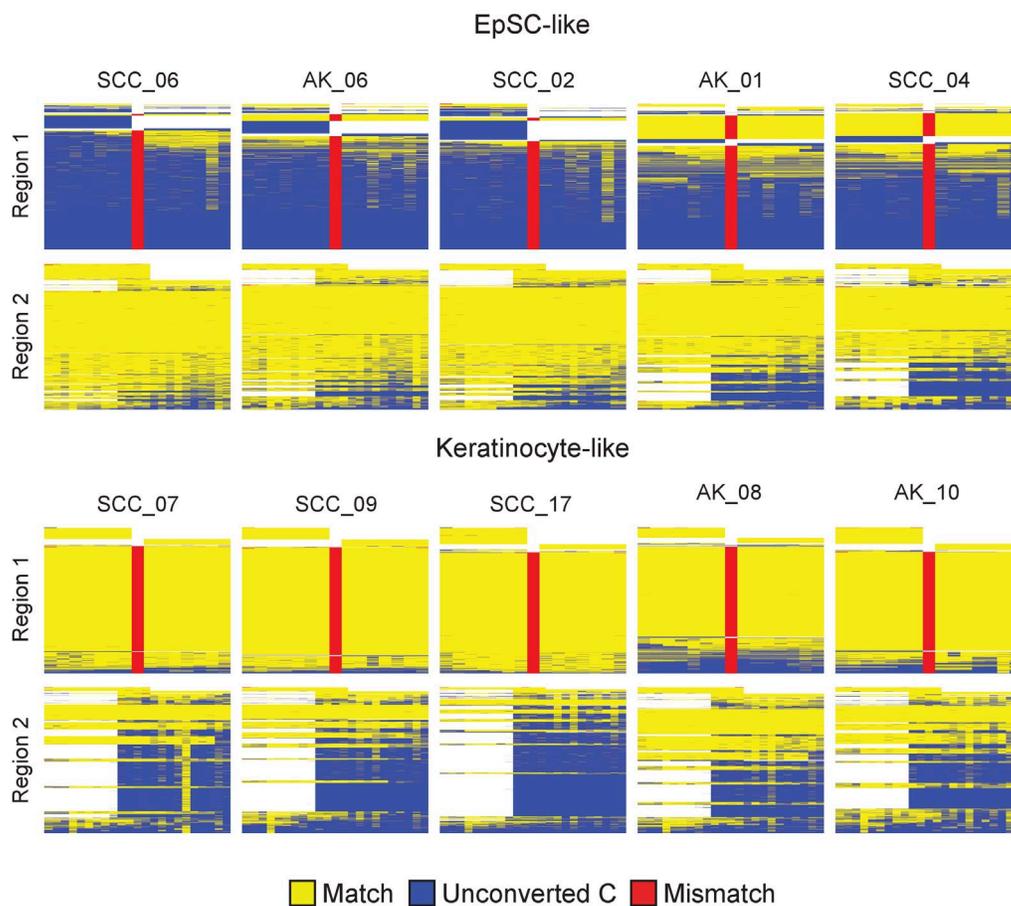


**Figure S1. Epidermal DNA methylation datasets display high tumor purity. (A)** Boxplot showing the tumor purity of all healthy and epidermal tumors as estimated by MethylResolver. Tumor samples were separated by cell-of-origin subclass. **(B)** Average tumor purity-adjusted fraction of leukocyte populations in healthy and epidermal tumor cell-of-origin subclasses as estimated by MethylResolver (Arneson, Yang and Wang, 2020).

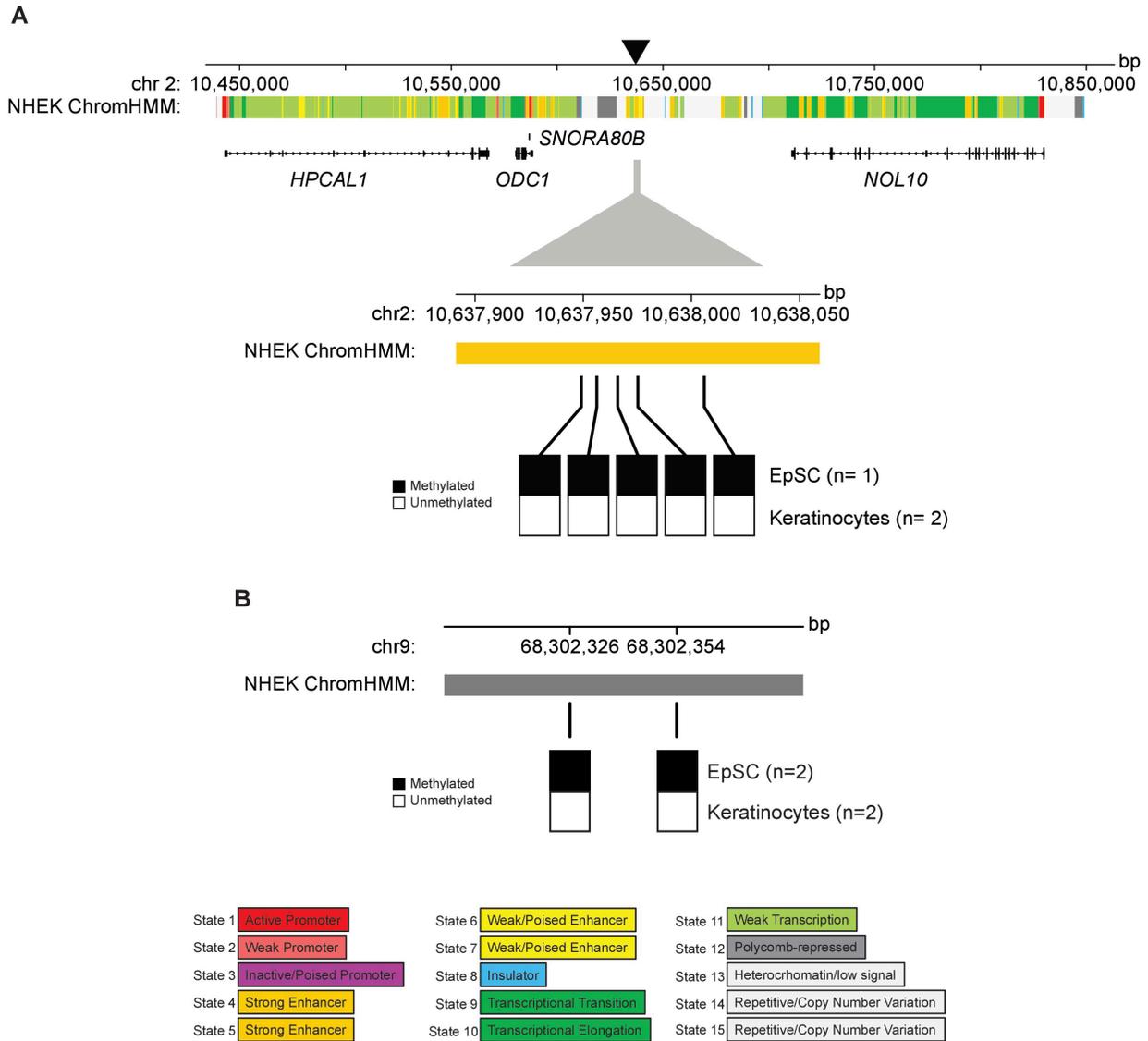


**Figure S2. Comparison of EPIC DNA methylation profiles in FF-FFPE matched samples.** Density plots displaying the comparison between methylomes obtained with Infinium MethylationEPIC BeadArrays in three matched FF-FFPE epidermal tumor samples. The correlation coefficient ( $r$ ) is shown for each comparison.

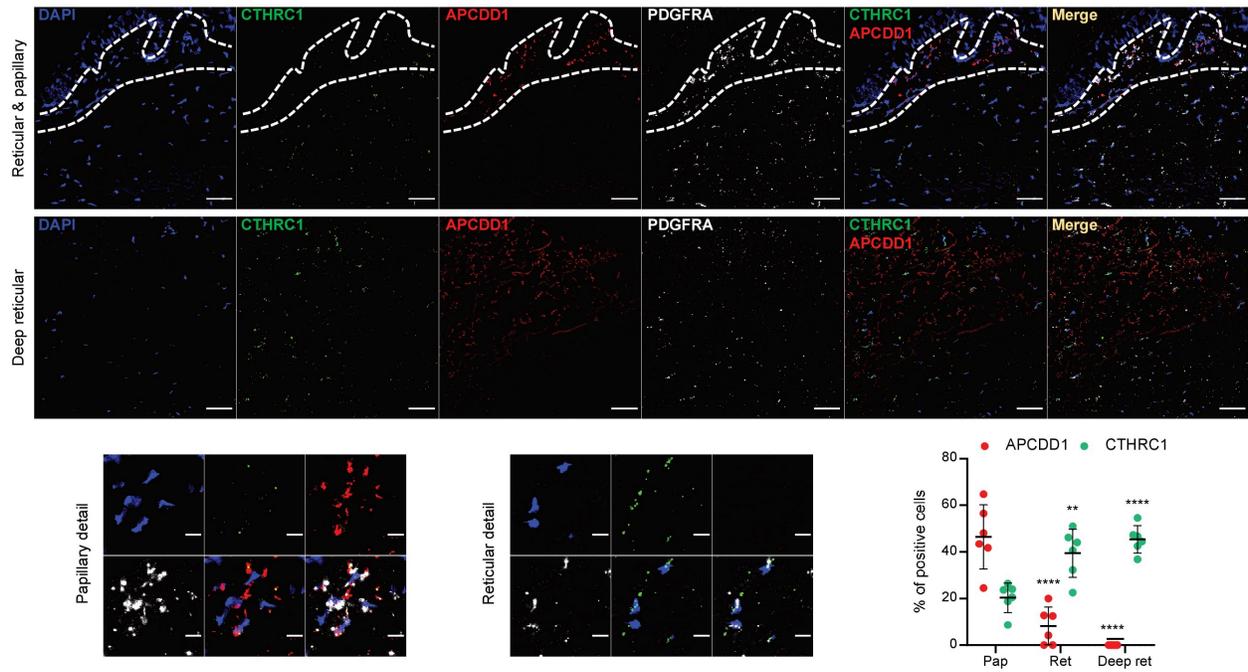
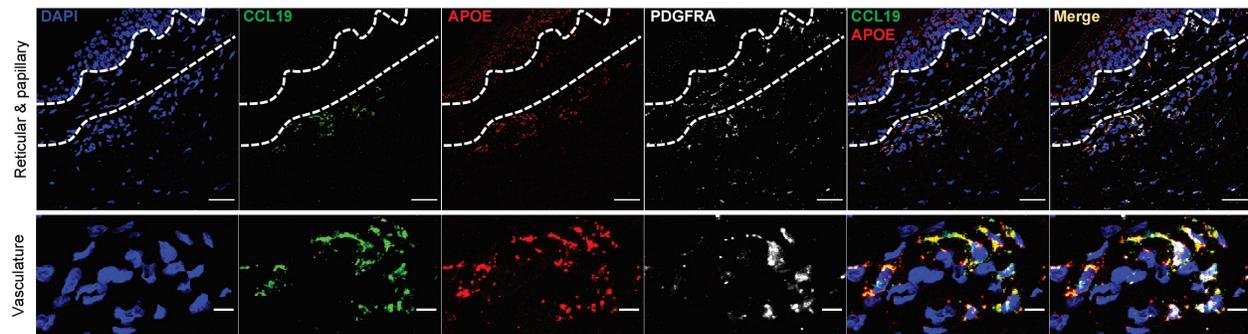
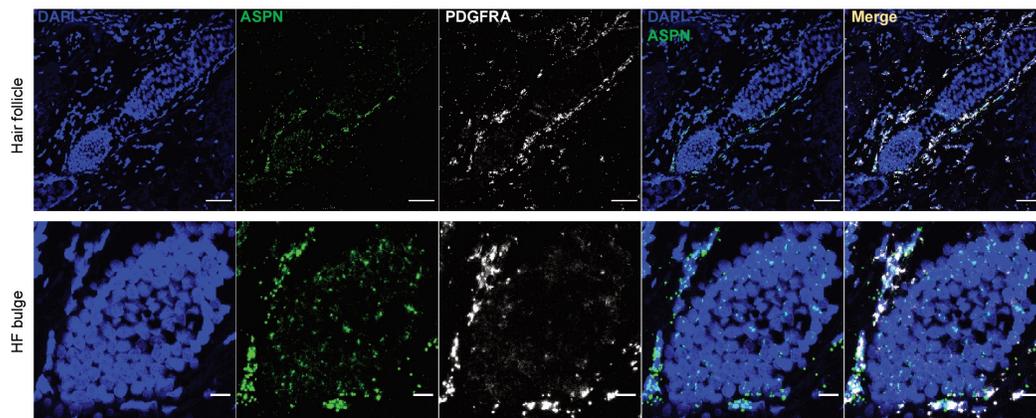
---



**Figure S3. Amplicon-based targeted sequencing strategy for KC stratification.** Plots displaying the methylation values per CpG (column) and read (row) in both regions for the 10 FF samples used for testing the approach. Unmethylated cytosines are shown in yellow while methylated ones are shown in blue. Red indicates mapping mismatches. Plots generated using BisAMP (Bormann *et al.*, 2019).

