Decoding epidermis reconstruction through time-course multi-omics data integration

Bou Samra, Elias[1*], Tacheau, Charlotte[1], Scott-Boyer, Marie-Pier[2], Bodein, Antoine[2], Foucher, Aude[1], Abed, Kahina[1], Segaud, Valery[3], Thillou, Fabienne[3], Boyera, Nathalie[3], Misra, Namita[1], Droit, Arnaud[2], Perin, Olivier[1], **Cavusoglu, Nükhet[1]**

[1] L'Oréal Research and Innovation, 1 avenue E. Schueller, 93600, Aulnay-sous-Bois, France.
[2] Molecular Medicine Department, CHU de Québec Research Center, Université Laval, Québec, QC, Canada.
[3] EPISKIN SA, 4, rue Alexander Fleming, 69366 Lyon, France.

*Nükhet Cavusoglu, L'Oréal Research and Innovation, 1 avenue E. Schueller, 93600, Aulnay-sous-Bois, France, nukhet.cavusoglu@rd.loreal.com

Abstract

The development of epidermis and its homeostasis require tightly regulated molecular processes governed by cooperation and interaction of molecules, such as transcripts, proteins, and metabolites. The complexity of these regulations makes clear that studying the system at a single molecular level gives only limited information, hence integrating and inferring regulatory interactions from these data are crucial.

Methods: To investigate the potential of multi-omics integration to identify the sequential events leading to epidermis differentiation, we generated concurrent high-throughput RNA-seq, and quantitative proteomics and metabolomics from a commercial reconstructed human epidermis model (SkinEthic Episkin Co, Lyon, France). In this regard, datasets were first collected at eleven timepoints over twenty-eight days of culture, integrated and clustered. Then, a multi-layer network-based analysis using both data-driven and knowledge-driven building methods was used to map the regulatory interactions between clustered molecules.

Results: Cluster analysis identified eight distinct subgroups of molecules which showed distinctive temporal patterns characterized by their initial levels of abundance and variation over the time-course. Functional enrichment analyses allowed to capture precise biological events occurring during keratinocyte differentiation process. In addition, multi-layered network approach enabled to identify sub-networks of pathways associated with keratinocyte differentiation and expanded the sub-networks to additional markers that have evidence of interactive or regulatory mechanisms relevant for sub-networks characterization.

Conclusions: By integrating longitudinal multi-omics datasets, we were able to detect temporal relationships and interactions between biomolecules which will provide us with a

comprehensive list of known and previously unrecognized components of the epidermal reconstruction process.

Key words: multi-omics data, integrative approach, networks, epidermis reconstruction

Introduction

The development of epidermis and its homeostasis require tightly regulated molecular processes that involves epidermal keratinocyte proliferation, differentiation, and apoptosis to name a few of them. Such processes are governed by a dynamic cooperation and interaction of biomolecules, namely transcripts, proteins, and metabolites, and each of these biomolecules embed essential information about the regulatory programs orchestrating skin's features.

To study epidermis development, researchers are increasingly using 3D cell cultures. Three-dimensional skin models offer the potential to study various skin disorders and evaluate the efficacy and safety of active compounds *in vitro*. Previous work investigating the longitudinal patterns of calcium-induced differentiation of primary human keratinocytes in 2D monolayer cells or 3D reconstituted human epidermis (RHE) models have mainly focused on studying this process at a single-omics layer, e.g., transcriptomics or proteomics [1-4]. These studies allowed to detect biomolecules of one type, and thus captured changes only for a subset of the components of a particular pathway. More recently, multi-omics datasets were generated and used to study molecular mechanisms, for instance the unfolded protein response signaling pathway by which human keratinocytes undergo to terminal differentiation [5]. While different omics layers existed, the datasets were analyzed separately, and the conclusions were compared and combined. Thus, the capture of the complementary effects and synergistic interactions between the different omics layers and biomolecules has lacked. To address this issue, we have recently developed an analytical framework for the integration of longitudinal multi-omics datasets that relies on multi-omics kinetic clustering approach [6] and multi-layers network-based analysis using both data-driven and knowledge-driven methods [7]. In this study, we tailored these approaches to comprehensively profile in a holistic view the landscape of multi-layer biomolecules and expression dynamics across a densely sampled time course of epidermal reconstruction. We were able to facilitate the interpretation of multi-layers systems and to detect within temporal relationships and interactions between molecules. To our knowledge, this is the first longitudinal multi-omics integrative study to provide a comprehensive view of known and previously unrecognized components of the epidermal reconstitution process.

Materials & Methods

SkinEthic^{TM} Reconstructed Human Epidermis culture

The reconstructed human epidermis (RHE) equivalent cultures were prepared at EPISKIN Laboratories (Lyon, France) following their standard protocol. Human neonatal keratinocytes from four individual donors were previously cultivated on plastic in monolayer culture with a chemically defined medium to prepare banks. Keratinocytes were then seeded onto polycarbonate filter inserts. Two hours after seeding, cultures were raised to the air-liquid interface in a chemically defined medium. The air-liquid culture continued for a total of 28 days (time 1, 3, 5, 7, 10, 12, 14, 17, 21, 25 and 28 days) to produce the fully differentiated epidermal equivalents.

Hematoxylin and eosin (HE) analysis

For all conditions, tissues were fixed in a formalin solution (4%W/V) before to be dehydrated and embedded in paraffin. Paraffin blocks were cut in 5µm thickness sections using the microtome for histology. After mounting the cuts on a blade, the sections were stained with hemalun-eosin (HE) on a Sakura robot and were scanned using the Nanozoomer (NDP scan software 3.3) for imaging.

Total RNA extraction

Tissue samples were homogenized in QIAzol Lysis Reagent (Qiagen, Germantown, MD, USA) using a rotor-stator homogenizer. Total RNA was extracted using the miRNeasy micro kit on-column DNase (Qiagen, Hilden, DE) treatment following the manufacturer's instructions. Quantity of RNA was measured using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and total RNA quality was assayed on an Agilent BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA).

