Intrinsic Differences between the Open Chromatin Regions of Oral and Epidermal Keratinocytes

Yao XIAO[1], Huan LIU[1,2]

Objective: *To investigate and characterise the differences between the open chromatin regions of oral and epidermal keratinocytes.*
Methods: *Human immortalised oral epithelial cell lines (HIOECs) were used as the standard model for oral keratinocytes, and primary normal human epidermal keratinocytes (NHEKs) were chosen as the model for epidermal keratinocytes. Assay for transposase accessible chromatin using sequencing (ATAC-seq) and H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) were used to evaluate the dynamic changes in open chromatin regions and active enhancers during oral keratinocyte differentiation. In silico prediction and dual-luciferase assays were used to evaluate the enriched motifs and maintain enhancer activity in specific enriched HIOECs. Integration and comparison of HIOEC ATAC-seq with NHEK ATAC-seq were used to identify oral keratinocyte–enriched open chromatin regions along with key motifs governing differential enhancer activity. The genomic regulatory elements and GWAS overlap algorithm was used to compare the annotation rate of HIOEC-overlapped craniofacial enhancers with other craniofacial enhancers for orofacial cleft–associated variants.*
Results: *During the differentiation of HIOECs, 14933 open chromatin regions became more accessible. Grainyhead-like (GRHL) and Krüppel-like factor (KLF) motifs were overrepresented in maintaining HIOEC-specific activity. Compared with NHEKs, 16161 open chromatin regions were uniquely accessible in HIOECs. Within these regions, the C/EBP motif governed HIOEC-specific enhancer regulating SOX2 and PITX2, which enhanced oral keratinocyte wound healing. When intersected with human craniofacial super-enhancers, open chromatin regions in HIOECS can better annotate the common variants associated with orofacial cleft.*
Conclusion: *The intrinsic differences between the open chromatin regions of human oral and epidermal keratinocytes are directly maintained by a set of transcription factors.*
Key words: *keratinocytes, open chromatin regions, transcription factors*
Chin J Dent Res 2020;23(2):119–130; doi: doi: 10.3290/j.cjdr.a44748

1 State Key Laboratory Breeding Base of Basic Science of Stomatology (Hubei-MOST) and Key Laboratory for Oral Biomedicine of Ministry of Education (KLOBM), School and Hospital of Stomatology, Wuhan University, Wuhan, P.R. China.
2 Department of Periodontology, School of Stomatology, Wuhan University, Wuhan, P.R. China.

Corresponding author: Dr Huan LIU, State Key Laboratory Breeding Base of Basic Science of Stomatology (Hubei-MOST) and Key Laboratory for Oral Biomedicine of Ministry of Education (KLOBM), School and Hospital of Stomatology, Wuhan University, Wuhan 430079, P.R. China. Email: liu.huan@whu.edu.cn

This research was supported by grants from the Young Elite Scientist Sponsorship Program by CAST (NO. 2017QNRC001) (HL), from the Natural Science Foundation of Hubei Province (NO. 2017CFB515) (HL) and from the National Natural Science Foundation of China (NO. 81771057; 81400477) (HL).

Cutaneous wound healing is a complex sequential biological process in which multiple types of cells are involved[1]. Re-epithelialisation is the most critical step in this process, involving the restoration, migration and terminal differentiation of epidermal cells. Keratinocytes are essential for effective wound re-epithelialisation[2].

Although the skin and oral mucosa are both covered by keratinocytes, their morphology and physiological characteristics, including wound healing ability and migration of keratinocytes from these two types of tissues, vary greatly[3,4]. Previous studies have revealed that in comparison to skin wounds, oral wounds exhibit a lower inflammatory response with lower neutrophil,

macrophage and T-cell infiltration[3,4] and differential expression of proinflammatory, profibrotic cytokines such as TGF-β1[5]. However, little is known about the transcriptional regulation for all these genes, i.e., the potential determinants for the fate of different keratinocytes.