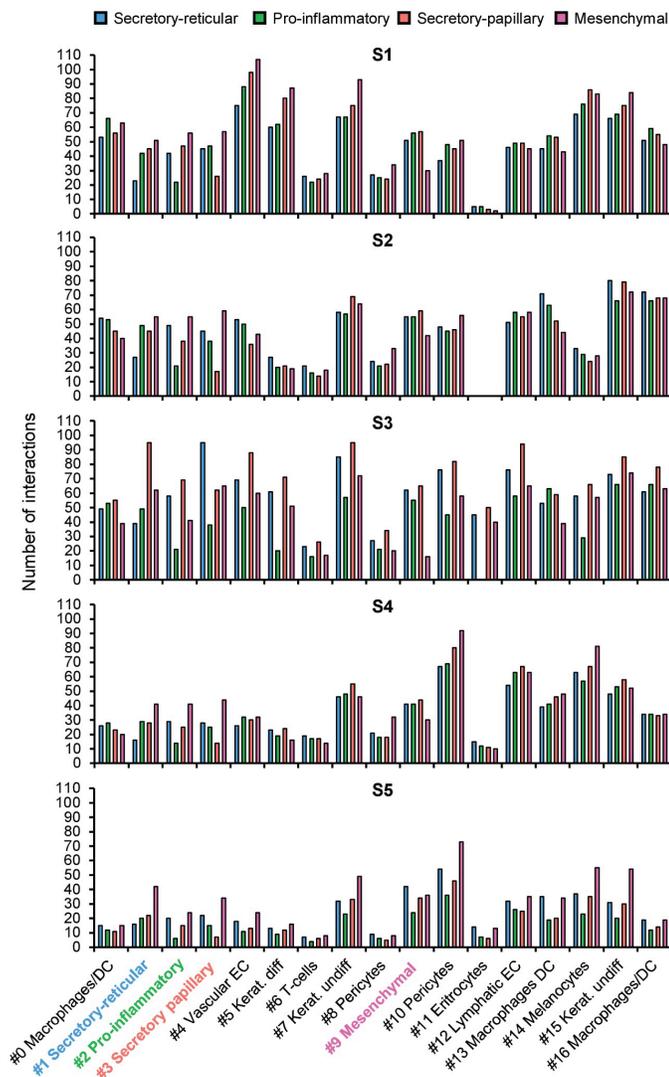


**Figure S4. Representative differentially methylated regions between EpSC and differentiated keratinocytes in the sci-MET dataset. (A)** Differentially methylated region located in a strong enhancer (as designated in NHEKs), in close proximity to *HPCAL1* and *ODC1*, two genes with known differential expression upon keratinocyte differentiation. **(B)** Differentially methylated region located in a polycomb-repressed region (as designated in NHEKs).

**A** Secretory-reticular and secretory-papillary**B** Pro-inflammatory**C** Mesenchymal

**Figure S5. Fibroblast subpopulations validation in old skin using mRNA-FISH. (A)** Representative confocal images for mRNA detection of *CTHRC1* (green) and *APCDD1* (red), markers for secretory-reticular and secretory-papillary subpopulations, respectively. Details from the papillary and reticular dermis are shown in the lower panel (left and center). Quantification of positive cells for each gene and dermal region is shown in the lower right panel. **(B)** Representative confocal images for mRNA detection of *CCL19* (green) and *APOE* (red), markers for pro-inflammatory fibroblasts. A detail of a vessel is shown in the lower panel. **(C)** Representative confocal images for mRNA detection of *ASPN* (green), a marker gene for mesenchymal fibroblasts. A detail of the hair follicle bulb is shown in the lower panel. In **(A)** and **(B)** dashed line denotes the papillary dermis while in **(C)**, it denotes the dermal papilla. Nuclei were counterstained with DAPI and fibroblast identity was assessed by mRNA detection of *PDGFRA* (gray) in all assays. Each experiment was performed in three old FFPE skin sections (54-86 y/o). Images are shown at 40x original magnification. Scale bar, 50  $\mu\text{m}$  (main images), and 10  $\mu\text{m}$  (details). Statistical analyses in **(A)** were performed using a two-way ANOVA test (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .); error bars represent the standard deviation. Pap: papillary dermis, Ret: reticular dermis, Deep ret: deep reticular dermis, HF: hair follicle, DP: dermal papilla.

---



**Figure S6. Fibroblast interactome analyses in individual skin samples.** Bar plots depicting the total number of predicted interactions between each fibroblast subpopulation and the rest of the cell types present in human skin, based on their transcriptomic profiles, in each individual sample. Coloring and numbering are according to the original unsupervised clustering performed by Seurat. The results indicate an intermediate phenotype of the 53 y/o sample (S3).



## 7.2. Supplementary Tables

	Age	Gender	Location	Cells	Avg. CpGs/cell	Max CpGs/cell	Sample type
Donor 1	62	Male	Ilioinguinal	186	0.79%	1.24%	Fresh
Donor 2	72	Male	Face	671	0.054%	0.48%	FF
Donor 3	61	Male	Head/neck	644	0.043%	0.18%	FF
Donor 4	70	Male	Face	547	0.033%	0.14%	FF

**Table S1. Overview of the human epidermal analyzed with sci-MET.** The table depicts biological features and sequencing parameters for the fresh sample used in the experiment described in the Results, section 3.2.3, as well as for the three FF epidermal samples used in the experiment described in the Additional results, section 7.3. The data shown for Donor 2-4 refer to the sequencing details of the library containing approximately 2000 cells.

Sample ID	Gender	Age	Skin type	Reads per sample	N° of cells	Reads per cell	Genes per cell
Donor 1	Male	25	Fair	322,091,192	2784	102,904	1343
Donor 2	Male	27	Fair	338,738,780	2670	119,737	1111
Donor 3	Male	53	Fair	359,776,321	3324	107,976	1718
Donor 4	Male	70	Fair	378,219,220	2144	170,215	1388
Donor 5	Male	69	Fair	370,342,531	4535	81,411	872

**Table S2. Overview of the sun-protected healthy skin samples used in the scRNA-seq analysis.** The table depicts biological features and sequencing statistics for the five skin samples used in the scRNA-seq experiment described in the Results, section 3.3.

## 7.3. Additional results

### 7.3.1. sci-MET protocol optimization for human epidermis isolated from frozen material.

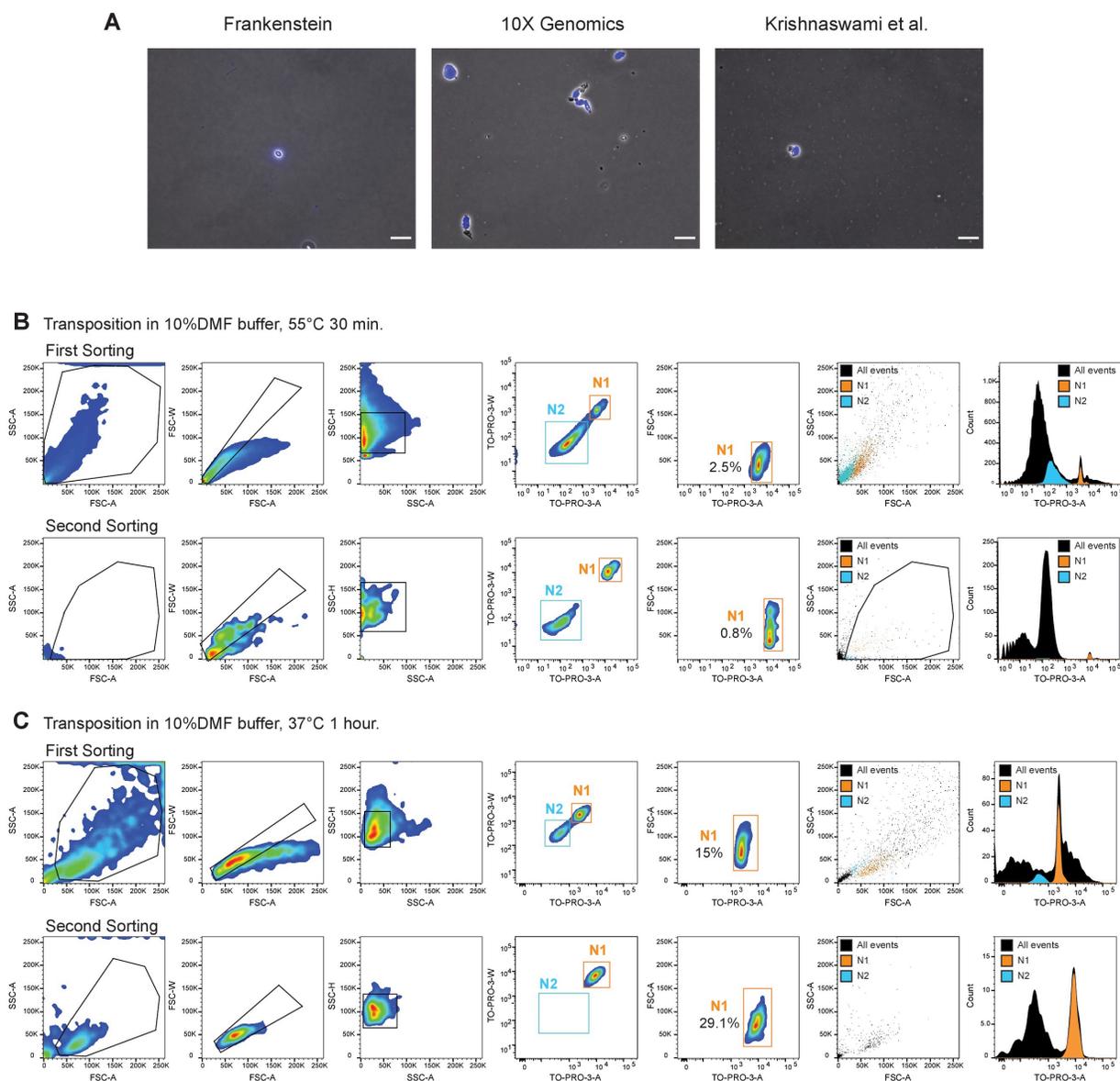
Fresh primary human samples are not always available. Also, obtaining sufficient available nuclei to sort a full 96-well plate in the first sorting step (necessary condition to maintain barcode probabilities and ensure proper single-cell resolution) from a unique sample might not always be possible. Thus, collecting FF samples and combining them into one plate might be a more convenient option to achieve successful sci-MET experiments. The protocol described for fresh epidermal samples in the Results section 3.2.2 could not be applied to FF human epidermis samples and thus the sample preparation protocol had to be optimized for this type of samples.

Healthy whole human skin samples were immersed in liquid nitrogen directly after resection, therefore containing epidermis and dermis, and were stored at -80 until further processing. For all the following experiments, the epidermis was separated from the dermis using the heat-split protocol (see Methods, section 6.2.1.1). First, three methods for nuclear isolation from FF samples were tested. These included the Frankenstein protocol (Martelotto, 2020), a 10X Genomics demonstrated protocol used for single-cell ATAC sequencing (10X Genomics), and the protocol proposed by Krishnaswami et. al. (Krishnaswami *et al.*, 2016). Although the Frankenstein protocol resulted in the best nuclear separation from cellular debris, the 10X Genomics protocol yielded a higher nuclei recovery (Figure S8A). Thus, the 10X Genomics protocol was used for the subsequent experiments. These three nuclear isolation protocols are described in detail in the Supplementary Methods, section 7.4.2.

The concentration of DMF in the tagmentation buffer was again tested for FF samples with 10% DMF yielding the best recovery. Nonetheless, the number of nuclei present in the second sorting step was significantly reduced compared to fresh samples, which prevented the sorting of full plates (Figure S8B). To test whether lowering the transposition temperature could reduce nuclear loss, a sci-MET experiment was carried out, performing the tagmentation at 37°C for 1 hour and with the 10% DMF buffer. These conditions yielded a good amount of nuclei (Figure S8C), and were subsequently used for further experiments using FF skin samples.

Three human epidermis samples from sun-exposed head and neck regions were used for sorting a full 96-well plate (32 wells/sample). Nuclei were then tagmented with the above-mentioned conditions and sci-MET library preparation was performed. A full plate library was successfully sequenced with a mapping rate of 90% to the human genome. Of the 2,112 possible single-cells present in the library (96well x 22 nuclei), a total of 1,633 were detected with at least

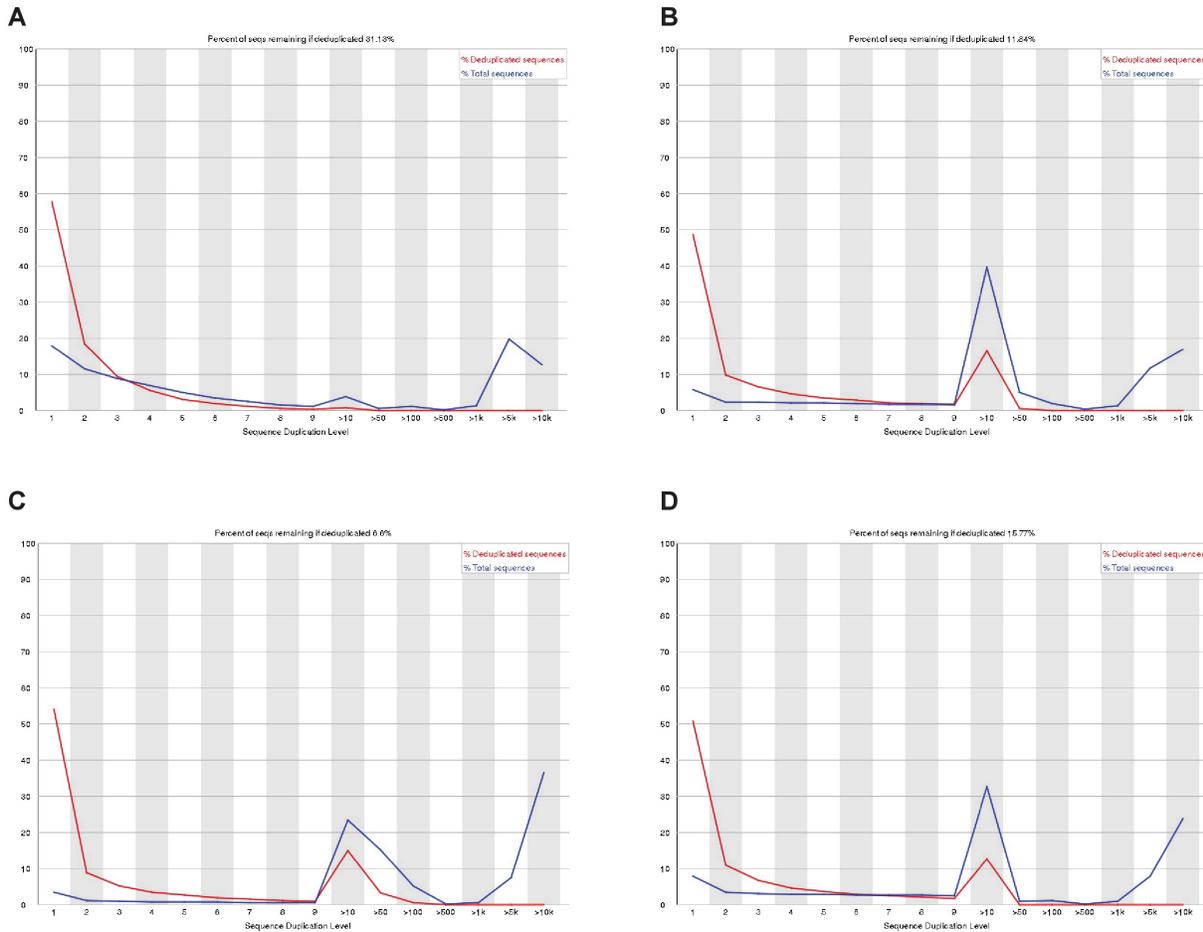
30,000 reads. All three donors were represented similarly in the dataset and displayed comparable average CpG coverage per cell and average methylation ratios (Table S1).