Total RNA-Sequencing

The Illumina TruSeq stranded Total RNA library prep kit with Ribo-Zero Gold (Illumina Inc., San Diego, CA, USA) was used to prepare RNA sequencing libraries, according to manufacturer's instruction. Briefly, 1 ug of total RNA was used. Ribosomal RNA (rRNA) including both cytoplasmic and mitochondrial rRNA was removed using biotinylated, target-specific oligos combined with Ribo-Zero rRNA removal beads. Following purification with Agencourt AMPure XP beads (Beckman Coutler, Missisauga, Ontario, Canada), the RNA was

fragmented using divalent cations under elevated temperature. The fragmented RNA was used as a template for cDNA synthesis by reverse transcriptase with random primers. Strand specificity was achieved by replacing dTTP with dUTP. This cDNA was further converted into double stranded DNA that was end-repaired to incorporate the specific index adaptor for multiplexing. Followed a purification, an amplification for 15 cycles was performed using a polymerase unable to incorporate past dUTP. Therefore, the second strand is quenched during amplification. The quality of final amplified librairies were examined with a DNA screentape D1000 on a TapeStation 2200 and the quantification was done on the QBit 3.0 fluorometer (ThermoFisher Scientific, Canada). Subsequently, RNA-seq librairies with unique index were pooled together in equimolar ratio and the pool was sequenced on a high ouput flowcell on an HiSeq 2500 system at the Next-Generation Sequencing Plateform, Genomics Center, CHU de Québec-Université Laval Research Center, Québec City, Canada for paired-end 125 pb sequencing. After sequencing, raw data were obtained in the fastq format. The software FastQC [8] was used for validating the quality of the data and Trimmomatic [9] was used for trimming of the adapter content and overrepresented sequences. The mRNAs were aligned to the EnsDb.Hsapiens.v86 transcriptome with the Kallisto tool [10]. The R-package tximport [11] was used to convert the transcript quantification to gene quantification. Only protein-coding genes with a minimum of five read detected across all samples were kept for further analysis. Final normalization was performed using the Relative Log Expression (RLE) method [12, 13] and data were log2 transformed.

Protein extraction and LC-MS analysis

Briefly, RHE were grinded via inox beads beating (3min, 30 Hz) in presence of DOC buffer (0,5% sodium deoxycholate, 50 mM ammonium bicarbonate, 50 mM DTT, 1µM Pepstatin, and cocktail of inhibitors). Supernatant was filter (0.45µm), then precipitated with 4 vol of ice-cold acetone. After resuspension in 1% sodium deoxycholate/50mM ammonium bicarbonate, proteins were quantified using the Bradford method. Finally, 20 µg were subjected to denaturation, reduction, alkylation, and digestion via Trypsin. Peptides were marked via isobaric labelling using TMT10-plex and samples were pooled according to the label. For alignment between injection, a reference pool containing all samples was produced and labelled with TMT10. Labeled pooled samples were fractionated by reverse phase-HPLC at High pH into 14 fractions. Each fraction was separated on Dionex Ultimate 3000 nano RSLC (15µx500mm C18 3µ 100A) using an acetonitrile gradient (5 to 40% acetonitrile in 90min) before injection via electrospray (positive mode, 2100V) into the mass spectrometer

OrbitrapTM FusionTM (ThermoFisher Scientific, Canada) for MS3 data acquisition. Precursor data acquisition was performed in 350 to 1500 m/z range, followed by CID fragmentation in ion trap and isolation and analysis of 10 MS2 fragments that are then further fragmented via HCD fragmentation for detection in orbitrap (MS3). Bioinformatic analysis was performed using ProteomeDiscoverer 2.1 against the Human Uniprot database.

Metabolite extraction and LC-MS analysis

Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Samples are extracted with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation to recover chemically diverse metabolites. The resulting extract is divided for the following analysis: two separate reverse phases (RP)/UPLC-MS/MS methods using positive ion mode electrospray ionization (ESI), one RP/UPLC-MS/MS using negative ion mode ESI, one HILIC/UPLC-MS/MS using negative ion mode ESI.

All methods utilize a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a ThermoScientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract is dried then reconstituted in solvents compatible to each of the four methods. One aliquot is analyzed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract is gradient-eluted from a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 µm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). A second aliquot is also analyzed using acidic positive ion conditions but is chromatographically optimized for more hydrophobic compounds. A third aliquot is analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts are gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot is analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 µm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternates between MS and data-dependent MSn scans using dynamic exclusion. The scan range varies slightly between methods, but covers approximately 70-1000 m/z.

TimeOmics approach and clustering

The timeOmics approach was used to model and cluster longitudinal multiomics molecules with similar expression profiles over time. Briefly, to focus on time-varying features, filter on time profiles was applied to keep only molecules with the highest expression fold change between the lowest and highest point over the entire time course, as described in [14]. Every retained molecule was modelized over the time-course by considering the inter-individual variation using the Linear Mixed Model Spline framework [15]. Then the modelized expression profiles were clustered in groups of similar expressions over time. This was performed using various multivariate projection-based methods implemented in mixOmics [16].

Functional Annotation Clustering and Pathway Analysis

Over representation analysis (ORA) helps to find enriched and meaningful biological insights from interacting biomolecules. This task was achieved using gProfiler2 [17], on each mRNA and protein kinetic clusters. We focused on the three Gene Ontology (GO) terms: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC), and the two biological pathways: KEGG and Reactome. P-values were corrected with gProfiler2 custom multiple testing (Benjamini-Hochberg FDR) correction algorithm [18] and only significant terms were considered (FDR < 0.05).

Multi-layered network reconstruction (NetOmics)

The first layer to be reconstructed was the gene inference network from the mRNA. We used the ARACNe algorithm to build a network by kinetic cluster but also to build an entire network composed of all the mRNA. The second layer was the PPI network. Proteins were connected to each other using the BioGRID interaction database. As for the genes, PPI sub-networks were built for each kinetic cluster, but we also built an entire PPI network composed of all the proteins. These first two layers were combined thanks to two types of links. First, protein-coding information linked genes to their corresponding proteins. Second, TF-regulated information from TF2DNA and TTRUST databases [19, 20] linked proteins to genes. In addition, KEGG pathway database was also used to link protein enzymes to metabolite reactions.