Regulation of transcription is mediated by transcription factors (TFs) that bind to specific motifs within the cis-regulatory elements in the genome DNA and eventually influence gene expression. Sets of TFs have been regarded as cell fate determinants; for instance, ectopic expression of SOX2, Oct4, KLF4 and c-myc could reprogram somatic cells into pluripotent stem cells[6]. We and other groups have previously shown epidermal keratinocyte differentiation to be regulated by a set of dynamic gene regulatory networks with multiple TF modules involved (reviewed by Klein and Andersen[7]). During re-epithelialisation, MAF and MAFB serve to initiate differentiation of epidermal progenitor cells; later, GRHL3, KLF4, ZNF750 and PRDM1 are upregulated and promote terminal differentiation of keratinocytes and wound healing. Recently, SOX2 and PITX2 were proved to be uniquely expressed during oral mucosa wound healing, and forced expression of these two factors can reprogram epidermal keratinocytes and enhance their wound healing ability in vivo[8].

Similar to the role of TFs in cell fate determination, cell type–specific open chromatin landscapes have also been found to be associated with cell fate[9]. Many cell type–specific open chromatin regions are intergenic and bound by cell type–specific TFs. Among these regions, active enhancers, flanked by H3K27ac modification[10], are highly related to gene activation. Despite findings about dynamic changes in TF sets governing epidermal keratinocyte differentiation, little is known about how TFs regulate oral keratinocyte differentiation and whether there are different sets of TFs differentiating oral keratinocytes from epidermal keratinocytes. Thus, we hypothesised that oral keratinocyte–specific TFs could bind to the oral keratinocyte–specific active enhancers, making oral keratinocytes different from epidermal keratinocytes. In the present study, we first analysed the landscape of open chromatin regions in oral keratinocytes. We performed ATAC-seq and H3K27ac ChIP-seq to reveal the identical active enhancers governing terminal differentiation of oral keratinocytes. By comparing the open chromatin regions of epidermal and oral keratinocytes, we identified regions with differential enhancer activity along with TF motifs differentially enriched in different cells. Moreover, we found that nucleosome-free regions (NFRs) enriched in oral keratinocytes can better annotate common variants for orofacial cleft than the other commonly used epigenome datasets.

Materials and methods

Cell culture

The HIOEC lines[11] and NHEKs (purchased from Lonza, Cologne, Germany) were maintained in Keratinocyte SFM (1x) (Life Technologies, Carlsbad, CA, USA) supplemented with epidermal growth factor (EGF) and bovine pituitary extract (Life Technologies). Human embryonic kidney 293FT cells (purchased from Invitrogen [Life Technologies]) were maintained in Dulbecco's Modified Eagle Medium (DMEM) (Hyclone, Pittsburgh, PA, USA) supplemented with 10% foetal bovine serum (Hyclone). Human embryonic palatal mesenchymal cells (HEPMs; CRL-1486, purchased from ATCC [Gaithersburg, MD, USA] were maintained in α-MEM (Hyclone) supplemented with 10% foetal bovine serum (Hyclone). To induce keratinisation of HIOECs, when cells reached 90% confluency, the medium was supplemented with Ca^{2+} for three days. All cells were incubated in 5% CO_2 incubators. All the cell lines used in this study were tested for mycoplasma contamination and were free of contamination.

Assay for transposase-accessible chromatin using sequencing (ATAC-seq) library preparation and high-throughput sequencing

We prepared the ATAC-seq library according to a previously published protocol[12]. Cells that were cultured in 24-well plates were subjected to 50 µL of cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40; all components purchased from Sigma [St Louis, MO, USA]) in a centrifuge at 500 × g for 15 minutes at 4°C. Pelleted nuclei were resuspended in 50 µL of tagmentation reaction mix (25 µL Nextera TD Buffer, 2.5 µL Nextera TD Enzyme and 22.5 µL H_2O, all from the Nextera DNA Sample Preparation Kit [Illumina]). We performed tagmentation at 37°C for 30 minutes in a thermocycler (Eppendorff, Hamburg, Germany). Immediately after the reaction was completed, the DNA was purified using a MinElute PCR Purification Kit (QIAGEN, Germantown, MD, USA). Eluted DNA was subjected to PCR (polymerase chain reaction) amplification and library indexing using the NEBNext® High-Fidelity 2X PCR Master Mix (New England Biolabs, Ipswich, MA, USA) with different customised Nextera PCR primer pairs for library index. PCR was