**Figure S8. sci-MET sample preparation protocol optimization for FF human epidermis. (A)** Representative images of nuclei after nucleosome depletion, isolated from FF healthy human epidermis using the Frankenstein (Martelotto, 2020), 10X Genomics, or Kishnaswami *et al.*, 2016 protocol. Images are shown at 10x original magnification. Scale bar, 100  $\mu$ m. **(B-C)** Representative FACS plots displaying the gating strategy followed for sorting single nuclei from FF healthy human epidermis. Upper plots display the first sorting step while lower plots display the second sorting step, after nuclear tagmentation performed at 55°C for 30 minutes **(B)** or 37°C for 1 hour **(C)**.

To increase CpG coverage, libraries containing only 200 cells were also generated from a second plate obtained from the same patients. Nonetheless, sequencing of this library revealed a high percentage of PCR duplicates that was masked by the high number of cells included in the first library, and therefore impairing downstream analyses (Figure S9A and S9B). This duplication rate was not significantly reduced by generating a library containing 200 cells obtained from the same patients and using 18 cycles in the index PCR step instead of 21 (Figure S9C).

Sequencing of a library with three new samples and following the same strategy but with freshly prepared transposomes yielded the same results, thus suggesting that the PCR duplication was not a consequence of using transposomes stored over a month (Figure S9D). Together, these findings suggest that lowering the temperature of transposition improves nuclear integrity but lowers transposition efficiency, and thus further work needs to be done to optimize the sci-MET protocol for FF human epidermal samples.



**Figure S9. sci-MET libraries from FF epidermis samples display high rates of PCR duplication.** **(A)** Plot displaying the sequence duplication levels in sci-MET libraries from the epidermis of three donors and containing 2,000 cells. **(B)** Plot displaying the sequence duplication levels in sci-MET libraries from the epidermis of the same three donors as in (A) but containing 200 cells. **(C)** Plot displaying the sequence duplication levels in sci-MET libraries from the epidermis of the same three donors as in (A) but containing 200 cells and generated with 18 PCR cycles. **(D)** Plot displaying the sequence duplication levels in sci-MET libraries from the epidermis of three different donors and containing 200 cells. The blue line depicts the distribution of duplication levels of all sequences. The red line depicts the de-duplicated sequences and the proportions refer to the amount of de-duplicated reads that belong to a certain duplication level in the original data.

---

## 7.4. Supplementary Methods

### 7.4.1. Nuclear isolation protocols for sci-MET (fresh human epidermis)

Apart from the nuclear isolation protocol used for final experiments (Wysocka, Reilly and Herr, 2001), three additional nuclear isolation protocols were tested to optimize the sci-MET protocol for fresh human epidermis samples. In all these experiments, the epidermis was isolated from fresh whole skin samples, and single-cell suspension was obtained with the Epidermis Dissociation kit for human material (Miltenyi Biotec). Isolated nuclei were quantified and observed under a microscope to assess the best method. The protocol described in Wysocka, Reilly and Herr, 2001 displayed a better yield than the three protocols described in this section and, thus, was used in further experiments.

#### 7.4.1.1. Buro, Shah and Henriksen, 2010

After cells were recovered, pellets were resuspended in 4 ml pre-chilled Swelling Buffer (25 mM HEPES pH 7.8, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 0.1% IGEPAL® CA-630, 1 mM DTT and 1X protease inhibitors cocktail). Cells were incubated on ice for 10 minutes, followed by homogenization with a Douncer (20 strokes with tight pestle). Nuclei were then recovered by centrifugation at 2,500 rpm for 5 minutes at 4°C.

#### 7.4.1.2. Neely and Bao, 2019

After cells were recovered, pellets were resuspended in 200 µl Nuclei extraction Buffer A per 1 million cells (NEBA; 10 mM HEPES pH 7.4, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl and 1X protease inhibitors cocktail). Then, an equal volume of Nuclei Extraction Buffer B (NEBB; 10 mM HEPES pH 7.4, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 0.4% IGEPAL CA-630 and 1X protease inhibitors cocktail) was added and the sample was mixed by gentle pipetting. Cells were incubated on ice for 2 minutes, and nuclei were subsequently recovered by centrifugation at 500 g for 5 minutes at 4°C.

#### 7.4.1.3. Corces *et al.* 2017

After cells were recovered, pellets were resuspended in ATAC-RSB buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>) supplemented with 0.1% IGEPAL® CA-630, 0.1% Tween-20 and 0.01% Digitonin in a 50µl/50.000 cells ratio. Cells were incubated on ice for 3 minutes, and 1 ml of ATAC-RSB buffer containing 0.1% Tween-20 was then added. Nuclei were recovered by centrifugation at 500 g for 10 minutes at 4°C.

### 7.4.2. Nuclear isolation protocols for sci-MET (FF human epidermis)

Fresh-frozen whole human skin samples were collected and kept at -80°C until further processing. The epidermis was isolated using heat split as described in the Methods section 6.2.1.1. Three nuclear isolation protocols were tested for this type of samples and are described hereunder. Isolated nuclei were quantified and observed under a microscope to assess the best method. Even though the Frankenstein protocol provided a cleaner nuclei isolation, the 10X Genomics protocol displayed a better yield, and therefore was used in the experiments described in the Additional results section 7.3.1.

#### 7.4.2.1. 10X Genomics applied protocol

The isolated epidermis was directly immersed in 500 µl of 0.1X Lysis Buffer (Table S3), and homogenized using a pellet pestle (15-20 strokes). Samples were incubated on ice for 15 minutes with mixing with a wide-bore pipette tip 10 times after 5 minutes of incubation. Then, 500 µl of cold Wash Buffer (Table S3) were added to the samples and mixed by pipetting. Lysates were passed through a 70 µm cell strainer and subsequently through a 40 µm to remove cellular debris. Nuclei were then recovered by centrifugation at 500 g for 5 minutes at 4°C.

	<b>Recipe</b>
Wash Buffer	10 mM Tris-HCl pH 7.4 10 mM NaCl 3 mM MgCl <sub>2</sub> 1% BSA 0.1% Tween-20
Lysis Dilution Buffer	10 mM Tris-HCl pH 7.4 10 mM NaCl 3mM MgCl <sub>2</sub> 1% BSA
1X Lysis Buffer	10 mM Tris-HCl pH 7.4 10 mM NaCl 3 mM MgCl <sub>2</sub> 0.1% Tween-20 0.1% IGEPAL® CA-630 0.01% Digitonin 1% BSA
0.1X Lysis Buffer	<i>For 2 ml:</i> 200 ml 1X Lysis Buffer 1.8 ml Lysis Dilution Buffer

**Table S3. Nuclear isolation buffers from 10X Genomics demonstrated protocol.** The table depicts the recipes for the buffers described in the 10X Genomics demonstrated protocol “Nuclei Isolation from Mouse Brain Tissue for Single Cell ATAC Sequencing” and used for nuclear isolation of FF human epidermis in the sci-MET experiments.

### 7.4.2.2. Frankenstein protocol (Martelotto, 2020)

The isolated epidermis was transferred to a tube containing 500  $\mu$ l of pre-chilled Nuclei EZ Lysis Buffer (Sigma-Aldrich). Tissue was homogenized with 20 strokes using a pellet pestle followed by the addition of 1 ml of Nuclei EZ lysis Buffer (Sigma-Aldrich). Samples were then incubated for 5 minutes on ice before being passed through a 70  $\mu$ m cell strainer. Cells were centrifuged at 500 g for 5 minutes at 4°C, and the pellet was resuspended in 1.5 ml Nuclei EZ Lysis Buffer (Sigma-Aldrich) and incubated for 5 minutes on ice. Samples were centrifuged at 500 g for 5 minutes at 4°C, and 500  $\mu$ l of Nuclei Wash and Resuspension Buffer (1% BSA, 0.2U/ $\mu$ l RNaseA Inhibitor in 1X PBS) were subsequently added without resuspending the pellets, which were incubated 5 minutes on ice. After incubation, 1 ml of Nuclei Wash and Resuspension Buffer was added, and nuclei were recovered by centrifugation at 500 g for 5 minutes at 4°C.

### 7.4.2.3. Krishnaswami *et al.*, 2016

A glass Dounce homogenizer and two pestles (tight and loose) were pre-chilled on ice. The isolated epidermis sample was transferred to the Douncer, containing 1 ml of cold Homogenization Buffer (Table S4). Tissue was homogenized on ice by performing five strokes with the loose pestle followed by 15 strokes with the tight pestle. The sample was then passed through a 40  $\mu$ m cell strainer, and nuclei were recovered by centrifugation at 500 g for 5 minutes at 4°C.

Recipe	
Nuclei isolation medium #1 (NIM1)	1.5 M sucrose 1M KCl 1M MgCl <sub>2</sub> 1M Tris buffer, pH 8.0
Nuclei isolation medium #2 (NIM2)	<i>For 5 ml:</i> 4,895 $\mu$ l NIM1 5 $\mu$ l 1 mM DTT 100 $\mu$ l 50X protease inhibitor
Homogenization buffer	<i>For 1.5 ml:</i> 1,452 $\mu$ l NIM2 15 $\mu$ l RNase inhibitor (40U/ $\mu$ l) 15 $\mu$ l Superasin (20U/ $\mu$ l)* 15 $\mu$ l Triton X-100 (10% v/v)

**Table S4. Nuclear isolation buffers from Krishnaswami *et al.*, 2016 protocol.** The table depicts the recipes for the buffers described in the nuclei isolation protocol published in Krishnaswami *et al.*, 2016 and used for nuclear isolation of FF human epidermis in the sci-MET experiments. \* *Not added in the test experiment.*

## 7.5. List of publications

**Solé-Boldo, L.**, Raddatz, G., Schütz, S., Mallm J-P., Rippe, K., Lonsdorf, A.S., Rodríguez-Paredes, M., Lyko, F. Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming. *Commun Biol* **3**, 188 (2020). <https://doi.org/10.1038/s42003-020-0922-4>

## 8. References

- Adey, A. *et al.* (2014) “In vitro, long-range sequence information for de novo genome assembly via transposase contiguity,” *Genome Research*, 24(12), pp. 2041–2049. doi: 10.1101/gr.178319.114.
- Albibas, A. A. *et al.* (2018) “Subclonal Evolution of Cancer-Related Gene Mutations in p53 Immunopositive Patches in Human Skin,” *Journal of Investigative Dermatology*, 138(1), pp. 189–198. doi: 10.1016/j.jid.2017.07.844.
- Al-Eitan, L. N. *et al.* (2020) “Genome-wide identification of methylated CpG sites in nongenital cutaneous warts,” *BMC Medical Genomics*, 13(1), p. 100. doi: 10.1186/s12920-020-00745-6.
- Amini, S. *et al.* (2014) “Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing,” *Nature Genetics*, 46(12), pp. 1343–1349. doi: 10.1038/ng.3119.
- Amôr, N. G., Santos, P. S. da S. and Campanelli, A. P. (2021) “The Tumor Microenvironment in SCC: Mechanisms and Therapeutic Opportunities,” *Frontiers in Cell and Developmental Biology*, 9(February), pp.1-12. doi: 10.3389/fcell.2021.636544.
- An, J., Rao, A. and Ko, M. (2017) “TET family dioxygenases and DNA demethylation in stem cells and cancers,” *Experimental and Molecular Medicine*, 49(e323). doi: 10.1038/emm.2017.5.
- Angermueller, C. *et al.* (2017) “DeepCpG: Accurate prediction of single-cell DNA methylation states using deep learning,” *Genome Biology*, 18(1), pp. 1–13. doi: 10.1186/s13059-017-1189-z.
- Apalla, Z., Lallas, A., *et al.* (2017) “Epidemiological trends in skin cancer.,” *Dermatology practical & conceptual*, 7(2), pp. 1–6. doi: 10.5826/dpc.0702a01.
- Apalla, Z., Nashan, D., *et al.* (2017) “Skin Cancer: Epidemiology, Disease Burden, Pathophysiology, Diagnosis, and Therapeutic Approaches,” *Dermatology and Therapy*, pp. 5–19. doi: 10.1007/s13555-016-0165-y.
- Arneson, D., Yang, X. and Wang, K. (2020) “MethylResolver—a method for deconvoluting bulk DNA methylation profiles into known and unknown cell contents,” *Communications Biology*, 3(1), pp. 1–13. doi: 10.1038/s42003-020-01146-2.
- Aryee, M. J. *et al.* (2014) “Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays,” *Bioinformatics*, 30(10), pp. 1363–1369. doi: 10.1093/bioinformatics/btu049.
- Ascensión, A. M. *et al.* (2021) “Human Dermal Fibroblast Subpopulations Are Conserved across Single-Cell RNA Sequencing Studies,” *Journal of Investigative Dermatology*, 141(7), pp. 1735-1744. doi: 10.1016/j.jid.2020.11.028.
- Aunan, J. R., Cho, W. C. and Søreide, K. (2017) “The biology of aging and cancer: A brief overview of shared and divergent molecular hallmarks,” *Aging and Disease*, 8(5), pp. 628–642. doi: 10.14336/AD.2017.0103.
- Avgustinova, A. and Benitah, S. A. (2016) “Epigenetic control of adult stem cell function,” *Nature Reviews Molecular Cell Biology*, 17(10), pp. 643–658. doi: 10.1038/nrm.2016.76.
- Azimi, A. *et al.* (2018) “Differential proteomic analysis of actinic keratosis , Bowen ’ s disease and cutaneous squamous cell carcinoma by label-free LC – MS / MS,” *Journal of Dermatological Science*, 91(1), pp. 69–78. doi: 10.1016/j.jdermsci.2018.04.006.