Results

Longitudinal multi-omics datasets collection

To characterize the sequential events leading to epidermis differentiation at a multi-molecular level, we performed concurrent longitudinal high-throughput transcriptomics, quantitative proteomics and metabolomics in a commercial reconstructed human epidermis (RHE) model (SkinEthic Episkin Co, Lyon, France). In this regard, human neonatal keratinocytes from four individual donors were seeded onto polycarbonate filter inserts and followed over a 28-day time course reconstruction. Multi-omics datasets were collected at different timepoints as shown in Figure 1A. The time course of epidermis formation and maturation as shown using H&E staining revealed a fully stratified epithelium with various layers readily observable, including the basal, spinous, and granular layers as early as 5 days in culture. The stratum corneum initially appears thin and compact but matures over time (Figure 1B). Principal-component analysis (PCA) on RNA-Seq data showed high consistency between biological replicates and strong time-dependence for global gene expression patterns, in particular days 1 to 7 (Figure 1C).
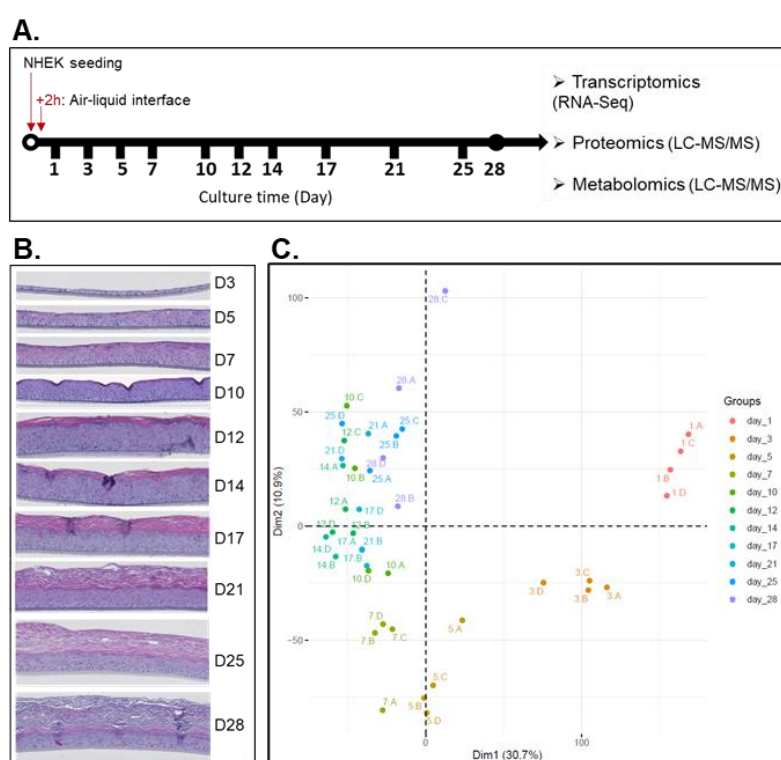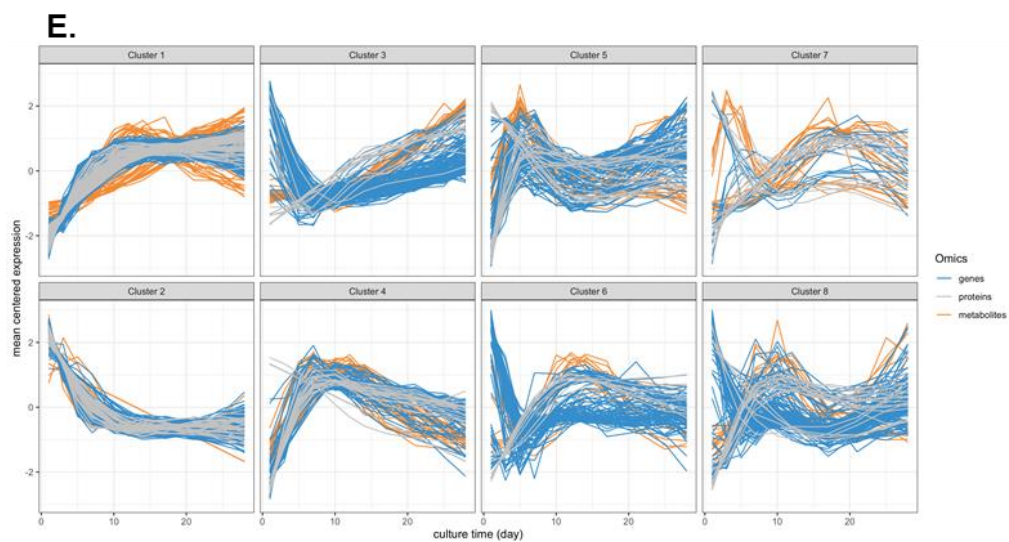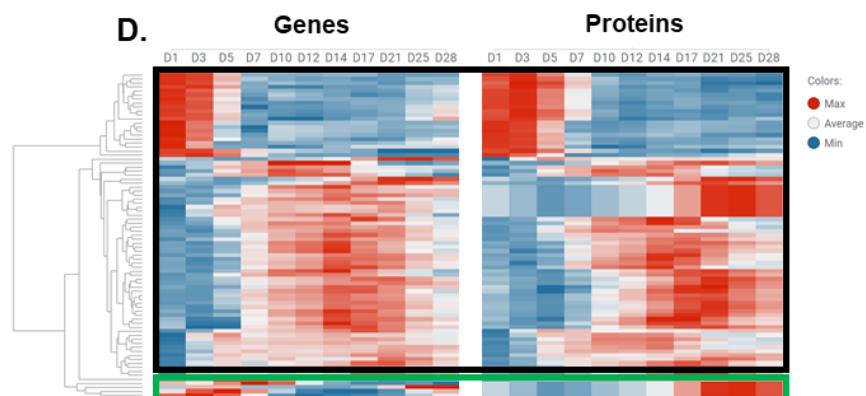