performed as follows: 72°C for 5 minutes; 98°C for 30 seconds; 11 cycles of 98°C for 10 seconds, 63°C for 30 seconds and 72°C for 1 minute; and hold at 4°C. The PCR product was purified with 1.8 × volume (90 μL for each sample) of Ampure XP beads (Beckman Coulter, Brea, CA, USA) to produce 18 μL of final library. Library quality was assessed using 1 μL of the final purified DNA on a BioAnalyzer 2100 High Sensitivity DNA Chip (Agilent Technologies, Santa Clara, CA, USA). All DNA libraries that exhibited a nucleosome pattern in the BioAnalyzer 2100 assay passed the pre-sequencing QC process and were pooled for high-throughput sequencing in HiSeq 2500, HiSeq4000 or HiSeq X Ten (Illumina, provided by Annoroad Genomics Company, Beijing, China).

Raw ATAC-seq fastq files that passed QC by FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) were trimmed with Trimmomatic v0.38[13] (using NexteraPE-PE.fa as adapter file) and mapped to hg19 reference genome build using Bowtie 2 (Johns Hopkins University, Baltimore, MD, USA)[14] (default parameters plus parameters -X2000 and -m1). After removal of duplicated reads with Picard (http://broadinstitute.github.io/picard/) and sorting with SAMtools[15], we identified insert fragments shorter than 100 bp as NFRs using a customised Python script (depending on pysam) as previously described[12]. To assess the reproducibility of the biological replicates in each experiment and generate bigwig output for visualisation, we utilised deepTools 2.0[16]. Peak calling was performed with MACS2 v2.1.1[17] (parameter: -- nomodel -- nolambda -- gsize 2.7e9 -- keep-dup all -- slocal 10000).

Chromatin immunoprecipitation of H3K27 acetylation (H3K27ac) combined with high-throughput sequencing (ChIP-seq)

HIOEC cells were seeded at 1×10^5 cells per 100 mm plate (1 plate per biological replicate), grown to 90%–100% confluency (the medium was refreshed every other day) and subjected to 1.2 mM Ca^{2+} in a culture medium for 3 days. Cells were washed with ice-cold phosphate buffered saline (PBS, Hyclone) and fixed with 1% paraformaldehyde (PFA) for 10 minutes at room temperature. PFA was then quenched in 134M Glycine (Sigma) for 5 minutes at room temperature, and cells were collected with a cell scraper in ice-cold PBS and using centrifuge. The cell pellets were resuspended with 5 mM PIPES pH 8.5, 85 mM KCl, 1% (v/v) IGEPAL CA-630, 50mM NaF, 1 mM phenylmethylsulphonyl fluoride (PMSF), 1mM phenylarsine oxide, 5 mM sodium orthovanadate and cOmplete™ Protease Inhibitor Cocktail (Sigma). After sonication and quantification using Qubit 3.0 (Life Technologies), chromatin immunoprecipitation was performed using 4ug of anti-histone H3, acetylated lysine 27 (H3K27ac [ab4729]) (Abcam, Cambridge, UK) per 500,000 cells. ChIP-seq libraries were indexed using a VAHTS Universal DNA Library Prep Kit (ND606-01, Vazyme Biotech, Nanjing, China) and with TrueSeq3 PE primer sets. After quantification and qualification, 150 bp paired-end sequencing was performed using the HiSeq X Ten sequencer (Illumina, provided by Annoroad Genomics, Beijing, China). Output sequences were trimmed using Trimmomatic v0.38[13] (parameter: ILLUMINACLIP: TrueSeq3PE-PE.fa:2:30:10:8:TRUE LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:30). All trimmed, paired reads were aligned to human genome assembly 19 (hg19) using Bowtie (default parameters)[14]. Peaks were called using MACS2 v2.1.1[17]. Visualisation and post-mapping QC were performed in the same way as for ATAC-seq management described above.

Integration and comparison of different ATAC-seq and H3K27ac ChIP-seq datasets

NHEK ATAC-seq results were downloaded from GSE6738218 and processed as described above. NHEK H3K27ac ChIP-seq raw data were downloaded from Roadmap Epigenomics Project[10] (http://www.roadmapepigenomics.org/). To compare the ATAC-seq reads between HIOEC D0 and D3 or HIOECs and NHEKs, we performed differential consensus peak calling using the R package DiffBind[19] with a fold-change threshold of 2, and FDR < 0.01. To integrate H3K27ac ChIP-seq with ATAC-seq results, we aligned all the H3K27ac ChIP-seq reads to different sets of NFRs using deepTools v2.0[16].