- Azimi, A. *et al.* (2020) “Data Independent Acquisition Proteomic Analysis Can Discriminate between Actinic Keratosis, Bowen’s Disease, and Cutaneous Squamous Cell Carcinoma,” *Journal of Investigative Dermatology*, 140(1), pp. 212–222.e11. doi: 10.1016/j.jid.2019.06.128.
- Bailleul, B. *et al.* (1990) “Skin hyperkeratosis and papilloma formation in transgenic mice expressing a ras oncogene from a suprabasal keratin promoter,” *Cell*, 62(4), pp. 697–708. doi: 10.1016/0092-8674(90)90115-U.
- Bartoš, V. and Kullová, M. (2018) “Non-melanoma Skin Cancer – a Clinicopathological Study of Patients with Basal Cell Carcinoma and Squamous Cell Carcinoma,” *Klinická Onkologie*, 31(1), pp. 40–45. doi: 10.14735/amko201840.
- Baylin, S. B. and Jones, P. A. (2016) “Epigenetic determinants of cancer,” *Cold Spring Harbor Perspectives in Biology*, 8(9). doi: 10.1101/cshperspect.a019505.
- Beerman, I. *et al.* (2013) “Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging,” *Cell Stem Cell*, 12(4), pp. 413–425. doi: 10.1016/j.stem.2013.01.017.
- Bell, C. G. *et al.* (2019) “DNA methylation aging clocks: Challenges and recommendations,” *Genome Biology*, 20(1), pp. 1–24. doi: 10.1186/s13059-019-1824-y.
- Bergen, V. *et al.* (2020) “Generalizing RNA velocity to transient cell states through dynamical modeling,” *Nature Biotechnology*, 38(12), pp. 1408–1414. doi: 10.1038/s41587-020-0591-3.
- Berman, B. P. *et al.* (2012) “Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamin-associated domains,” *Nature Genetics*, 44(1), pp. 40–46. doi: 10.1038/ng.969.
- Bert, S. A. *et al.* (2013) “Regional Activation of the Cancer Genome by Long-Range Epigenetic Remodeling,” *Cancer Cell*, 23(1), pp. 9–22. doi: 10.1016/j.ccr.2012.11.006.
- Bielajew, B. J., Hu, J. C. and Athanasiou, K. A. (2020) “Collagen: quantification, biomechanics and role of minor subtypes in cartilage,” *Nature Reviews Materials*, 5(10), pp. 730–747. doi: 10.1038/s41578-020-0213-1.
- Blanpain, C. (2013) “Tracing the cellular origin of cancer,” *Nature Cell Biology*, 15(2), pp. 126–134. doi: 10.1038/ncb2657.
- Blanpain, C. and Fuchs, E. (2009) “Epidermal homeostasis: a balancing act of stem cells in the skin,” *Nature reviews. Molecular cell biology*, 10(3), pp. 207–17. doi: 10.1038/nrm2636.
- Bock, C. *et al.* (2012) “DNA Methylation Dynamics during In Vivo Differentiation of Blood and Skin Stem Cells,” *Molecular Cell*, 47(4), pp. 633–647. doi: 10.1016/j.molcel.2012.06.019.
- Bond, D. R. *et al.* (2020) “Single-cell epigenomics in cancer: Charting a course to clinical impact,” *Epigenomics*, 12(13), pp. 1139–1151. doi: 10.2217/epi-2020-0046.
- Bonilla, X. *et al.* (2016) “Genomic analysis identifies new drivers and progression pathways in skin basal cell carcinoma,” *Nature Genetics*, 48(4), pp. 398–406. doi: 10.1038/ng.3525.
- Bormann, F. *et al.* (2018) “Cell-of-Origin DNA Methylation Signatures Are Maintained during Colorectal Carcinogenesis,” *Cell Reports*, 23(11), pp. 3407–3418. doi: 10.1016/j.celrep.2018.05.045.
- Bormann, F. *et al.* (2019) “BisAMP: A web-based pipeline for targeted RNA cytosine-5 methylation analysis,” *Methods*, 156, pp. 121–127. doi: 10.1016/j.ymeth.2018.10.013.
- Bostick, M. *et al.* (2007) “UHRF1 plays a role in maintaining DNA methylation in mammalian cells,” *Science*, 317(5845), pp. 1760–1764. doi: 10.1126/science.1147939.

## 8. References

---

- Botting, R. A. *et al.* (2017) "Phenotypic and functional consequences of different isolation protocols on skin mononuclear phagocytes," *Journal of Leukocyte Biology*, 101(6), pp. 1393–1403. doi: 10.1189/jlb.4a1116-496r.
- Bowden, G. T. (2004) "Prevention of non-melanoma skin cancer by targeting ultraviolet-B-light signalling," *Nature Reviews Cancer*, 4(1), pp. 23–35. doi: 10.1038/nrc1253.
- Braasch, D. A. and Corey, D. R. (2001) "Locked nucleic acid (LNA): Fine-tuning the recognition of DNA and RNA," *Chemistry and Biology*, 8(1), pp. 1–7. doi: 10.1016/S1074-5521(00)00058-2.
- Brinkhuizen, T. *et al.* (2012) "Epigenetic Changes in Basal Cell Carcinoma Affect SHH and WNT Signaling Components," *PLoS ONE*, 7(12), p. e51710. doi: 10.1371/journal.pone.0051710.
- Budden, T. *et al.* (2021) "Female immunity protects from cutaneous squamous cell carcinoma," *Clinical Cancer Research*, 27(11), pp. 3215–3223. doi: 10.1158/1078-0432.ccr-20-4261.
- Buro, L. J., Shah, S. and Henriksen, M. A. (2010) "Chromatin immunoprecipitation (ChIP) to assay dynamic histone modification in activated gene expression in human cells," *Journal of Visualized Experiments*, (41), p. e2053. doi: 10.3791/2053.
- Burton, K. A., Ashack, K. A. and Khachemoune, A. (2016) "Cutaneous Squamous Cell Carcinoma: A Review of High-Risk and Metastatic Disease," *American Journal of Clinical Dermatology*, 17(5), pp. 491–508. doi: 10.1007/s40257-016-0207-3.
- Cameron, M. C. *et al.* (2019) "Basal cell carcinoma: Epidemiology; pathophysiology; clinical and histological subtypes; and disease associations," *Journal of the American Academy of Dermatology*, 80(2), pp. 303–317. doi: 10.1016/j.jaad.2018.03.060.
- Campisi, J. (2005) "Senescent cells, tumor suppression, and organismal aging: Good citizens, bad neighbors," *Cell*, 120(4), pp. 513–522. doi: 10.1016/j.cell.2005.02.003.
- di Carlo, S. E. and Peduto, L. (2018) "The perivascular origin of pathological fibroblasts," *Journal of Clinical Investigation*, 128(1), pp. 54–63. doi: 10.1172/JCI93558.
- Carmona, F. J. *et al.* (2014) "A comprehensive DNA methylation profile of epithelial-to-mesenchymal transition," *Cancer Research*, 74(19), pp. 5608–5619. doi: 10.1158/0008-5472.CAN-13-3659.
- Celià-Terrassa, T. *et al.* (2012) "Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells," *Journal of Clinical Investigation*, 122(5), pp. 1849–1868. doi: 10.1172/JCI59218.
- Chahoud, J. *et al.* (2016) "Association between  $\beta$ -genus human papillomavirus and cutaneous squamous cell carcinoma in immunocompetent individuals-a meta-analysis," *JAMA Dermatology*, 152(12), pp. 1354–1364. doi: 10.1001/jamadermatol.2015.4530.
- Chang, H. Y. *et al.* (2002) "Diversity, topographic differentiation, and positional memory in human fibroblasts," *Proceedings of the National Academy of Sciences of the United States of America*, 99(20), pp. 12877–12882. doi: 10.1073/pnas.162488599.
- Chang, Y., Li, H. and Guo, Z. (2014) "Mesenchymal stem cell-like properties in fibroblasts," *Cellular Physiology and Biochemistry*, 34(3), pp. 703–714. doi: 10.1159/000363035.
- Chen, J. T., Kempton, S. J. and Rao, V. K. (2016) "The economics of skin cancer: An analysis of medicare payment data," *Plastic and Reconstructive Surgery - Global Open*, 4(9). doi: 10.1097/GOX.0000000000000826.

- Chen, S. X., Zhang, L. J. and Gallo, R. L. (2019) "Dermal White Adipose Tissue: A Newly Recognized Layer of Skin Innate Defense," *Journal of Investigative Dermatology*, 139(5), pp. 1002–1009. doi: 10.1016/J.JID.2018.12.031.
- Chen, X. *et al.* (2018) "A rapid and robust method for single cell chromatin accessibility profiling," *Nature Communications*, 9(1), pp. 1–9. doi: 10.1038/s41467-018-07771-0.
- Cheng, J. *et al.* (2015) "History of allergy and atopic dermatitis in relation to squamous cell and basal cell carcinoma of the skin," *Cancer Epidemiology Biomarkers and Prevention*, 24(4), pp. 749–754. doi: 10.1158/1055-9965.EPI-14-1243.
- Cheng, J. B. *et al.* (2018) "Transcriptional Programming of Normal and Inflamed Human Epidermis at Single-Cell Resolution," *CellReports*, 25(4), pp. 871–883. doi: 10.1016/j.celrep.2018.09.006.
- Chitsazzadeh, V. *et al.* (2016) "Cross-species identification of genomic drivers of squamous cell carcinoma development across preneoplastic intermediates," *Nature Communications*, 7(1), pp. 1–17. doi: 10.1038/ncomms12601.
- Christenson, L. J. *et al.* (2005) "Incidence of basal cell and squamous cell carcinomas in a population younger than 40 years," *Journal of the American Medical Association*, 294(6), pp. 681–690. doi: 10.1001/jama.294.6.681.
- Chtanova, T. *et al.* (2005) "Identification of T Cell-Restricted Genes, and Signatures for Different T Cell Responses, Using a Comprehensive Collection of Microarray Datasets," *The Journal of Immunology*, 175(12), pp. 7837–7847. doi: 10.4049/jimmunol.175.12.7837.
- Chung, M. *et al.* (2021) "DNA methylation ageing clocks and pancreatic cancer risk: pooled analysis of three prospective nested case-control studies," *Epigenetics*, pp. 1–11. doi: 10.1080/15592294.2020.1861401.
- Cichorek, M. *et al.* (2013) "Skin melanocytes: Biology and development," *Postepy Dermatologii i Alergologii*, 30(1), pp. 30–41. doi: 10.5114/pdia.2013.33376.
- Ciriello, G. and Magnani, L. (2021) "The many faces of cancer evolution," *iScience*, 24(5), p. 102403. doi: 10.1016/j.isci.2021.102403.
- Cives, M. *et al.* (2020) "Non-melanoma skin cancers: Biological and clinical features," *International Journal of Molecular Sciences*, 21(15), pp. 1–24. doi: 10.3390/ijms21155394.
- Clark, S. J. *et al.* (2018) "ScNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells," *Nature Communications*, 9(1), pp. 1–9. doi: 10.1038/s41467-018-03149-4.
- Clayton, K. *et al.* (2017) "Langerhans cells-programmed by the epidermis," *Frontiers in Immunology*, 8(NOV), p. 1. doi: 10.3389/fimmu.2017.01676.
- Cohen, R. M. *et al.* (2008) "Red cell life span heterogeneity in hematologically normal people is sufficient to alter HbA1c," *Blood*, 112(10), pp. 4284–4291. doi: 10.1182/blood-2008-04-154112.
- Conforti, C. *et al.* (2021) "Sclerodermiform basal cell carcinomas vs. other histotypes: analysis of specific demographic, clinical and dermatoscopic features," *Journal of the European Academy of Dermatology and Venereology*, 35(1), pp. 79–87. doi: 10.1111/jdv.16597.
- Corces, M. R. *et al.* (2016) "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution," *Nature Genetics*, 48(10), pp. 1193–1203. doi: 10.1038/ng.3646.
- Corces, M. R. *et al.* (2017) "An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues," *Nature Methods*, 14(10), pp. 959–962. doi: 10.1038/nmeth.4396.

- Das, R. N. and Yaniv, K. (2020) “Discovering new progenitor cell populations through lineage tracing and in vivo imaging,” *Cold Spring Harbor Perspectives in Biology*, 12(10), pp. 1–25. doi: 10.1101/cshperspect.a035618.
- Davalos, V. *et al.* (2012) “Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis,” *Oncogene*, 31(16), pp. 2062–2074. doi: 10.1038/onc.2011.383.
- Deaton, A. M. and Bird, A. (2011) “CpG islands and the regulation of transcription,” *Genes and Development*, 25(10), pp. 1010–1022. doi: 10.1101/gad.2037511.
- Deng, C.-C. *et al.* (2021) “Single-cell RNA-seq reveals fibroblast heterogeneity and increased mesenchymal fibroblasts in human fibrotic skin diseases,” *Nature Communications*, 12(1), p. 3709. doi: 10.1038/s41467-021-24110-y.
- Deplus, R. *et al.* (2014) “Regulation of DNA methylation patterns by CK2-mediated phosphorylation of Dnmt3a,” *Cell Reports*, 8(3), pp. 743–753. doi: 10.1016/j.celrep.2014.06.048.
- Dika, E. *et al.* (2020) “Basal cell carcinoma: A comprehensive review,” *International Journal of Molecular Sciences*, 21(15), pp. 1–11. doi: 10.3390/ijms21155572.
- Ding, J. *et al.* (2020) “Systematic comparison of single-cell and single-nucleus RNA-sequencing methods,” *Nature Biotechnology*, 38(6), pp. 737–746. doi: 10.1038/s41587-020-0465-8.
- Dotto, G. P. and Rustgi, A. K. (2016) “Squamous Cell Cancers: A Unified Perspective on Biology and Genetics,” *Cancer Cell*, 29(5), pp. 622–637. doi: 10.1016/j.ccell.2016.04.004.
- Driskell, R. R. *et al.* (2011) “Hair follicle dermal papilla cells at a glance,” *Journal of Cell Science*, 124(8), pp. 1179–1182. doi: 10.1242/jcs.082446.
- Driskell, R. R. and Watt, F. M. (2015) “Understanding fibroblast heterogeneity in the skin,” *Trends in Cell Biology*, 25(2), pp. 92–99. doi: 10.1016/j.tcb.2014.10.001.
- Dulauroy, S. *et al.* (2012) “Lineage tracing and genetic ablation of ADAM12 + perivascular cells identify a major source of profibrotic cells during acute tissue injury,” *Nature Medicine*, 18(8), pp. 1262–1270. doi: 10.1038/nm.2848.
- Duran-Ferrer, M. *et al.* (2020) “The proliferative history shapes the DNA methylome of B-cell tumors and predicts clinical outcome,” *Nature Cancer*, 1(11), pp. 1066–1081. doi: 10.1038/s43018-020-00131-2.
- Edqvist, P. H. D. *et al.* (2015) “Expression of Human Skin-Specific Genes Defined by Transcriptomics and Antibody-Based Profiling,” *Journal of Histochemistry and Cytochemistry*, 63(2), pp. 129–141. doi: 10.1369/0022155414562646.
- Ehrlich, M. and Lacey, M. (2013) “DNA hypomethylation and hemimethylation in cancer,” *Advances in Experimental Medicine and Biology*, 754, pp. 31–56. doi: 10.1007/978-1-4419-9967-2\_2.
- Elizondo, D. M. *et al.* (2017) “Inhibition of allograft inflammatory factor-1 in dendritic cells restrains CD4+ T cell effector responses and induces CD25+Foxp3+ T regulatory subsets,” *Frontiers in Immunology*, 8(NOV). doi: 10.3389/fimmu.2017.01502.
- Elliott, K., Stephenson, T. J. and Messenger, A. G. (1999) “Differences in hair follicle dermal papilla volume are due to extracellular matrix volume and cell number: Implications for the control of hair follicle size and androgen responses,” *Journal of Investigative Dermatology*, 113(6), pp. 873–877. doi: 10.1046/j.1523-1747.1999.00797.x.