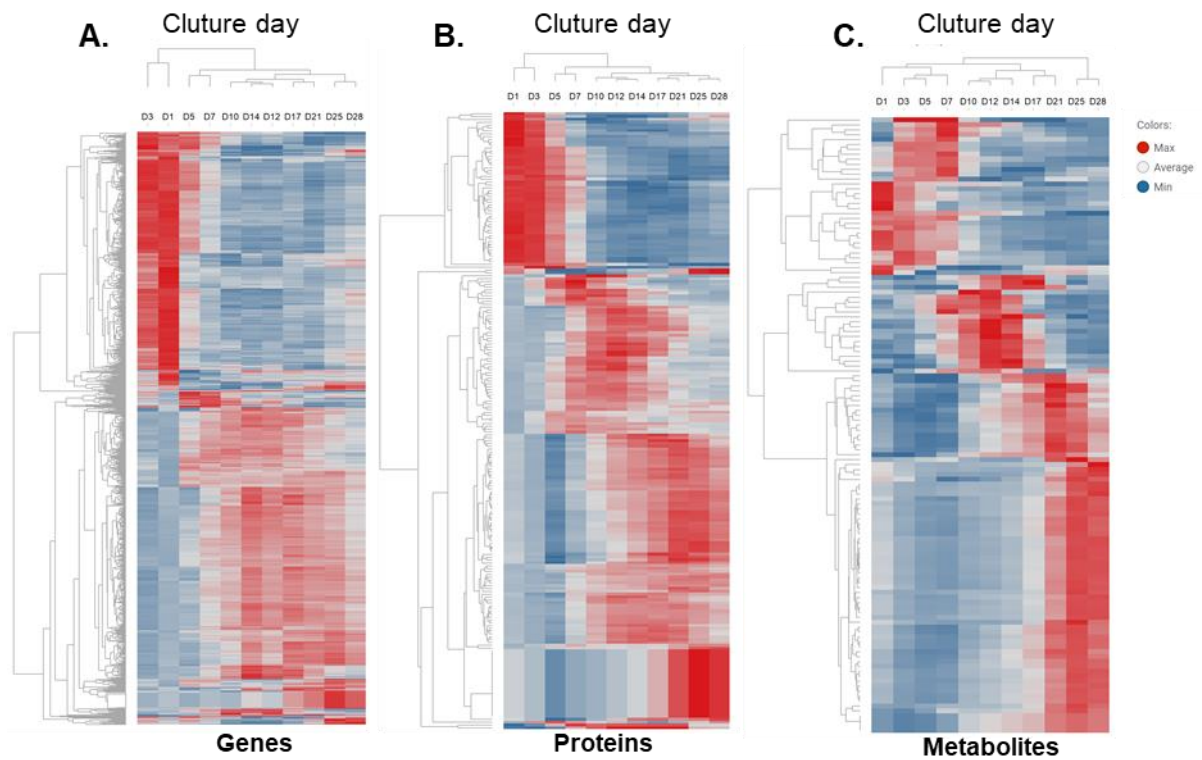**Figure 1. A.** Schematic of multi-omics data collected across the epidermal reconstruction time-course. **B.** Time course for maturation of the epidermis. Cultures collected at various time points (3–28 days) and analyzed by H&E. Representative cultures for each time point are shown. **C.** PCA of RNA-seq data highlight time as the primary axis of variation. A, B, C and D refers to the four individual donors.

Datasets modelling and clustering (timeOmics)

We next cataloged the dynamic changes occurring during epidermal reconstitution at the 3-omics level. To do so, the timeOmics approach was used as previously described [6] to cluster multi-omics molecules with similar expression profiles over time. Briefly, longitudinal datasets were first pre-processed and normalized to remove biomolecules with missing values and/or highly variable over time. Following data pre-processing, a 1.5- or 2-fold expression change based cut-off and an FDR < 5% on time points were used to remove proteins and metabolites, and genes, respectively which are not impacted by time effect. We thus identified 1012 genes, 222 proteins and 226 metabolites with a dynamic expression across the entire time course following calcium-induced keratinocytes differentiation. Heatmaps with their patterns of expression are shown in Figures 2A-C. Hierarchical clustering of time points during epidermal reconstitution revealed strong time-dependence for global expression patterns. For instance, gene and protein expression profiles were distinctly similar between days 1 and 3 compared with other time points. A similar pattern was observed for cultures at days 5 and 7, and the subsequent patterns of gene expression from days 10–28 were similar between them and distinct from the earlier time points. Within this last cluster, we distinguished sub-clustering of days 10-14, 17-21 and 25-28, corresponding simultaneously to optimum differentiation of RHE model (days 10-17) and it's degradation as the stratum corneum continues to thicken (days 21–28). Metabolite expression profiles on the other hand, were quite similar as those of genes and proteins except for day 1 which clustered separately from the later times points, and for days 5 and 7 who shared a similar pattern with day 10 (Figures 2A-C). Furthermore, we investigated expression profiles of biomolecules detected on both gene and protein levels. A total of 81 mRNA-protein pairs were thus analyzed whose expression ratios are illustrated in Figure 2D. A clustering analysis clearly identified patterns of temporal delay between mRNA and protein expression, supporting the fact that the relation between gene and protein expression changes is not linear, including patterns of gene and protein expression changing either in the same direction (black box) or in opposite directions (green box) (Figure 2D).

We then modelled the expression of every biomolecule including the variations of the four replicates for each of the eleven timepoints. This revealed 8 distinct subgroups of clusters characterized by their initial levels of abundance and variation during keratinocyte differentiation, with either a general increase or a general decrease in abundance across the time course (Figure 2E). The distribution of biomolecules among the different clusters is given in Table 1. To functionally annotate the subgroups of biomolecules, we employed a cluster-wide

enrichment analysis on genes and proteins using the gProfiler2 tool [17] (Tables 2-3 and data not shown). Cluster 1 comprised a total of 327 mRNAs, 62 proteins and 56 metabolites. Since it comprised the most of biomolecules (a total of 445 molecules out of 1460 (~30%), highlighted in grey in Table 1), we considered this cluster as an example for upcoming analyses to illustrate the obtained results. Biomolecule expression increased from the progenitor state (day 1) through days 7-10 and reached a plateau at later timepoints (Figure 2E). The most enriched genes in this cluster showed significant distribution among skin development, including epithelial differentiation, cornification, peptide cross-linking, lipid metabolism, and ceramide biosynthesis (Figure 2F and Table 2). Genes from the epidermal differentiation complex (EDC) such as filaggrin (FLG), loricrin (LOR), cornulin (CRNN), small proline rich protein (SPRRs) and late cornified envelope (LCEs), genes from the lipid biosynthesis metabolism such as ceramide synthases (CERS3 and CERS4) and lipoxygenases (ALOX12B and ALOXE3), and genes participating in the cornified envelope formation such as transglutaminases (TGM3 and TGM5) were all highly represented within this cluster (Table 2, in bold). In addition, we found significant enrichment for "regulation of protein tyrosine kinase activity" and "antimicrobial signaling". Focusing on specific functions of the modulated proteins of cluster 1, many of them were related to epidermis development, peptide cross-linking, immune function, export, secretion signaling, and small molecule metabolism (Figure 2G and Table 3). Interestingly, several proteins from the EDC were identical to those observed on the gene level, in particular FLG, LOR, and SPRR2E. However, other members from the EDC complex as the S100 calcium binding proteins S100A8, A9 and S100P were exclusively identified for proteins. Additionally, other molecules required to promote keratinocyte differentiation such as the non-apoptotic caspase CASP14 and the Homeodomain-only protein HOPX [21] were found to be expressed on both levels. Altogether, these findings are consistent with the formation of the skin barrier and other defenses of the epithelium, and thus are associated with keratinocyte late differentiation state. Moreover, they confirm that the consecutive integration of transcriptomics and proteomics data improves characterization of mechanisms related to epidermal reconstruction.