Motif enrichment assay

All the genomic regions of interest were trimmed into 500 bp centred with the summit of each peak. Using HOMER[20], we generated different lists of enriched motifs in each dataset, and TFs with a score higher than 0.8 were considered to match the potential enriched binding site.

Plasmid constructs and dual luciferase assay

Human genomic regions were amplified using human genomic DNA, which was extracted from HEPMs. Site-directed mutagenesis experiments were performed using overlapping PCR as described before[21]. Wildtype or mutant DNA fragments were inserted into

pGL3-promoter plasmids and Sanger sequencing was performed to validate the sequence. pGL3 plasmids inserted with different DNA fragments were coelectroporated with Renilla reporter into HIOECs, NHEKs or HEPMs with 4D-Nucleofector™ (Lonza). The Dual-Luciferase Reporter Assay System (Promega, Madison, WI, USA) and GloMax® 20/20 Luminometer (Promega) were employed to evaluate the ratio between firefly and Renilla activity. Three independent measurements were performed for each biological replicate, and all the results were presented as mean ± s.d. A Student *t* test was used to determine the statistical significance.

RNA isolation and quantitative real-time PCR (qRT-PCR)

Total RNA was extracted from HIOECs and NHEKs using TRIzol reagent (Invitrogen). RT Premix kit (Takara Bio Inc, Tokyo, Japan) was employed to synthesise the cDNA. Quantitative PCR (qPCR) was performed using the Applied Biosystems 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) and SYBER® Premix Ex TaqTM (Takara Bio Inc).

Data availability

All the data and materials related to this work are available on request.

Results

Landscape of open chromatin regions during differentiation of oral keratinocytes

To characterise the changes in open chromatin regions during differentiation of oral keratinocytes, we chose HIOECs as a standard cell line in the present study. We first confirmed the differentiation of HIOECs in vitro by using Ca^{2+} induction, and on day 3 of differentiation we found expression of KRT4 and KRT17 was strongly induced (data available on request). Thus, we chose day 0 and day 3 as two time points for genomic comparison. We first performed an on-plate ATAC-seq as described before[18] at two time points for three independent biological replicates (Fig 1a). Three biological replicates in each group were highly correlated with each other ($r2 > 0.99$ by Pearson test, data not shown), indicating strong data reproducibility. Depending on the nature of the open chromatin regions, we defined the reads from ATAC-seq that were shorter than 100 bp as NFRs[12]. By comparing the accessibility of NFRs between HIOEC day 0 and day 3, we found that after differentiation, the majority of the differential NFRs became more accessible (Figs 1b and 1d). In total, we identified that 14933 NFRs were significantly enriched on day 3 (data available on request). Using gene ontology (GO) enrichment analysis, we found the genes near these regions were significantly enriched for inflammation response and epidermal differentiation, correlating with the previously reported oral keratinocytes profile. For instance, compared with day 0, the NFRs near the KRT17 locus were more accessible on day 3 (Fig 1c), which was consistent with the induction of KRT17 during HIOEC in vitro differentiation.

Although it has been shown that the majority of the tissue-specific ATAC-seq peaks are active enhancers[22], we asked whether these differential NFRs exhibited different patterns of H3K27ac ChIP-seq, a marker for tissue-specific active enhancers. We then performed H3K27ac ChIP-seq for HIOEC on day 0 and day 3, then integrated the results with differential NFRs. As expected, the NFRs that were more accessible on day 3 were significantly enriched or flanked with H3K27ac ChIP-seq signals on day 3, but depleted with H3K27ac ChIP-seq on day 0 (Figs 1b and 1e). On the other hand, H3K27ac ChIP-seq signals in the NFRs enriched on day 0 did not exhibit any specific pattern (Fig 1b). Since active enhancers are NFRs flanked by H3K27ac modification, in the present study, we identified active enhancers modulating differentiation of oral keratinocytes.