- Enzerink, A. and Vaheri, A. (2011) "Fibroblast activation in vascular inflammation," *Journal of Thrombosis and Haemostasis*, 9(4) pp. 619–626. doi: 10.1111/j.1538-7836.2011.04209.x.
- Ernst, J. *et al.* (2011) "Mapping and analysis of chromatin state dynamics in nine human cell types," *Nature*, 473(7345), pp. 43–49. doi: 10.1038/nature09906.
- Fane, M. and Weeraratna, A. T. (2020) "How the ageing microenvironment influences tumour progression," *Nature Reviews Cancer*, 20(2), pp. 89–106. doi: 10.1038/s41568-019-0222-9.
- Farage, M. A. *et al.* (2008) "Intrinsic and extrinsic factors in skin ageing: A review," *International Journal of Cosmetic Science*, 30(2), pp. 87–95. doi: 10.1111/j.1468-2494.2007.00415.x.
- Finnegan, A. *et al.* (2019) "Single-cell transcriptomics reveals spatial and temporal turnover of keratinocyte differentiation regulators," *Frontiers in Genetics*, 10(JUL), p. 775. doi: 10.3389/fgene.2019.00775.
- Fisher, G. J. *et al.* (2009) "Collagen Fragmentation Promotes Oxidative Stress and Elevates Matrix Metalloproteinase-1 in Fibroblasts in Aged Human Skin," *The American Journal of Pathology*, 174(1), pp. 101–114. doi: 10.2353/AJPATH.2009.080599.
- Fisher, G. J. *et al.* (2016) "Reduction of fibroblast size/mechanical force down-regulates TGF- $\beta$  type II receptor: Implications for human skin aging," *Aging Cell*, 15(1), pp. 67–76. doi: 10.1111/accel.12410.
- Fitzmaurice, C. *et al.* (2019) "Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-Adjusted life-years for 29 cancer groups, 1990 to 2017: A systematic analysis for the global burden of disease study," *JAMA Oncology*, 5(12), pp. 1749–1768. doi: 10.1001/jamaoncol.2019.2996.
- Flavahan, W. A., Gaskell, E. and Bernstein, B. E. (2017) "Epigenetic plasticity and the hallmarks of cancer," *Science*, 357(6348). doi: 10.1126/science.aal2380.
- Franceschi, C. and Campisi, J. (2014) "Chronic inflammation (Inflammaging) and its potential contribution to age-associated diseases," *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, 69, pp. S4–S9. doi: 10.1093/gerona/glu057.
- Gao, C. F. *et al.* (2005) "Proliferation and invasion: Plasticity in tumor cells," *Proceedings of the National Academy of Sciences of the United States of America*, 102(30), pp. 10528–10533. doi: 10.1073/pnas.0504367102.
- Garcovich, S. *et al.* (2017) "Skin Cancer Epidemics in the Elderly as An Emerging Issue in Geriatric Oncology," *Aging and Disease*, 8(5), p. 643. doi: 10.14336/AD.2017.0503.
- Ghosh, K. and Capell, B. C. (2016) "The Senescence-Associated Secretory Phenotype: Critical Effector in Skin Cancer and Aging," *Journal of Investigative Dermatology*, 136(11), pp. 2133–2139. doi: 10.1016/j.jid.2016.06.621.
- Giangreco, A. *et al.* (2010) "Human skin aging is associated with reduced expression of the stem cell markers  $\beta$ 1 integrin and MCSP," *Journal of Investigative Dermatology*, 130(2), pp. 604–608. doi: 10.1038/jid.2009.297.
- Gilbert, M. T. P. *et al.* (2007) "The Isolation of Nucleic Acids from Fixed, Paraffin-Embedded Tissues-Which Methods Are Useful When?," *PLoS ONE*, 2(6), p. e537 doi: 10.1371/journal.pone.0000537.
- Gkoutela, S. *et al.* (2019) "Circulating Tumor Cell Clustering Shapes DNA Methylation to Enable Metastasis Seeding," *Cell*, 176(1–2), pp. 98–112.e14. doi: 10.1016/j.cell.2018.11.046.
- Glogau, R. G. (2000) "The risk of progression to invasive disease," *Journal of the American Academy of Dermatology*, 42(1 SUPPL. 1), pp. S23–S24. doi: 10.1067/mjd.2000.103339.

- Goldberg, M. *et al.* (2006) "Epigenetic silencing contributes to frequent loss of the fragile histidine triad tumour suppressor in basal cell carcinomas," *British Journal of Dermatology*, 155(6), pp. 1154–1158. doi: 10.1111/j.1365-2133.2006.07433.x.
- Golden, T. R., Hinerfeld, D. A. and Melov, S. (2002) "Oxidative stress and aging: beyond correlation, Aging Cell". *Aging Cell*, 1(2), pp. 117–123. Doi: 10.1046/j.1474-9728.2002.00015.x.
- Goll, M. G. *et al.* (2006) "Methylation of tRNA<sup>Asp</sup> by the DNA methyltransferase homolog Dnmt2," *Science*, 311(5759), pp. 395–398. doi: 10.1126/science.1120976.
- Gonzales, K. A. U. and Fuchs, E. (2017) "Skin and Its Regenerative Powers: An Alliance between Stem Cells and Their Niche," *Developmental Cell*, 43(4), pp. 387–401. doi: 10.1016/j.devcel.2017.10.001.
- Gosain, A. and DiPietro, L. A. (2004) "Aging and Wound Healing," *World Journal of Surgery*, 28(3), pp. 321–326. doi: 10.1007/s00268-003-7397-6.
- Greaves, M. and Maley, C. C. (2012) "Clonal evolution in cancer," *Nature*, 481(7381), pp. 306–313. doi: 10.1038/nature10762.
- Greenberg, M. V. C. and Bourc'his, D. (2019) "The diverse roles of DNA methylation in mammalian development and disease," *Nature Reviews Molecular Cell Biology*, 20, pp. 590–607. doi: 10.1038/s41580-019-0159-6.
- Greenblatt, M. S. *et al.* (1994) "Mutations in the p53 Tumor Suppressor Gene: Clues to Cancer Etiology and Molecular Pathogenesis," *Cancer Research*, 54(18), pp. 4855–4878.
- Gregory, P. A. *et al.* (2008) "The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1," *Nature Cell Biology*, 10(5), pp. 593–601. doi: 10.1038/ncb1722.
- Griffin, M. F. *et al.* (2020) "Understanding the impact of fibroblast heterogeneity on skin fibrosis," *DMM Disease Models and Mechanisms*, 13(6), p. dmm044164. doi: 10.1242/dmm.044164.
- Haniffa, M. A. *et al.* (2007) "Adult Human Fibroblasts Are Potent Immunoregulatory Cells and Functionally Equivalent to Mesenchymal Stem Cells," *The Journal of Immunology*, 179(3), pp. 1595–1604. doi: 10.4049/jimmunol.179.3.1595.
- Hansen, K. (2017) "IlluminaHumanMethylationEPICanno.ilm10b4.hg19: Annotation for Illumina's EPIC methylation arrays. R package version 0.6.0."
- Hashimoto, H. *et al.* (2012) "Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation," *Nucleic Acids Research*, 40(11), pp. 4841–4849. doi: 10.1093/nar/gks155.
- Haydont, V. *et al.* (2019) "Transcriptome profiling of human papillary and reticular fibroblasts from adult interfollicular dermis pinpoints the 'tissue skeleton' gene network as a component of skin chrono-ageing," *Mechanisms of Ageing and Development*, 179, pp. 60–77. doi: 10.1016/j.mad.2019.01.003.
- Haydont, V., Bernard, B. A. and Fortunel, N. O. (2019) "Age-related evolutions of the dermis: Clinical signs, fibroblast and extracellular matrix dynamics," *Mechanisms of Ageing and Development*, 177, pp. 150–156. doi: 10.1016/j.mad.2018.03.006.
- He, H. *et al.* (2020) "Single-cell transcriptome analysis of human skin identifies novel fibroblast subpopulation and enrichment of immune subsets in atopic dermatitis," *Journal of Allergy and Clinical Immunology*, 145(6), pp. 1615–1628. doi: 10.1016/j.jaci.2020.01.042.
- Heaton, H. and Lawrence, N. (2019) "Nonmelanoma skin cancer in women," *International Journal of Women's Dermatology*, 5(1), pp. 2–7. doi: 10.1016/j.ijwd.2018.08.007.

- Hecht, I. *et al.* (2015) "The motility-proliferation-metabolism interplay during metastatic invasion," *Scientific Reports*, 5(1), pp. 1–10. doi: 10.1038/srep13538.
- Heenan, P. J. and Bogle, M. S. (1993) "Eccrine differentiation in basal cell carcinoma," *Journal of Investigative Dermatology*, 100(3 SUPPL.), pp. S295–S299. doi: 10.1038/jid.1993.52.
- Heitzer, E. *et al.* (2010) "PTCH promoter methylation at low level in sporadic basal cell carcinoma analysed by three different approaches," *Experimental Dermatology*, 19(10), pp. 926–928. doi: 10.1111/j.1600-0625.2010.01120.x.
- Hernando, B. *et al.* (2021) "The effect of age on the acquisition and selection of cancer driver mutations in sun-exposed normal skin," *Annals of Oncology*, 32(3), pp. 412–421. doi: 10.1016/j.annonc.2020.11.023.
- Hervás-Marín, D. *et al.* (2019) "Genome wide DNA methylation profiling identifies specific epigenetic features in highrisk cutaneous squamous cell carcinoma," *PLoS ONE*, 14(12), p. e0223341. doi: 10.1371/journal.pone.0223341.
- Hoadley, K. A. *et al.* (2018) "Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer," *Cell*, 173(2), pp. 291-304.e6. doi: 10.1016/j.cell.2018.03.022.
- Hoek, K. S. *et al.* (2008) "In vivo switching of human melanoma cells between proliferative and invasive states," *Cancer Research*, 68(3), pp. 650–656. doi: 10.1158/0008-5472.CAN-07-2491.
- Hogervorst, M. *et al.* (2018) "A shift from papillary to reticular fibroblasts enables tumour–stroma interaction and invasion," *British Journal of Cancer*, (December 2017), pp. 1–9. doi: 10.1038/s41416-018-0024-y.
- Holliday, R. and Grigg, G. W. (1993) "DNA methylation and mutation," *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, 285(1), pp. 61–67. doi: 10.1016/0027-5107(93)90052-H.
- Hon, G. C. *et al.* (2014) "5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation," *Molecular Cell*, 56(2), pp. 286–297. doi: 10.1016/j.molcel.2014.08.026.
- Hou, Y. *et al.* (2016) "Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas," *Cell Research*, 26(3), pp. 304–319. doi: 10.1038/cr.2016.23.
- Hsu, Y. C., Li, L. and Fuchs, E. (2014) "Emerging interactions between skin stem cells and their niches," *Nature Medicine*, 20(8), pp. 847–856. doi: 10.1038/nm.3643.
- Huang, P. Y. and Balmain, A. (2014) "Modeling cutaneous squamous carcinoma development in the mouse," *Cold Spring Harbor Perspectives in Medicine*, 4(9). doi: 10.1101/cshperspect.a013623.
- Hwang, B., Lee, J. H. and Bang, D. (2018) "Single-cell RNA sequencing technologies and bioinformatics pipelines," *Experimental and Molecular Medicine*, 50:96. doi: 10.1038/s12276-018-0071-8.
- Ichim, T. E., O’Heeron, P. and Kesari, S. (2018) "Fibroblasts as a practical alternative to mesenchymal stem cells," *Journal of Translational Medicine*, 16:212. doi: 10.1186/s12967-018-1536-1.
- Iida, S. *et al.* (2020) "Verruca vulgaris and seborrheic keratosis exacerbated by immunosuppression," *Case Reports in Dermatological Medicine*, 2020:6682694. doi: 10.1155/2020/6682694.
- Ikehata, H. and Ono, T. (2011) "The mechanisms of UV mutagenesis," *Journal of Radiation Research*, 52(2), pp. 115–125. doi: 10.1269/jrr.10175.
- Inman, G. J. *et al.* (2018) "The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature," *Nature Communications*, 9(1), pp. 1–14. doi: 10.1038/s41467-018-06027-1.