A. Cluture day — Genes

B. Cluture day — Proteins

C. Cluture day — Metabolites

Colors:
Max
Average
Min

D. Genes   Proteins

Colors:
Max
Average
Min

E.

Cluster 1 Cluster 3 Cluster 5 Cluster 7
Cluster 2 Cluster 4 Cluster 6 Cluster 8

mean centered expression

culture time (day)

Omics
genes
proteins
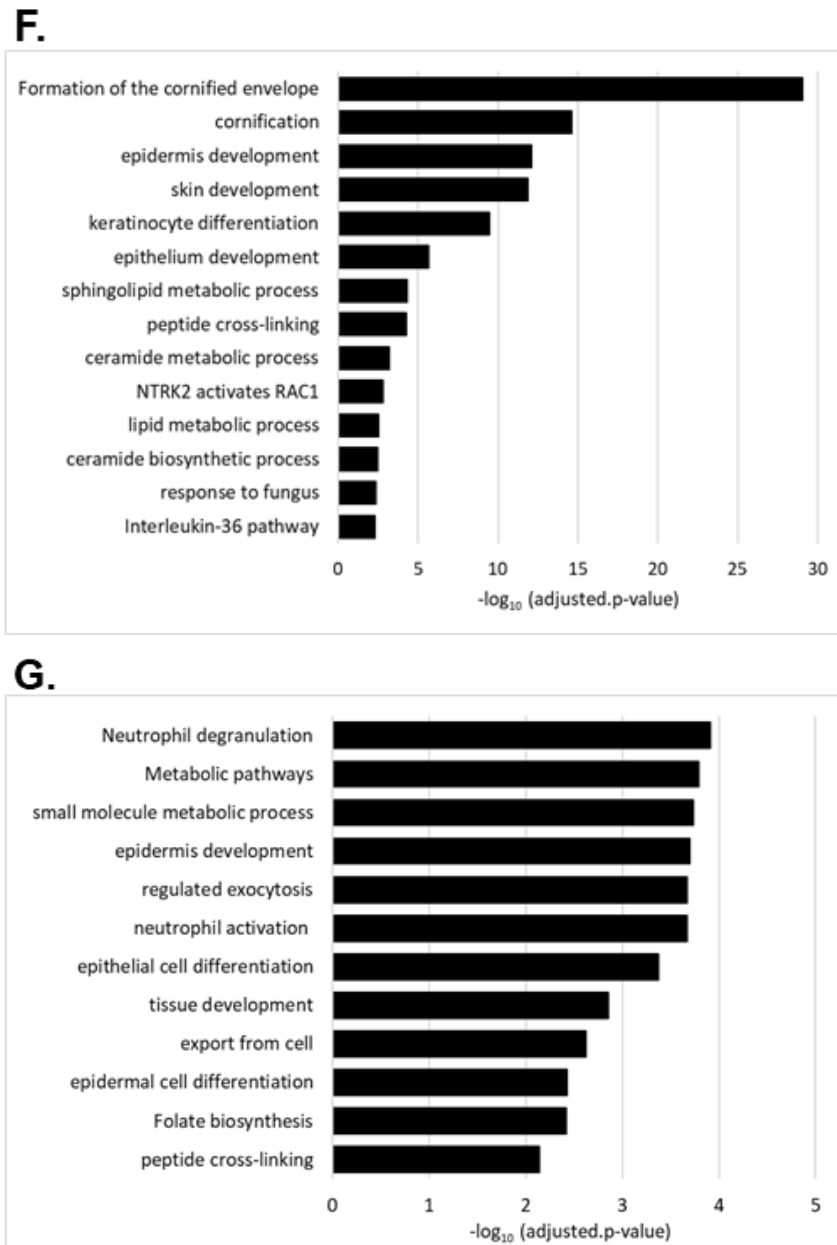metabolites

F.

G.

Figure 2. Hierarchical clustering of **(A)** gene, **(B)** protein and **(C)** metabolite expression profiles from RHE model demonstrates time dependence of expression changes. **D.** Heatmap on gene-protein pairs shows temporal delay between mRNA and protein expression. **E.** Clusters of gene, protein, and metabolite expression dynamics along differentiation. **F-G.** Top enriched pathways among genes **(F)** and proteins **(G)** of cluster 1.

	Genes	Proteins	Metabolites	Total/cluster
Cluster 1	327	62	56	445
Cluster 2	144	33	20	197
Cluster 3	159	38	49	246
Cluster 4	81	21	29	131
Cluster 5	87	23	23	133
Cluster 6	83	9	11	103
Cluster 7	16	16	21	53
Cluster 8	115	20	17	152

Table 1. Distribution of genes, proteins and metabolites among the 8 formed clusters.