KLF and GRHL motifs are overrepresented in NFRs enriched in differentiated HIOECs

TFs are supposed to determine or at least affect cell fate through binding to their motifs in the tissue- or cell-specific active cis-regulators. In order to figure out the major TF(s) governing HIOEC differentiation, we analysed the enriched motifs overrepresented in NFRs enriched in differentiated HIOECs.

Using HOMER, we identified 6 DNA motifs which corresponded to the specific TF families (Fig 2a). Among these families, cFos-JUN and TEAD are generic TFs overrepresented in most of the NFRs. The NFIC/C/EBP family, GRHL and KLF TFs have been shown to regulate epidermal cell differentiation and junction formation[7]. All NFRs with these motifs were assigned. Interestingly, of the 7764 NFRs occupied by GRHL motifs and 6195 NFRs occupied by KLF motifs, 3127 NFRs were co-occupied by both motifs. We analysed the function annotation for the nearby genes of these NFRs and found all the KLF-occupied NFRs were associated with genes for cytokine produc-

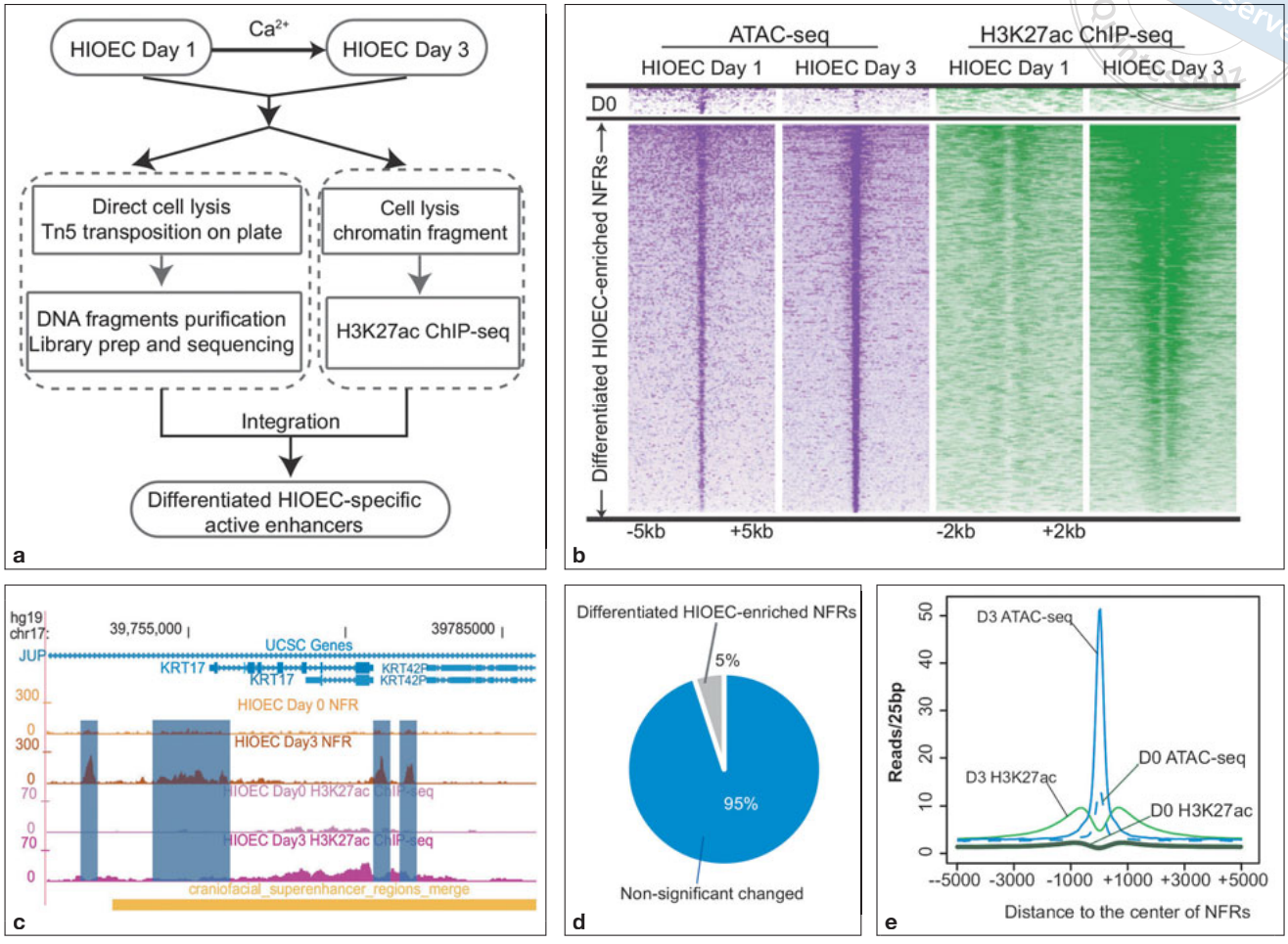


Fig 1 Dynamic change in accessibility of NFRs during HIOEC differentiation. **(a)** Workflow of on-plate ATAC-seq analysis and H3K27ac ChIP-seq during differentiation of HIOEC oral keratinocytes. **(b)** NFR summit-centred heatmap of ATAC-seq signal in HIOEC day 0 and HIOEC day 3 conditions, as well as H3K27ac ChIP-seq signal in the exact same regions for the differential ATAC-seq peaks. ATAC-seq peaks are sorted into the groups 'D0' and 'differentiated HIOEC-enriched NFRs'. **(c)** UCSC genome browser tracks showing different ATAC-seq (orange) and H3K27ac ChIP-seq (pink) signals on HIOEC day 0 (light colour) and HIOEC day 3 (dark colour). **(d)** Pie chart showing the distribution of all ATAC-seq peaks relative to differentiated HIOEC-enriched NFRs. **(e)** Plots of average density of ATAC-seq and H3K27ac ChIP-seq signals from HIOEC day 0 and day 3 conditions.