## 8. References

---

- Ishida, H. *et al.* (2002) “Comparative histochemical study of Bowen ’ s disease and actinic keratosis : preserved normal basal cells in Bowen ’ s disease,” *European Journal of Histochemistry*, 45(2), pp. 177–190. doi: 10.4081/1628.
- Ito, S. *et al.* (2011) “Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine,” *Science*, 333(6047), pp. 1300–1303. doi: 10.1126/science.1210597.
- Janson, D. G. *et al.* (2012) “Different gene expression patterns in human papillary and reticular fibroblasts,” *Journal of Investigative Dermatology*, 132(11), pp. 2565–2572. doi: 10.1038/jid.2012.192.
- Jayaraman, S. S. *et al.* (2014) “Mutational landscape of basal cell carcinomas by whole-exome sequencing,” *Journal of Investigative Dermatology*, 134(1), pp. 213–220. doi: 10.1038/jid.2013.276.
- Ji, A. L. *et al.* (2020) “Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma,” *Journal of Cleaner Production*, 182(2), pp. 497-514.e22. doi: 10.1016/j.cell.2020.05.039.
- Kanzaki, R. and Pietras, K. (2020) “Heterogeneity of cancer-associated fibroblasts: Opportunities for precision medicine,” *Cancer Science*, 111(8), pp. 2708–2717. doi: 10.1111/cas.14537.
- Kao, G. F. (1986) “Carcinoma Arising in Bowen’s Disease,” *Archives of Dermatology*, 122(10), pp. 1124–1126. doi: 10.1001/archderm.1986.01660220042010.
- Karaayvaz, M. *et al.* (2018) “Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq,” *Nature Communications*, 9(1), pp. 1–10. doi: 10.1038/s41467-018-06052-0.
- Karemaker, I. D. and Vermeulen, M. (2018) “Single-Cell DNA Methylation Profiling: Technologies and Biological Applications,” *Trends in Biotechnology*, 36(9), pp. 952–965. doi: 10.1016/j.tibtech.2018.04.002.
- Kasper, M. *et al.* (2011) “Wounding enhances epidermal tumorigenesis by recruiting hair follicle keratinocytes,” *Proceedings of the National Academy of Sciences of the United States of America*, 108(10), pp. 4099–4104. doi: 10.1073/pnas.1014489108.
- Kiesslich, T., Pichler, M. and Neureiter, D. (2013) “Epigenetic control of epithelial-mesenchymal-transition in human cancer,” *Molecular and Clinical Oncology*, 1(1), pp. 3–11. doi: 10.3892/mco.2012.28.
- Kim, D., Chung, K. B. and Kim, T. G. (2020) “Application of single-cell RNA sequencing on human skin: Technical evolution and challenges,” *Journal of Dermatological Science*, 99(2), pp. 74–81. doi: 10.1016/j.jdermsci.2020.06.002.
- Kligman, A. M. (1979) “Perspectives and problems in cutaneous gerontology,” *Journal of Investigative Dermatology*, 73(1), pp. 39–46. doi: 10.1111/1523-1747.ep12532758.
- Kling, T. *et al.* (2017) “Validation of the MethylationEPIC BeadChip for fresh-frozen and formalin-fixed paraffin-embedded tumours,” *Clinical Epigenetics*, 9(1), pp. 1–6. doi: 10.1186/s13148-017-0333-7.
- Kolarsick, P. A. J., Kolarsick, M. A. and Goodwin, C. (2011) “Anatomy and Physiology of the Skin,” *Journal of the Dermatology Nurses’ Association*, 3(4), pp. 203–213. doi: 10.1097/JDN.0b013e3182274a98.
- Korgavkar, K., Xiong, M. and Weinstock, M. (2013) “Changing incidence trends of cutaneous T-cell lymphoma,” *JAMA Dermatology*, 149(11), pp. 1295–1299. doi: 10.1001/jamadermatol.2013.5526.
- Korosec, A. *et al.* (2019) “Lineage Identity and Location within the Dermis Determine the Function of Papillary and Reticular Fibroblasts in Human Skin,” *Journal of Investigative Dermatology*, 139(2), pp. 342–351. doi: 10.1016/j.jid.2018.07.033.
- Kossard, S. and Rosen, R. (1992) “Cutaneous Bowen’s disease: An analysis of 1001 cases according to age, sex, and site,” *Journal of the American Academy of Dermatology*, 27(3), pp. 406–410. doi: 10.1016/0190-9622(92)70208-W.

- Krishnaswami, S. R. *et al.* (2016) "Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons," *Nature Protocols*, 11(3), pp. 499–524. doi: 10.1038/nprot.2016.015.
- Krüger, K., Blume-Peytavi, U. and Orfanos, C. E. (1999) "Basal cell carcinoma possibly originates from the outer root sheath and/or the bulge region of the vellus hair follicle," *Archives of Dermatological Research*, 291(5), pp. 253–259. doi: 10.1007/s004030050405.
- Krutmann, J. *et al.* (2017) "The skin aging exposome." *Journal of Dermatological Science*, 85(2017), pp. 152–161. doi: 10.1016/j.jdermsci.2016.09.015.
- Kulis, M. *et al.* (2013) "Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer," *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1829(11), pp. 1161–1174. doi: 10.1016/j.bbagr.2013.08.001.
- Lambert, A. W., Pattabiraman, D. R. and Weinberg, R. A. (2017) "Emerging Biological Principles of Metastasis," *Cell*, 168(4), pp. 670–691. doi: 10.1016/j.cell.2016.11.037.
- Lambert, S. R. *et al.* (2014) "Key differences identified between actinic keratosis and cutaneous squamous cell carcinoma by transcriptome profiling," *British Journal of Cancer*, 110(2), pp. 520–529. doi: 10.1038/bjc.2013.760.
- Langton, A. K. *et al.* (2012) "Differential expression of elastic fibre components in intrinsically aged skin," *Biogerontology*, 13(1), pp. 37–48. doi: 10.1007/s10522-011-9332-9.
- Langton, A. K. *et al.* (2016) "The impact of intrinsic ageing on the protein composition of the dermal-epidermal junction," *Mechanisms of Ageing and Development*, 156, pp. 14–16. doi: 10.1016/j.mad.2016.03.006.
- Lapouge, G. *et al.* (2011) "Identifying the cellular origin of squamous skin tumors," *Proceedings of the National Academy of Sciences of the United States of America*, 108(18), pp. 7431–7436. doi: 10.1073/pnas.1012720108.
- Latif, N. *et al.* (2015) "Expression of smoothmuscle cellmarkers and co-activators in calcified aortic valves," *European Heart Journal*, 36(21), pp. 1335–1345. doi: 10.1093/eurheartj/ehf547.
- Latil, M. *et al.* (2017) "Cell-Type-Specific Chromatin States Differentially Prime Squamous Cell Carcinoma Tumor-Initiating Cells for Epithelial to Mesenchymal Transition," *Cell Stem Cell*, 20(2), pp. 191–204.e5. doi: 10.1016/j.stem.2016.10.018.
- Lee, L.-Y. and Liu, S.-X. (2020) "Pathogenesis of Photoaging in Human Dermal Fibroblasts," *International Journal of Dermatology and Venereology*, 3(1), pp. 37–42. doi: 10.1097/jd9.0000000000000068.
- Lee, S. J. *et al.* (2015) "Generation of pure lymphatic endothelial cells from human pluripotent stem cells and their therapeutic effects on wound repair," *Scientific Reports*, 5:11019. doi: 10.1038/srep11019.
- Lee, Y. I. *et al.* (2021) "Cellular senescence and inflammaging in the skin microenvironment," *International Journal of Molecular Sciences*, 22:3849. doi: 10.3390/ijms22083849.
- Leeuwenberg, J. F. *et al.* (1992) "E-selectin and intercellular adhesion molecule-1 are released by activated human endothelial cells in vitro.," *Immunology*, 77(4), pp. 543–549.
- Li, A. *et al.* (2018) "Chondrogenic properties of collagen type XI, a component of cartilage extracellular matrix," *Biomaterials*, 173, pp. 47–57. doi: 10.1016/j.biomaterials.2018.05.004.
- Li, J. *et al.* (2012) "Progressive alopecia reveals decreasing stem cell activation probability during aging of mice with epidermal deletion of DNA methyltransferase 1," *Journal of Investigative Dermatology*, 132(12), pp. 2681–2690. doi: 10.1038/jid.2012.206.

- Li, Y. *et al.* (2009) "Methylation-associated silencing of S100A4 expression in human epidermal cancers," *Experimental Dermatology*, 18(10), pp. 842–848. doi: 10.1111/j.1600-0625.2009.00922.x.
- Li, Y., Sawalha, A. H. and Lu, Q. (2009) "Aberrant DNA methylation in skin diseases," *Journal of Dermatological Science*, 54(3), pp. 143–149. doi: 10.1016/j.jdermsci.2009.01.009.
- López-Otín, C. *et al.* (2013) "The hallmarks of aging," *Cell*, 153(6), p. 1194. doi: 10.1016/j.cell.2013.05.039.
- Lowenstein, S. E. *et al.* (2017) "Risk prediction tools for keratinocyte carcinoma after solid organ transplantation: a review of the literature," *British Journal of Dermatology*, pp. 1–6. doi: 10.1111/bjd.15889.
- Luebberding, S., Krueger, N. and Kerscher, M. (2013) "Age-related changes in skin barrier function - Quantitative evaluation of 150 female subjects," *International Journal of Cosmetic Science*, 35(2), pp. 183–190. doi: 10.1111/ics.12024.
- Lyko, F. (2018) "The DNA methyltransferase family: A versatile toolkit for epigenetic regulation," *Nature Reviews Genetics*, 19(2), pp. 81–92. doi: 10.1038/nrg.2017.80.
- Lynch, M. D. and Watt, F. M. (2018) "Fibroblast heterogeneity: implications for human disease," *Journal of Clinical Investigation*, 128(1), pp. 26–35. doi: 10.1172/JCI93555.
- Madeleine, M. M. *et al.* (2017) "Epidemiology of keratinocyte carcinomas after organ transplantation," *British Journal of Dermatology*, pp. 1–9. doi: 10.1111/bjd.15931.
- Maeda, K. (2017) "New method of measurement of epidermal turnover in humans," *Cosmetics*, 4(47). doi: 10.3390/cosmetics4040047.
- Maksimovic, S. *et al.* (2014) "Epidermal Merkel cells are mechanosensory cells that tune mammalian touch receptors," *Nature*, 509(7502), pp. 617–621. doi: 10.1038/nature13250.
- la Manno, G. *et al.* (2018) "RNA velocity of single cells," *Nature*, 560(7719), pp. 494–498. doi: 10.1038/s41586-018-0414-6.
- Marks, R., Rennie, G. and Selwood, T. S. (1988) "MALIGNANT TRANSFORMATION OF SOLAR KERATOSES TO SQUAMOUS CELL CARCINOMA," *The Lancet*, 331(8589), pp. 795–797. doi: 10.1016/S0140-6736(88)91658-3.
- Martelotto, L. (2020) "Human Cell Atlas Method Development Community." doi: 10.17504/protocols.io.3fkjkw.
- McAnulty, R. J. (2007) "Fibroblasts and myofibroblasts: Their source, function and role in disease," *International Journal of Biochemistry and Cell Biology*, 39(4), pp. 666–671. doi: 10.1016/j.biocel.2006.11.005.
- Michel, M. *et al.* (1996) "Keratin 19 as a biochemical marker of skin stem cells in vivo and in vitro: Keratin 19 expressing cells are differentially localized in function of anatomic sites, and their number varies with donor age and culture stage," *Journal of Cell Science*, 109(5), pp. 1017–1028. doi: 10.1242/jcs.109.5.1017.
- Mine, S. *et al.* (2008) "Aging alters functionally human dermal papillary fibroblasts but not reticular fibroblasts: A new view of skin morphogenesis and aging," *PLoS ONE*, 3(12), p. e4066 doi: 10.1371/journal.pone.0004066.
- Miroshnikova, Y. A. *et al.* (2018) "Adhesion forces and cortical tension couple cell proliferation and differentiation to drive epidermal stratification," *Nature Cell Biology*, 20(1), pp. 69–80. doi: 10.1038/s41556-017-0005-z.
- Mizumoto, N. and Takashima, A. (2004) "CD1a and langerin: acting as more than Langerhans cell markers," *Journal of Clinical Investigation*, 113(5), pp. 658–660. doi: 10.1172/jci21140.

- Mofidi, A. *et al.* (2018) “The economic burden of occupational non-melanoma skin cancer due to solar radiation,” *Journal of Occupational and Environmental Hygiene*, 15(6), pp. 481–491. doi: 10.1080/15459624.2018.1447118.
- Moll, R., Divo, M. and Langbein, L. (2008) “The human keratins: Biology and pathology,” *Histochemistry and Cell Biology*, 129(6), pp. 705–733. doi: 10.1007/s00418-008-0435-6.
- Moran, S. *et al.* (2016) “Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis,” 2045(16), pp. 1–9. doi: 10.1016/S1470-2045(16)30297-2.
- Moran, S., Arribas, C. and Esteller, M. (2016) “Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences,” *Epigenomics*, 8(3), pp. 389–399. doi: 10.2217/epi.15.114.
- Moreci, R. S. and Lechler, T. (2020) “Epidermal structure and differentiation,” *Current Biology*, 30(4), pp. R144–R149. doi: 10.1016/j.cub.2020.01.004.
- Mulqueen, R. *et al.* (2019) “Improved single-cell ATAC-seq reveals chromatin dynamics of in vitro corticogenesis,” *bioRxiv*, p. 637256. doi: 10.1101/637256.
- Mulqueen, R. M. *et al.* (2018) “Highly scalable generation of DNA methylation profiles in single cells,” *Nature Biotechnology*, 36(5), pp. 428–431. doi: 10.1038/nbt.4112.
- Murao, K. *et al.* (2006) “Epigenetic abnormalities in cutaneous squamous cell carcinomas: Frequent inactivation of the RB1/p16 and p53 pathways,” *British Journal of Dermatology*, 155(5), pp. 999–1005. doi: 10.1111/j.1365-2133.2006.07487.x.
- Nagarajan, P. *et al.* (2019) “Keratinocyte carcinomas: Current concepts and future research priorities,” *Clinical Cancer Research*, 25(8), pp. 2379–2391. doi: 10.1158/1078-0432.CCR-18-1122.
- Nath, A. and Bild, A. H. (2021) “Leveraging Single-Cell Approaches in Cancer Precision Medicine,” *Trends in Cancer*, 7(4), pp. 359–372. doi: 10.1016/j.trecan.2021.01.007.
- Nauroy, P. *et al.* (2017) “Human Dermal Fibroblast Subpopulations Display Distinct Gene Signatures Related to Cell Behaviors and Matrisome,” *Journal of Investigative Dermatology*, 137(8), pp. 1787–1789. doi: 10.1016/j.jid.2017.03.028.
- Neely, A. E. and Bao, X. (2019) “Nuclei Isolation Staining (NIS) Method for Imaging Chromatin-Associated Proteins in Difficult Cell Types,” *Current Protocols in Cell Biology*, 84(1), p. e94. doi: 10.1002/cpcb.94.
- Neves, R. *et al.* (2010) “Role of DNA methylation in miR-200c/141 cluster silencing in invasive breast cancer cells,” *BMC Research Notes*, 3, p. 219. doi: 10.1186/1756-0500-3-219.
- Oakes, C. C. *et al.* (2016) “DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia,” *Nature Genetics*, 48(3), pp. 253–264. doi: 10.1038/ng.3488.
- de Oliveira, N. F. P., de Souza, B. F. and de Castro Coêlho, M. (2020) “UV Radiation and Its Relation to DNA Methylation in Epidermal Cells: A Review,” *Epigenomes*, 4(4), p. 23. doi: 10.3390/epigenomes4040023.
- Ooi, S. K. T. *et al.* (2007) “DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA,” *Nature*, 448(7154), pp. 714–717. doi: 10.1038/nature05987.
- Pacifico, A. *et al.* (2008) “Loss of CDKN2A and p14ARF expression occurs frequently in human nonmelanoma skin cancers,” *British Journal of Dermatology*, 158(2), pp. 291–297. doi: 10.1111/j.1365-2133.2007.08360.x.