term_id	term_name	-log10(p-val)	Genes
REAC:R-HSA-6809371	Formation of the cornified envelope	29,07	**TGM5**,CASP14,KRT23,DSG1,DSC1,**FLG**,**LCE2B**,**SPRR2G**,**LCE3D**,KRT80,KRT1,KRT78,KRT72,**LCE1D**,KRT2,LIPM,**LCE3E**,KRT73,**LCE5A**,**LCE1E**,**LCE1A**,**LCE4A**,**LCE2A**,**LCE2C**,**LCE2D**,KRT77,**LCE1B**,**LCE1C**,**LOR**,**SPRR2E**,LIPN,LIPK,**LCE6A**,LCE1F
GO:0070268	cornification	14,61	**TGM5**,CASP14,KRT23,DSG1,DSC1,**FLG**,**CERS3**,**SPRR2G**,**LCE3D**,KRT80,KRT1,KRT78,KRT72,KRT2,LIPM,KRT73,**LCE1A**,KRT77,**LOR**,**SPRR2E**,LIPN,LIPK
GO:0008544	epidermis development	12,09	USH2A,STS,**TGM5**,CASP14,KRT23,TGM3,DSG1,DSC1,**FLG2**,**FLG**,**CERS3**,**LCE2B**,**SPRR2G**,**LCE3D**,ATP7A,KRT80,KRT1,ACER1,KRT78,KRT72,KRT2,LIPM,KRT73,**LCE1A**,KRTDAP,KRT77,C1orf68,**LOR**,**SPRR2E**,LIPN,LIPK
GO:0043588	skin development	11,86	**TGM5**,CASP14,KRT23,**TGM3**,DSG1,DSC1,**FLG2**,**FLG**,**CERS3**,**SPRR2G**,**LCE3D**,ATP7A,KRT80,KRT1,ACER1,KRT78,KRT72,KRT2,LIPM,**ALOXE3**,**ALOX12B**,KRT73,**LCE1A**,KRT77,**LOR**,**SPRR2E**,LIPN,LIPK,ASPRV1
GO:0030216	keratinocyte differentiation	9,48	**TGM5**,CASP14,KRT23,**TGM3**,DSG1,DSC1,**FLG**,**CERS3**,**SPRR2G**,**LCE3D**,KRT80,KRT1,ACER1,KRT78,KRT72,KRT2,LIPM,KRT73,**LCE1A**,KRT77,**LOR**,**SPRR2E**,LIPN,LIPK
GO:0060429	epithelium development	5,66	USH2A,DLX3,CTSH,**TGM5**,CASP14,UPK1A,KRT23,WNT3,POF1B,**TGM3**,DSG1,DSC1,ANXA9,**FLG2**,**FLG**,GJA1,**CERS3**,**SPRR2G**,**LCE3D**,CITED2,ATP7A,KRT80,KRT1,ACER1,KRT78,KRT72,KRT2,LIPM,PSAPL1,KRT73,**LCE1A**,KRT77,**LOR**,**SPRR2E**,LIPN,LIPK,EPPK1
GO:0006665	sphingolipid metabolic process	4,34	**CERS4**,STS,SMPD3,ALDH5A1,ELOVL4,**CERS3**,SGPP2,ELOVL7,ACER1,DEGS2,**ALOXE3**,**ALOX12B**,SPTSSB
GO:0018149	peptide cross-linking	4,29	**TGM5**,**TGM3**,**FLG**,KRT1,KRT2,**LOR**,**SPRR2E**
GO:0006672	ceramide metabolic process	3,19	**CERS4**,SMPD3,ALDH5A1,**CERS3**,ACER1,DEGS2,**ALOXE3**,**ALOX12B**,SPTSSB
REAC:R-HSA-9032759	NTRK2 activates RAC1	2,80	DOCK3,NTRK2,BDNF
GO:0006629	lipid metabolic process	2,50	DGAT2,RORA,APOB,**CERS4**,PTGS1,SORBS1,STS,SMPD3,ALDH5A1,PRLR,ELOVL4,PLBD1,CD36,TM7SF2,CERS3,TTC39B,PLA2G2F,ABCG1,SGPP2,PLB1,HPGD,ELOVL7,FAAH2,ACER1,DEGS2,SDR16C5,CES4A,PSAPL1,ALOXE3,ALOX12B,PNPLA1,TNFAIP8L3,CYP4F12,THEM5,SPTSSB,BCO2,ABCA4
GO:0046513	ceramide biosynthetic process	2,47	**CERS4**,SMPD3,**CERS3**,DEGS2,ALOXE3,ALOX12B,SPTSSB
GO:0009620	response to fungus	2,38	ARG1,IL36RN,RNASE7,CLEC7A,NLRP10
REAC:R-HSA-9014826	Interleukin-36 pathway	2,32	IL36RN,IL36B,IL1F10

Table 2. Top enriched pathways of genes in cluster 1

Trem ID	Term Name	-log10(p-val)	Proteins
REAC:R-HSA-6798695	Neutrophil degranulation	3,91	FTH1,ARG1,ALAD,GM2A,S100P,SERPINB3,DSC1,ASAH1,GGH,CALML5,DPP7
KEGG:01100	Metabolic pathways	3,79	IMPA2,ALOX15B,AKR1B10,GGCT,CA2,ARG1,ALDH2,CKB,ALAD,CBR1,GOT1,HAL,ACAA2,GALK1,ASAH1,HAGH,FAHD1
GO:0044281	small molecule metabolic process	3,74	SDR16C5,IMPA2,ALOX15B,AKR1B10,CYB5A,SOD1,ARG1,ALDH2,CKB,ETFA,CBR1,GOT1,GLRX,HAL,ACAA2,GALK1,ASAH1,BLMH,HAGH,GGH,ACOT13
GO:0008544	epidermis development	3,70	SOD1,IVL,FLG,CASP14,KLK7,TGM3,DSC1,ASAH1,C1orf68,CALML5
GO:0002283	neutrophil activation	3,67	FTH1,ARG1,ALAD,GM2A,S100P,SERPINB3,DSC1,ASAH1,GGH,CALML5,DPP7
GO:0045055	regulated exocytosis	3,67	SOD1,FTH1,ARG1,ALAD,GM2A,S100P,SERPINB3,DSC1,ASAH1,ECM1,GGH,CALML5,DPP7
GO:0030855	epithelial cell differentiation	3,37	ALOX15B,SOD1,IVL,CBR1,FLG,CASP14,RAB25,TGM3,DSC1,ASAH1,PSAPL1
GO:0009888	tissue development	2,85	ALOX15B,SOD1,IVL,CBR1,FLG,SERPINB3,CASP14,KLK7,RAB25,TGM3,DSC1,ASAH1,ECM1,C1orf68,PSAPL1,CALML5
GO:0140352	export from cell	2,63	ALOX15B,SOD1,FTH1,ARG1,ALAD,GM2A,S100P,SERPINB3,RAB25,DSC1,ASAH1,ECM1,GGH,CALML5,DPP7
GO:0009913	epidermal cell differentiation	2,43	SOD1,IVL,FLG,CASP14,TGM3,DSC1,ASAH1
KEGG:00790	Folate biosynthesis	2,42	AKR1B10,CBR1,GGH
GO:0018149	peptide cross-linking	2,14	IVL,FLG,TGM3

Table 3. Top enriched pathways of proteins in cluster 1

Longitudinal multi-omics data Network construction (NetOmics)