tion (Fig 2f). However, for the NFRs co-occupied by KLF and GRHL, their nearby genes were specifically enriched for epidermal formation, such as membrane permeability and junction formation (Fig 2g). These results showed that, in terms of NFR accessibility, KLF and GRHL served as a major TF set governing HIOEC differentiation with potential interaction.

Interestingly, we also noticed that CTCF, an over-represented motif in insulator DNA sequences, was also enriched in differentiated HIOEC NFRs (10%). Given the generic function of CTCF in NFRs as an insulator protein[23], we hypothesised that those occupied by CTCF motifs are not active enhancers. By integrating the H3K27ac ChIP-seq signals in the NFRs occupied by CTCF, KLF and GRHL, we found the H3K27ac ChIP-seq signals were significantly lower than those in KLF- or GRHL- occupied NFRs (Figs 2c to 2e) ($P<10–600$ by Kolmogorov-Smirnov test). Moreover, only a small proportion of the CTCF-occupied NFRs overlapped with either KLF- or GRHL-occupied NFRs (Fig 2b), indicating their distinct function in the open chromatin regions during HIOEC differentiation.

Given the enrichment of KLF and GRHL motifs in differentiated HIOEC NFRs, we asked whether they are necessary for maintaining enhancer activity of the epithelium-related genes. We chose three representative enhancers: one in the upstream of KRT14 (E1, hg19: chr17:39735946-39736442) (Fig 3a), with the presence of KLF motifs; one in the upstream of SFN (E2, chr1:27171296-27171673) (Fig 3c), with