## 8. References

---

- Padilla, R. S. *et al.* (2010) "Gene expression patterns of normal human skin, actinic keratosis, and squamous cell carcinoma: A spectrum of disease progression," *Archives of Dermatology*, 146(3), pp. 288–293. doi: 10.1001/archdermatol.2009.378.
- Paquet-Fifield, S. *et al.* (2009) "A role for pericytes as microenvironmental regulators of human skin tissue regeneration," *Journal of Clinical Investigation*, 119(9), pp. 2795–2806. doi: 10.1172/JCI38535.
- Park, S. M. *et al.* (2008) "The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2," *Genes and Development*, 22(7), pp. 894–907. doi: 10.1101/gad.1640608.
- Paulson, K. G. *et al.* (2018) "Merkel cell carcinoma: Current US incidence and projected increases based on changing demographics," *Journal of the American Academy of Dermatology*, 78(3), pp. 457–463.e2. doi: 10.1016/j.jaad.2017.10.028.
- Pawlina, W. & Ross, M. (2015) *Histology: A Text and Atlas with Correlated Cell and Molecular Biology*. Wolters Kluwer Health.
- Peltonen, S. *et al.* (1999) "A novel component of epidermal cell-matrix and cell-cell contacts: Transmembrane protein type XIII collagen," *Journal of Investigative Dermatology*, 113(4), pp. 635–642. doi: 10.1046/j.1523-1747.1999.00736.x.
- Peterson, S. C. *et al.* (2015) "Basal cell carcinoma preferentially arises from stem cells within hair follicle and mechanosensory niches," *Cell Stem Cell*, 16(4), pp. 400–412. doi: 10.1016/j.stem.2015.02.006.
- Pfeifer, G. P., Tang, M. S. and Denissenko, M. F. (2000) "Mutation hotspots and DNA methylation," *Current Topics in Microbiology and Immunology*, 249, pp. 1–19. doi: 10.1007/978-3-642-59696-4\_1.
- Philippeos, C. *et al.* (2018) "Spatial and Single-Cell Transcriptional Profiling Identifies Functionally Distinct Human Dermal Fibroblast Subpopulations," *Journal of Investigative Dermatology*, 138(4), pp. 811–825. doi: 10.1016/j.jid.2018.01.016.
- Picelli, S. *et al.* (2014) "Tn5 transposase and tagmentation procedures for massively scaled sequencing projects," *Genome Research*, 24(12), pp. 2033–2040. doi: 10.1101/gr.177881.114.
- Pixberg, C. F. *et al.* (2017) "Analysis of DNA methylation in single circulating tumor cells," *Oncogene*, 36(23), pp. 3223–3231. doi: 10.1038/onc.2016.480.
- Pontiggia, L. *et al.* (2009) "Markers to evaluate the quality and self-renewing potential of engineered human skin substitutes in vitro and after transplantation," *Journal of Investigative Dermatology*, 129(2), pp. 480–490. doi: 10.1038/jid.2008.254.
- Puisieux, A. *et al.* (2018) "Cellular Pliancy and the Multistep Process of Tumorigenesis," *Cancer Cell*, 33(2), pp. 164–172. doi: 10.1016/j.ccell.2018.01.007.
- Quan, T. *et al.* (2013) "Elevated matrix metalloproteinases and collagen fragmentation in photodamaged human skin: Impact of altered extracellular matrix microenvironment on dermal fibroblast function," *Journal of Investigative Dermatology*, 133(5), pp. 1362–1366. doi: 10.1038/jid.2012.509.
- Rachidi, W. *et al.* (2007) "Sensing radiosensitivity of human epidermal stem cells," *Radiotherapy and Oncology*, 83(3), pp. 267–276. doi: 10.1016/j.radonc.2007.05.007.
- Rahmani, W. *et al.* (2014) "Hair Follicle Dermal Stem Cells Regenerate the Dermal Sheath, Repopulate the Dermal Papilla, and Modulate Hair Type," *Developmental Cell*, 31(5), pp. 543–558. doi: 10.1016/j.devcel.2014.10.022.

- Ratushny, V. *et al.* (2012) “Review series: From keratinocyte to cancer : the pathogenesis and modeling of cutaneous squamous cell carcinoma,” *The Journal of Clinical Investigation*, 122(2), pp. 464–72. doi: 10.1172/JCI57415.464.
- Reeves, M. Q. *et al.* (2018) “Multicolour lineage tracing reveals clonal dynamics of squamous carcinoma evolution from initiation to metastasis,” *Nature Cell Biology*, 20(6), pp. 699–709. doi: 10.1038/s41556-018-0109-0.
- Ren, X. and Kuan, P. F. (2019) “methylGSA: a Bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing,” *Bioinformatics (Oxford, England)*. Edited by J. Hancock, 35(11), pp. 1958–1959. doi: 10.1093/bioinformatics/bty892.
- Reynolds, G. *et al.* (2021) “Developmental cell programs are co-opted in inflammatory skin disease,” *Science*, 371(6527). doi: 10.1126/science.aba6500.
- Ricard-Blum, S. (2011) “The Collagen Family,” *Cold Spring Harbor Perspectives in Biology*, 3(1), pp. 1–19. doi: 10.1101/cshperspect.a004978.
- Rinaldi, L. *et al.* (2016) “Dnmt3a and Dnmt3b Associate with Enhancers to Regulate Human Epidermal Stem Cell Homeostasis,” *Cell Stem Cell*, 19(4), pp. 491–501. doi: 10.1016/j.stem.2016.06.020.
- Rinaldi, L. *et al.* (2017) “Loss of dnmt3a and dnmt3b does not affect epidermal homeostasis but promotes squamous transformation through PPAR-g,” *eLife*, 6. doi: 10.7554/eLife.21697.
- Rinn, J. L. *et al.* (2006) “Anatomic demarcation by positional variation in fibroblast gene expression programs,” *PLoS Genetics*, 2(7), pp. 1084–1096. doi: 10.1371/journal.pgen.0020119.
- Io Riso, P. *et al.* (2020) “A cell-of-origin epigenetic tracer reveals clinically distinct subtypes of high-grade serous ovarian cancer,” *Genome Medicine*, 12(1), p. 94. doi: 10.1186/s13073-020-00786-7.
- Rittié, L. and Fisher, G. J. (2015) “Natural and sun-induced aging of human skin,” *Cold Spring Harbor Perspectives in Medicine*, 5(1). doi: 10.1101/cshperspect.a015370.
- Rivera-Gonzalez, G., Shook, B. and Horsley, V. (2014) “Adipocytes in skin health and disease,” *Cold Spring Harbor Perspectives in Medicine*, 4(3). doi: 10.1101/cshperspect.a015271.
- Rodríguez-Paredes, M. *et al.* (2018) “Methylation profiling identifies two subclasses of squamous cell carcinoma related to distinct cells of origin,” *Nature Communications*, 9(1), pp. 1–9. doi: 10.1038/s41467-018-03025-1.
- Rodríguez-Paredes, M. and Esteller, M. (2011) “Cancer epigenetics reaches mainstream oncology,” *Nature Medicine*, 17(3), pp. 330–339. doi: 10.1038/nm.2305.
- Rognoni, E. and Watt, F. M. (2018) “Skin Cell Heterogeneity in Development, Wound Healing, and Cancer,” *Trends in Cell Biology*, xx, pp. 1–14. doi: 10.1016/j.tcb.2018.05.002.
- Roskos, K. v., Maibach, H. I. and Guy, R. H. (1989) “The effect of aging on percutaneous absorption in man,” *Journal of Pharmacokinetics and Biopharmaceutics*, 17(6), pp. 617–630. doi: 10.1007/BF01062121.
- Saglam, O. *et al.* (2008) “Immunohistochemical staining of palisading basal cells in Bowen’s disease and basal involvement in actinic keratosis: Contrasting staining patterns suggest different cells of origin,” *American Journal of Dermatopathology*, 30(2), pp. 123–126. doi: 10.1097/DAD.0b013e3181658062.
- Salzer, M. C. *et al.* (2018) “Identity Noise and Adipogenic Traits Characterize Dermal Fibroblast Aging,” *Cell*, 175(6), pp. 1575–1590.e22. doi: 10.1016/j.cell.2018.10.012.

## 8. References

---

- Sand, M. *et al.* (2019) “Dicer sequencing, whole genome methylation profiling, mRNA and smallRNA sequencing analysis in basal cell carcinoma,” *Cellular Physiology and Biochemistry*, 53(5), pp. 760–773. doi: 10.33594/000000171.
- Satpathy, A. T. *et al.* (2019) “Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion,” *Nature Biotechnology*, 37(8), pp. 925–936. doi: 10.1038/s41587-019-0206-z.
- Saxonov, S., Berg, P. and Brutlag, D. L. (2006) “A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters,” *Proceedings of the National Academy of Sciences of the United States of America*, 103(5), pp. 1412–1417. doi: 10.1073/pnas.0510310103.
- Schafer, I. A. *et al.* (1985) “Comparative observation of fibroblasts derived from the papillary and reticular dermis of infants and adults: Growth kinetics, packing density at confluence and surface morphology,” *Mechanisms of Ageing and Development*, 31(3), pp. 275–293. doi: 10.1016/0047-6374(85)90095-8.
- Schmitz, L. (2019) “Histological classification of cutaneous squamous cell carcinomas with different severity,” 33, pp. 11–15. doi: 10.1111/jdv.15950.
- Schonherr, E. *et al.* (1993) “Differences in decorin expression by papillary and reticular fibroblasts in vivo and in vitro,” *Biochemical Journal*, 290(3), pp. 893–899. doi: 10.1042/bj2900893.
- Sellheyer, K. (2011) “Basal cell carcinoma: Cell of origin, cancer stem cell hypothesis and stem cell markers,” *British Journal of Dermatology*, 164(4), pp. 696–711. doi: 10.1111/j.1365-2133.2010.10158.x.
- Sen, G. L. *et al.* (2010) “DNMT1 maintains progenitor function in self-renewing somatic tissue,” *Nature*, 463(7280), pp. 563–567. doi: 10.1038/nature08683.
- Shen, H. and Laird, P. W. (2013) “Interplay between the cancer genome and epigenome,” *Cell*, 153(1), pp. 38–55. doi: 10.1016/j.cell.2013.03.008.
- Shin, J. W. *et al.* (2019) “Molecular mechanisms of dermal aging and antiaging approaches,” *International Journal of Molecular Sciences*, 20:2126. doi: 10.3390/ijms20092126.
- Shin, W. *et al.* (2020) “Dysfunction of Hair Follicle Mesenchymal Progenitors Contributes to Age-Associated Hair Loss,” *Developmental Cell*, 53(2), pp. 185–198.e7. doi: 10.1016/j.devcel.2020.03.019.
- Smith, R. S. *et al.* (1997) “Fibroblasts as sentinel cells. Synthesis of chemokines and regulation of inflammation.” *The American journal of pathology*, 151(2), pp. 317–322.
- Smith, Z. D. and Meissner, A. (2013) “DNA methylation: Roles in mammalian development,” *Nature Reviews Genetics*, 14(3), pp. 204–220. doi: 10.1038/nrg3354.
- Solanas, G. and Benitah, S. A. (2013) “Regenerating the skin: a task for the heterogeneous stem cell pool and surrounding niche,” *Nature reviews. Molecular cell biology*, 14(11), pp. 737–48. doi: 10.1038/nrm3675.
- Sorrell, J. M. and Caplan, A. I. (2004) “Fibroblast heterogeneity: More than skin deep,” *Journal of Cell Science*, 117(5), pp. 667–675. doi: 10.1242/jcs.01005.
- Sos, B. C. *et al.* (2016) “Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay,” *Genome Biology*, 17(1), p. 20. doi: 10.1186/s13059-016-0882-7.
- South, A. P. *et al.* (2014) “NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis,” *Journal of Investigative Dermatology*, 134(10), pp. 2630–2638. doi: 10.1038/jid.2014.154.
- de Souza, C. P. E. *et al.* (2020) “Epiclomal: Probabilistic clustering of sparse single-cell DNA methylation data,” *PLoS Computational Biology*, 16(9), p. e1008270. doi: 10.1371/journal.pcbi.1008270.