Furthermore, we highlighted the interactions between the multiple omics data by building a 3-layers network for each cluster: a gene inference layer from the mRNA using the ARACNe algorithm [22], a protein-protein interaction layer build from BioGRID known interactions, and a metabolite layer from KEGG pathways. The first two layers were combined thanks to two types of links. First, protein-coding information linked genes to their corresponding proteins. Second, TF-regulated information from TF2DNA and TTRUST [19, 20] databases linked proteins to genes. Focusing on cluster 1, the constructed network was composed of 420 gene-gene interactions, 87 gene-protein interactions, 320 protein-protein interaction and 126 protein-metabolite interactions (Figure 3A). For instance, beyond being clustered together, we found connections (edges) between the HOPX protein/transcription factor (TF) (red circle, Figure 3A and Figure 3B) and the FLG and LOR genes reflecting a functional interaction between these biomolecules. Indeed, this is concordant with previous reports showing that experimental HOPX depletion in keratinocytes results in LOR and FLG gene expression regulation [21, 23]. Then, we focused on the glutaredoxin (GLRX) gene-protein pair in the network (Figure 3C, red arrows). GLRX is a small molecular-weight protein and a member of the thioredoxin family. Previous reports have shown that it plays a protective role against oxidative stress in lung and neuronal cells [24]. Moreover, glutaredoxin deficiency has been shown to promote activation of the transforming growth factor beta pathway in airway epithelial cells, and suggest a mechanism that involves the plasticity of basal cells, the stem cells of the airways [25]. GLRX is connected to the superoxide dismutase (SOD1) protein (blue arrow) which is an important antioxidative enzyme that protects skin from oxidative stress. SOD1 is itself connected to a group of molecules, in particular the transaminase enzyme GOT1, the glyoxalase HAGH and the creatine kinase CKB (yellow arrows), all of them described to be associated with the oxidative metabolism [26-28]. For example, increased expression of GOT1 fuels oxidative metabolism of cells cultured in low pH by enhancing the non-canonical glutamine metabolic pathway, and cell survival in low pH is reduced upon depletion of GOT1 due to increased intracellular ROS levels [26]. Glyoxalase pathway is a well-conserved antioxidant defense system and responds to the constantly changing cellular environment. An impaired glyoxalase system can result in increased oxidative and dicarbonyl stress and is implicated in various disease conditions [27]. Collectively, these results provide a potential role for the induction of antioxidative pathway during keratinocyte differentiation which needs to be further explored.

The identification of this pathway was facilitated thanks to net construction from interacting molecules of distinct cellular omics layers.
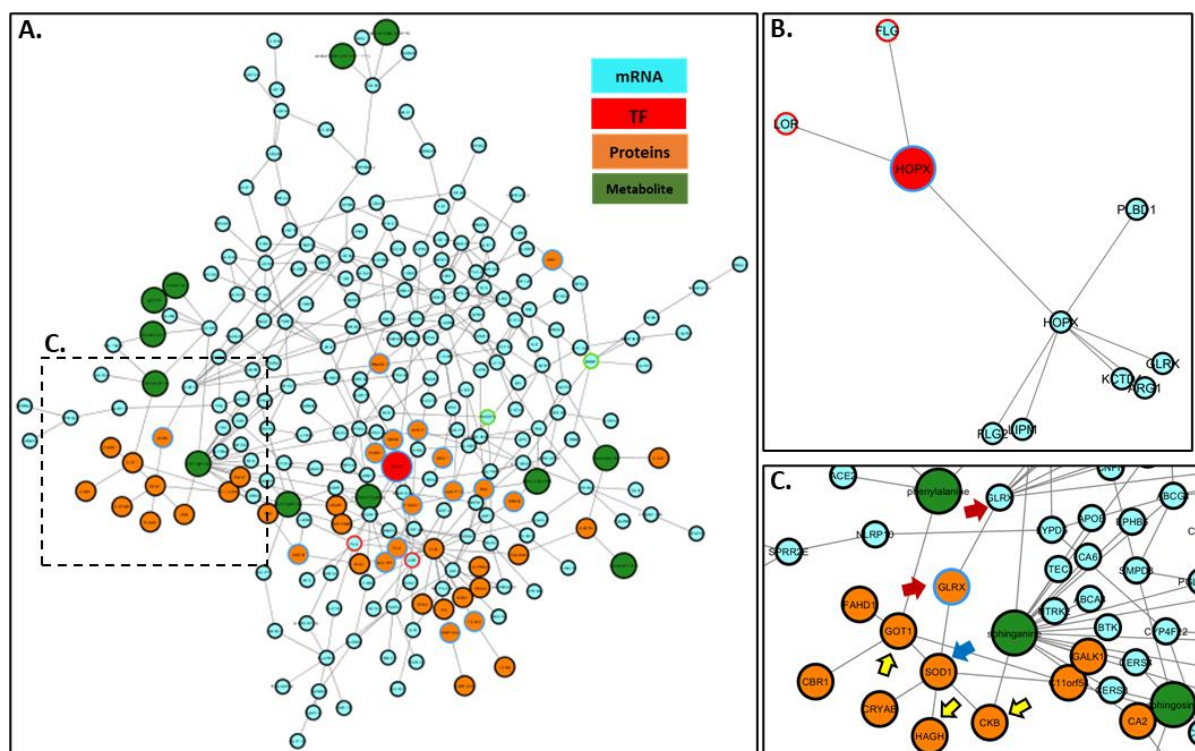

Figure 3. Network representation of biomolecules belonging to cluster 1. A. 3-layer network was built with a gene inference layer network from the mRNA using the ARACNe algorithm, a protein-protein interaction layer build from BioGRID known interactions, and a metabolite layer from KEGG pathways. The first two layers were combined thanks to two types of links. First, protein-coding information linked genes to their corresponding proteins. Second, TF-regulated information from TF2DNA and TTRUST databases linked proteins to genes. **B**. Focus on HOPX network. **C**. Focus on glutaredoxin network and connection to other biomolecules.

Discussion

It is well known that the differentiation process of keratinocytes is regulated by several biological signaling pathways [1-4], but up to now, a capture of interactions between different molecular layers composing these pathways has lacked. Here, we present a resource to help deciphering the regulatory mechanisms occurring during epidermal reconstruction.