- Sriram, G., Bigliardi, P. L. and Bigliardi-Qi, M. (2015) "Fibroblast heterogeneity and its implications for engineering organotypic skin models in vitro," *European Journal of Cell Biology*, 94(11), pp. 483–512. doi: 10.1016/j.ejcb.2015.08.001.
- Stang, A., Khil, L. and Kaj, H. (2019) "Incidence and mortality for cutaneous squamous cell carcinoma : comparison across three continents," *Journal of the European Academy of Dermatology and Venereology*, 33(S8), pp. 6–10. doi: 10.1111/jdv.15967.
- Stewart, C. A. *et al.* (2020) "Single-cell analyses reveal increased intratumoral heterogeneity after the onset of therapy resistance in small-cell lung cancer," *Nature Cancer*, 1(4), pp. 423–436. doi: 10.1038/s43018-019-0020-z.
- Stortchevoi, A., Kamelamela, N. and Levine, S. S. (2020) "SPRI beads-based size selection in the range of 2-10kb," *Journal of Biomolecular Techniques*, 31(1), pp. 7–10. doi: 10.7171/jbt.20-3101-002.
- Stuart, T. *et al.* (2019) "Comprehensive Integration of Single-Cell Data," *Cell*, 177(7), pp. 1888-1902.e21. doi: 10.1016/j.cell.2019.05.031.
- Stunova, A. and Vistejnova, L. (2018) "Dermal fibroblasts—A heterogeneous population with regulatory function in wound healing," *Cytokine and Growth Factor Reviews*, 39(February), pp. 137–150. doi: 10.1016/j.cytogfr.2018.01.003.
- Tabib, T. *et al.* (2018) "SFRP2/DPP4 and FMO1/LSP1 Define Major Fibroblast Populations in Human Skin," *Journal of Investigative Dermatology*, 138(4), pp. 802–810. doi: 10.1016/j.jid.2017.09.045.
- Tan, S. T. *et al.* (2018) "Basal cell carcinoma arises from interfollicular layer of epidermis," *Journal of Oncology*, 2018. doi: 10.1155/2018/3098940.
- Teschendorff, A. E. (2020) "A comparison of epigenetic mitotic-like clocks for cancer risk prediction," *Genome Medicine*, 12(1), p. 56. doi: 10.1186/s13073-020-00752-3.
- Teschendorff, A. E. *et al.* (2020) "EPISCORE: Cell type deconvolution of bulk tissue DNA methylomes from single-cell RNA-Seq data," *Genome Biology*, 21(1), p. 1. doi: 10.1186/s13059-020-02126-9.
- Thomson, J. *et al.* (2021) "The Genomic Landscape of Actinic Keratosis," *Journal of Investigative Dermatology*, 141(7), pp. 1664–1674.e7. doi: 10.1016/j.jid.2020.12.024.
- Tigges, J. *et al.* (2014) "The hallmarks of fibroblast ageing," *Mechanisms of Ageing and Development*, 138(1), pp. 26–44. doi: 10.1016/j.mad.2014.03.004.
- Toll, A. *et al.* (2016) "MiR-204 silencing in intraepithelial to invasive cutaneous squamous cell carcinoma progression," *Molecular Cancer*, 15(1), p. 1. doi: 10.1186/s12943-016-0537-z.
- Tsuboi, R. *et al.* (2020) "Autologous cell-based therapy for male and female pattern hair loss using dermal sheath cup cells: A randomized placebo-controlled double-blinded dose-finding clinical study," *Journal of the American Academy of Dermatology*, 83(1), pp. 109–116. doi: 10.1016/j.jaad.2020.02.033.
- Vandiver, A. R. *et al.* (2015) "Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin," *Genome Biology*, 16(1). doi: 10.1186/s13059-015-0644-y.
- Veit, G. *et al.* (2011) "Collagen XXIII, novel ligand for integrin  $\alpha 2\beta 1$  in the epidermis," *Journal of Biological Chemistry*, 286(31), pp. 27804–27813. doi: 10.1074/jbc.M111.220046.
- Vento-Tormo, R. *et al.* (2018) "Single-cell reconstruction of the early maternal–fetal interface in humans," *Nature*, 563(7731), pp. 347–353. doi: 10.1038/s41586-018-0698-6.
- Visvader, J. E. (2011) "Cells of origin in cancer," *Nature*, 469(7330), pp. 314–322. doi: 10.1038/nature09781.

## 8. References

---

- Vorstandlechner, V. *et al.* (2020) “Deciphering the functional heterogeneity of skin fibroblasts using single-cell RNA sequencing,” *FASEB Journal*, 34(3), pp. 3677–3692. doi: 10.1096/fj.201902001RR.
- Waldera Lupa, D. M. *et al.* (2015) “Characterization of Skin Aging-Associated Secreted Proteins (SAASP) Produced by Dermal Fibroblasts Isolated from Intrinsically Aged Human Skin,” *Journal of Investigative Dermatology*, 135(8), pp. 1954–1968. doi: 10.1038/jid.2015.120.
- Wang, Q. *et al.* (2017) “Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment,” *Cancer Cell*, 32(1), pp. 42–56.e6. doi: 10.1016/j.ccell.2017.06.003.
- Wang, S. *et al.* (2020) “Single cell transcriptomics of human epidermis identifies basal stem cell transition states,” *Nature Communications*, 11(1), pp. 1–14. doi: 10.1038/s41467-020-18075-7.
- Wang, T. *et al.* (2020) “Dysfunctional epigenetic aging of the normal colon and colorectal cancer risk,” *Clinical Epigenetics*, 12(1), p. 5. doi: 10.1186/s13148-019-0801-3.
- Wang, W. *et al.* (2012) “Collagen XXIV (Col24a1) promotes osteoblastic differentiation and mineralization through TGF- $\beta$ /smads signaling pathway,” *International Journal of Biological Sciences*, 8(10), pp. 1310–1322. doi: 10.7150/ijbs.5136.
- Wehner, M. R. *et al.* (2012) “Indoor tanning and non-melanoma skin cancer: Systematic review and meta-analysis,” *BMJ (Online)*, 345(7877). doi: 10.1136/bmj.e5909.
- Wei, L. *et al.* (2021) “Ultradeep sequencing differentiates patterns of skin clonal mutations associated with sun-exposure status and skin cancer burden,” *Science Advances*, 7(1), p. eabd7703. doi: 10.1126/sciadv.abd7703.
- White, A. C. *et al.* (2011) “Defining the origins of Ras/p53-mediated squamous cell carcinoma,” *Proceedings of the National Academy of Sciences of the United States of America*, 108(18), pp. 7425–7430. doi: 10.1073/pnas.1012670108.
- Wiench, M. *et al.* (2011) “DNA methylation status predicts cell type-specific enhancer activity,” *EMBO Journal*, 30(15), pp. 3028–3039. doi: 10.1038/emboj.2011.210.
- Wierzbinska, J. A. *et al.* (2020) “Methylome-based cell-of-origin modeling (Methyl-COOM) identifies aberrant expression of immune regulatory molecules in CLL,” *Genome Medicine*, 12(1), p. 29. doi: 10.1186/s13073-020-00724-7.
- Wiklund, E. D. *et al.* (2011) “Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer,” *International Journal of Cancer*, 128(6), pp. 1327–1334. doi: 10.1002/ijc.25461.
- Wikramanayake, T. C., Stojadinovic, O. and Tomic-Canic, M. (2014) “Epidermal Differentiation in Barrier Maintenance and Wound Healing,” *Advances in Wound Care*, 3(3), pp. 272–280. doi: 10.1089/wound.2013.0503.
- Wlaschek, M. *et al.* (2021) “Connective Tissue and Fibroblast Senescence in Skin Aging,” *Journal of Investigative Dermatology*, 141(4), pp. 985–992. doi: 10.1016/j.jid.2020.11.010.
- Wotherspoon, D., Rogerson, C. and O’shaughnessy, R. F. L. (2020) “Perspective: Controlling epidermal terminal differentiation with transcriptional bursting and rna bodies,” *Journal of Developmental Biology*, 8(4), pp. 1–17. doi: 10.3390/jdb8040029.
- Wysocka, J., Reilly, P. T. and Herr, W. (2001) “Loss of HCF-1–Chromatin Association Precedes Temperature-Induced Growth Arrest of tsBN67 Cells,” *Molecular and Cellular Biology*, 21(11), pp. 3820–3829. doi: 10.1128/mcb.21.11.3820-3829.2001.

- Xi, Y. and Li, W. (2009) "BSMAP: Whole genome bisulfite sequence MAPping program," *BMC Bioinformatics*, 10(1), pp. 1–9. doi: 10.1186/1471-2105-10-232.
- Yang, Z. *et al.* (2016) "Correlation of an epigenetic mitotic clock with cancer risk," *Genome Biology*, 17(1), p. 205. doi: 10.1186/s13059-016-1064-3.
- Yooyongsatit, S. *et al.* (2015) "Patterns and functional roles of LINE-1 and Alu methylation in the keratinocyte from patients with psoriasis vulgaris," *Journal of Human Genetics*, 60(7), pp. 349–355. doi: 10.1038/jhg.2015.33.
- Youssef, K. K. *et al.* (2010) "Identification of the cell lineage at the origin of basal cell carcinoma," *Nature Cell Biology*, 12(3), pp. 299–305. doi: 10.1038/ncb2031.
- Youssef, K. K. *et al.* (2012) "Adult interfollicular tumour-initiating cells are reprogrammed into an embryonic hair follicle progenitor-like fate during basal cell carcinoma initiation," *Nature Cell Biology*, 14(12), pp. 1282–1294. doi: 10.1038/ncb2628.
- Yu, S. L. and Lee, S. K. (2017) "Ultraviolet radiation: DNA damage, repair, and human disorders," *Molecular and Cellular Toxicology*, 13(1), pp. 21–28. doi: 10.1007/s13273-017-0002-0.
- Zackheim, H. S. (1963) "Origin of the human basal cell epithelioma.," *The Journal of investigative dermatology*, 40(6), pp. 283–297. doi: 10.1038/jid.1963.152.
- Zhang, X. *et al.* (2016) "DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells," *Nature Genetics*, 48(9), pp. 1014–1023. doi: 10.1038/ng.3610.
- Zheng, S. C., Widschwendter, M. and Teschendorff, A. E. (2016) "Epigenetic drift, epigenetic clocks and cancer risk," *Epigenomics*, 8(5), pp. 705–719. doi: 10.2217/epi-2015-0017.
- Zhu, S. *et al.* (2017) "Advances in single-cell RNA sequencing and its applications in cancer research," *Oncotarget*, 8(32), pp. 53763–53779. doi: 10.18632/oncotarget.17893.
- Ziegenhain, C. *et al.* (2017) "Comparative Analysis of Single-Cell RNA Sequencing Methods," *Molecular Cell*, 65, pp. 631-643.e4. doi: 10.1016/j.molcel.2017.01.023.
- Zouboulis, C. C. (2009) "The skin as an endocrine organ," *Dermato-Endocrinology*, 1(5), pp. 250–252. doi: 10.4161/derm.1.5.9499.



## 9. Acknowledgments

First and foremost, I would like to thank Prof. Dr. Frank Lyko for the opportunity to pursue this project and for his continuous support throughout these last four years. It was a great pleasure to be part of his team and I am deeply grateful for the scientific growth I accomplished thanks to his mentorship. I also have to thank Dr. Manuel Rodríguez-Paredes, as this work would not have been possible without his scientific and personal contributions. Not only did he provide priceless guidance during my time as a Ph.D. student, but he made the journey so much more joyful.

Over the past four years, I have been supported by many members of the A130 lab, which unquestionably helped this thesis to move forward successfully. On the bioinformatics site, I would like to thank Dr. Felix Bormann and Dr. Julian Gutekunst for their support with the DNA methylation analyses and for always having the door open to my endless questions, even after closing hours. Also, I would like to thank Dr. Günter Raddatz and Oliver Gilliam for their bioinformatics support and specially for their instrumental role in the single-cell analyses. On the wet lab site, I must thank Sabrina Schütz for her great help with the mRNA-FISH and IF assays and Tanja Musch for her valuable assistance in many of the lab experiments. Finally, I would like to extend my gratitude to Cansu Cirzi, Dr. Francesca Tuorto, Dr. Vitor Coutinho, and Dr. Florian Köhler, as well as to all current and former members of the A130 lab for the stimulating scientific discussions and their theoretical input.

Furthermore, the crucial contributions of the scOpenLab members have benefited this work extensively. I am especially grateful to Dr. Michelle S. Liberio as without her expertise the sci-MET project would not have succeeded. I am also thankful to her for sharing the long days and frustration with me and never giving up. I would also like to thank Dr. Jan-Philipp Mallm and Katharina Bauer for the countless discussions and helpful suggestions regarding the sci-MET method optimization. In addition, I would like to thank the members of the High-Throughput Core Facility for their support with the sci-MET sequencing protocol and the Genomics and Proteomics Core Facility members for their help with the DNA methylation experiments. I am also grateful for the important technical support I received from Dr. Daniel Hasche.

I would like to extend my sincere thanks to my Thesis Advisory Committee members, Prof. Dr. Ana Martin-Villalba and Dr. med. Anke S. Lonsdorf for their scientific input and guidance, as they undoubtedly helped to shape this thesis. Lastly, I would like to thank the additional members of my examination committee Prof. Dr. Michaela Frye and Dr. Michael Milsom, for their critical review of this work.

On a personal level, I am incredibly thankful for the people I shared this experience with. I could not have asked for a better partner in crime during my time at the DKFZ than Cansu, who I would like to thank for always being there for me. I also want to thank Sina and Ranja for always bringing happiness and joy in and out of the lab. I cannot forget the skin team, Manolo and Sabrina, to who I am genuinely thankful for all the laughs and fun we had together (and will continue to have). I would like to thank Malin, Laura, Manuel, Francesca, Felix, and Flo for the times we shared in the office where it never is cold. Lastly, I would like to thank all current and previous members of the A130 lab for creating such a nice environment and for all the trips, celebrations, and little moments we shared.

Lastly, I would like to thank my family and friends for their support, patience, and care. Thank you, Benedikt, for being my biggest support throughout this process and helping me navigate the stressful moments. Thank you, Miguel, Aida, and Anisa, for the unconditional support. Thanks to Heike, Rachel, Leo, and Alessandro, I am truly happy that destiny made us share that initial week and that I can count you as my friends.

*Gràcies a la família i als amics de sempre, per fer la distància més fàcil i per no deixar d'estar mai al meu costat.*