Dense longitudinal profiling of the transcriptome, proteome and metabolome throughout the differentiation process first enabled the identification of the dynamic trajectories of a set of 1012 genes, 222 proteins and 226 metabolites. We show that the trajectories of genes and proteins composing the cluster 1 were associated with a late differentiation state, and

interestingly, an agreement between transcriptomic and proteomic analyses on overrepresented pathways was obtained, with some peculiarities for each layer. The RHE model appears to be characterized by two phases, the initial formation of general epithelium (days 1-7) and the maturation of epidermal tissue and barrier formation (day 10 onwards). Previous studies exploring keratinocyte differentiation kinetics only by transcriptomic approaches on *in vitro* skin models reported similar results [2, 29]. Enrichment results of all clusters taken together will help to identify the molecular manipulation for a pathway of interest might be best studied by introducing changes to the culture condition at different time points of the epidermis reconstruction program.

Taking full advantage of the interactions between profiled omics and their interdependencies, we were able to identify an increase in the antioxidative stress players during the phases of epithelial development and barrier formation. Epidermis is particularly exposed to oxidative stress, either from environmental insults or because of specific impairments in antioxidant status resulting from pathologies or aging. Traditionally, antioxidant products are exogenously provided to neutralize pro-oxidant species. In our study, it appears that an endogenous antioxidant defense pathway is stimulated to ensure a proper differentiation process. This seems of particular interest for a positive maintenance of multiple features of epidermal organization and barrier formation.

Conclusion

We show in this study that the simultaneous analysis of interactions of biomolecules by means of multi-omics data clustering an integration is an invaluable approach to shed light on distinct stages underlying epidermis reconstruction processes, which is a powerful base of knowledge to develop new cosmetics and dermatologically active ingredients.

Conflict of Interest Statement.

NONE.

References.

1. Taylor, J.M., et al., *Dynamic and physical clustering of gene expression during epidermal barrier formation in differentiating keratinocytes.* PLoS One, 2009. **4**(10): p. e7651.

2. Bachelor, M., et al., *Transcriptional profiling of epidermal barrier formation in vitro.* J Dermatol Sci, 2014. **73**(3): p. 187-97.

3. Winget, J.M., et al., *Quantitative proteogenomic profiling of epidermal barrier formation in vitro.* J Dermatol Sci, 2015. **78**(3): p. 173-80.

4. Toufighi, K., et al., *Dissecting the calcium-induced differentiation of human primary keratinocytes stem cells by integrative and structural network analyses.* PLoS Comput Biol, 2015. **11**(5): p. e1004256.

5. Michaletti, A., et al., *Multi-omics profiling of calcium-induced human keratinocytes differentiation reveals modulation of unfolded protein response signaling pathways.* Cell Cycle, 2019. **18**(17): p. 2124-2140.

6. Bodein, A., et al., *timeOmics: an R package for longitudinal multi-omics data integration.* Bioinformatics, 2021.

7. Bodein, A., et al., *Interpretation of network-based integration from multi-omics longitudinal data.* Nucleic Acids Research, 2021. **50**(5): p. e27-e27.

8. Andrews, S., *FastQC A Quality Control Tool for High Throughput Sequence Data.* https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ 2010.

9. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.

10. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification.* Nat Biotechnol, 2016. **34**(5): p. 525-7.

11. Soneson, C., M.I. Love, and M.D. Robinson, *Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences.* F1000Res, 2015. **4**: p. 1521.

12. Anders, S., et al., *Count-based differential expression analysis of RNA sequencing data using R and Bioconductor.* Nat Protoc, 2013. **8**(9): p. 1765-86.

13. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biology, 2014. **15**(12): p. 550.

14. Bodein, A., et al., *A Generic Multivariate Framework for the Integration of Microbiome Longitudinal Studies With Other Data Types.* Frontiers in Genetics, 2019. **10**.

15. Straube, J., et al., *A Linear Mixed Model Spline Framework for Analysing Time Course 'Omics' Data.* PLoS One, 2015. **10**(8): p. e0134540.

16. Rohart, F., et al., *mixOmics: An R package for 'omics feature selection and multiple data integration.* PLoS Comput Biol, 2017. **13**(11): p. e1005752.

17. Raudvere, U., et al., *g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update).* Nucleic Acids Res, 2019. **47**(W1): p. W191-w198.

18. Reimand, J., et al., *g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W193-200.

19. Pujato, M., et al., *Prediction of DNA binding motifs from 3D models of transcription factors; identifying TLX3 regulated genes.* Nucleic Acids Res, 2014. **42**(22): p. 13500-12.

20. Han, H., et al., *TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions.* Nucleic Acids Res, 2018. **46**(D1): p. D380-d386.

21. Obarzanek-Fojt, M., et al., *Homeodomain-only protein HOP is a novel modulator of late differentiation in keratinocytes.* European Journal of Cell Biology, 2011. **90**(4): p. 279-290.

22. Margolin, A.A., et al., *ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context.* BMC Bioinformatics, 2006. **7 Suppl 1**(Suppl 1): p. S7.

23. Yang, J.M., et al., *Expression of the homeobox gene, HOPX, is modulated by cell differentiation in human keratinocytes and is involved in the expression of differentiation markers.* Eur J Cell Biol, 2010. **89**(7): p. 537-46.

24.   Peltoniemi, M., et al., *Expression of glutaredoxin is highly cell specific in human lung and is decreased by transforming growth factor-beta in vitro and in interstitial lung diseases in vivo.* Hum Pathol, 2004. **35**(8): p. 1000-7.

25.   Chia, S.B., et al., *Glutaredoxin deficiency promotes activation of the transforming growth factor beta pathway in airway epithelial cells, in association with fibrotic airway remodeling.* Redox Biology, 2020. **37**: p. 101720.

26.   Abrego, J., et al., *GOT1-mediated anaplerotic glutamine metabolism regulates chronic acidosis stress in pancreatic cancer cells.* Cancer Lett, 2017. **400**: p. 37-46.

27.   Aragonès, G., et al., *Glyoxalase System as a Therapeutic Target against Diabetic Retinopathy.* Antioxidants (Basel, Switzerland), 2020. **9**(11): p. 1062.

28.   Meyer, L.E., et al., *Mitochondrial creatine kinase activity prevents reactive oxygen species generation: antioxidant role of mitochondrial kinase-dependent ADP re-cycling activity.* J Biol Chem, 2006. **281**(49): p. 37361-71.

29.   Lopez-Pajares, V., et al., *A LncRNA-MAF:MAFB transcription factor network regulates epidermal differentiation.* Dev Cell, 2015. **32**(6): p. 693-706.