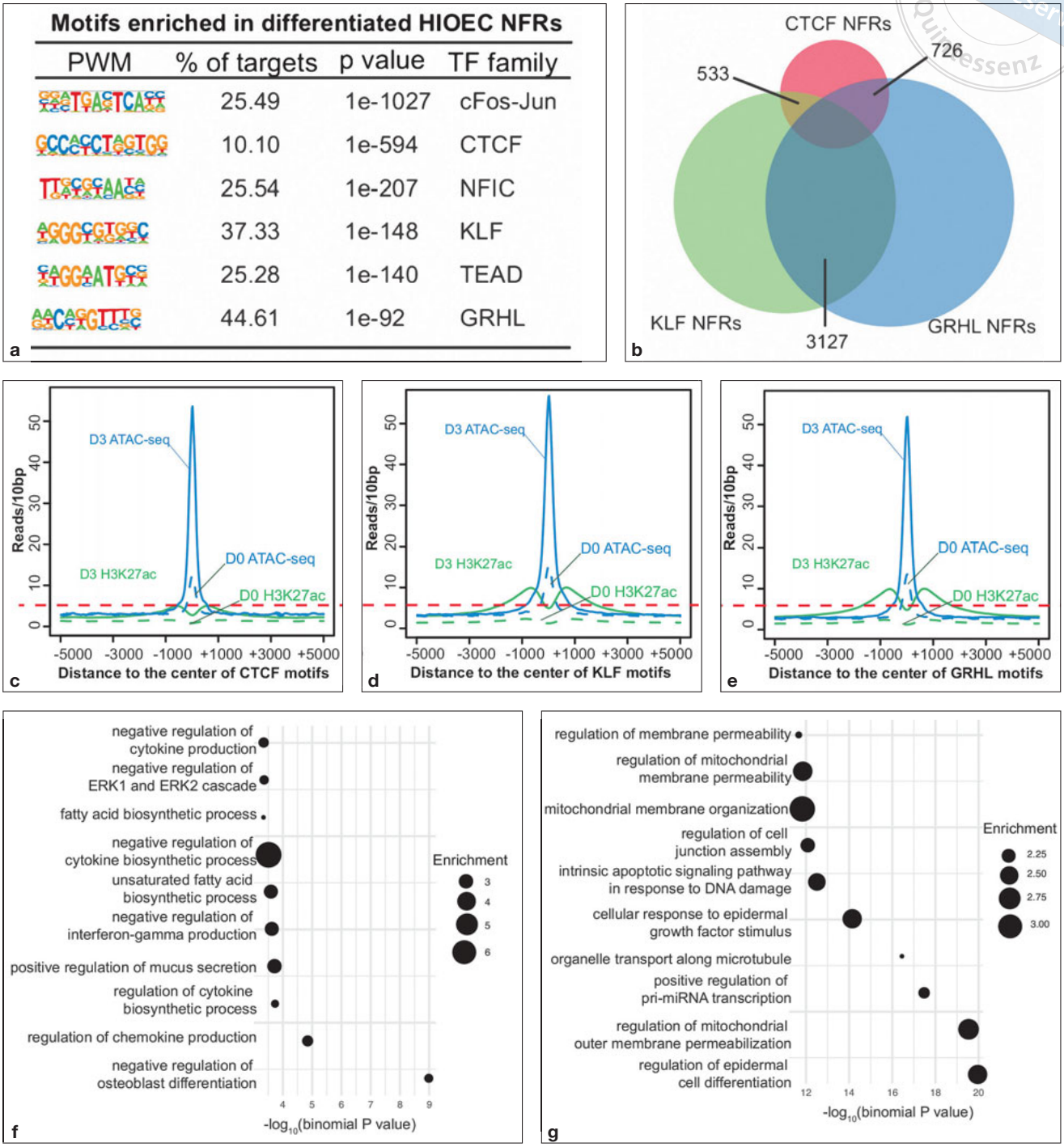


Fig 2 Overrepresented transcription factor motifs in differentiated HIOECs. **(a)** Enriched motifs in differentiated HIOEC-enriched NFRs. PWM, position weighted matrix; TF, transcription factor. **(b)** Venn diagram showing the intersection of differentiated HIOEC-enriched NFRs occupied with CTCF (CTCF NFRs), KLF (KLF NFRs) and GRHL motifs (GRHL NFRs). The numbers indicate the overlapped number of NFRs between two groups of NFRs. **(c)–(e)** Plots of average density of ATAC-seq and H3K27ac ChIP-seq signals in **(c)** CTCF NFRs, **(d)** KLF NFRs and **(e)** GRHL NFRs in HIOEC day 0 and day 3 conditions. The red dashed line across the three panels indicates the highest H3K27ac ChIP-seq signals in CTCF NFRs from HIOEC day 3 conditions. **(f)** Dot plot showing the GO enrichment assay for the nearby genes (within 100 bp) of KLF NFRs. **(g)** Dot plot showing the GO enrichment assay for the nearby genes (within 100 bp) of KLF-GRHL intersected NFRs.

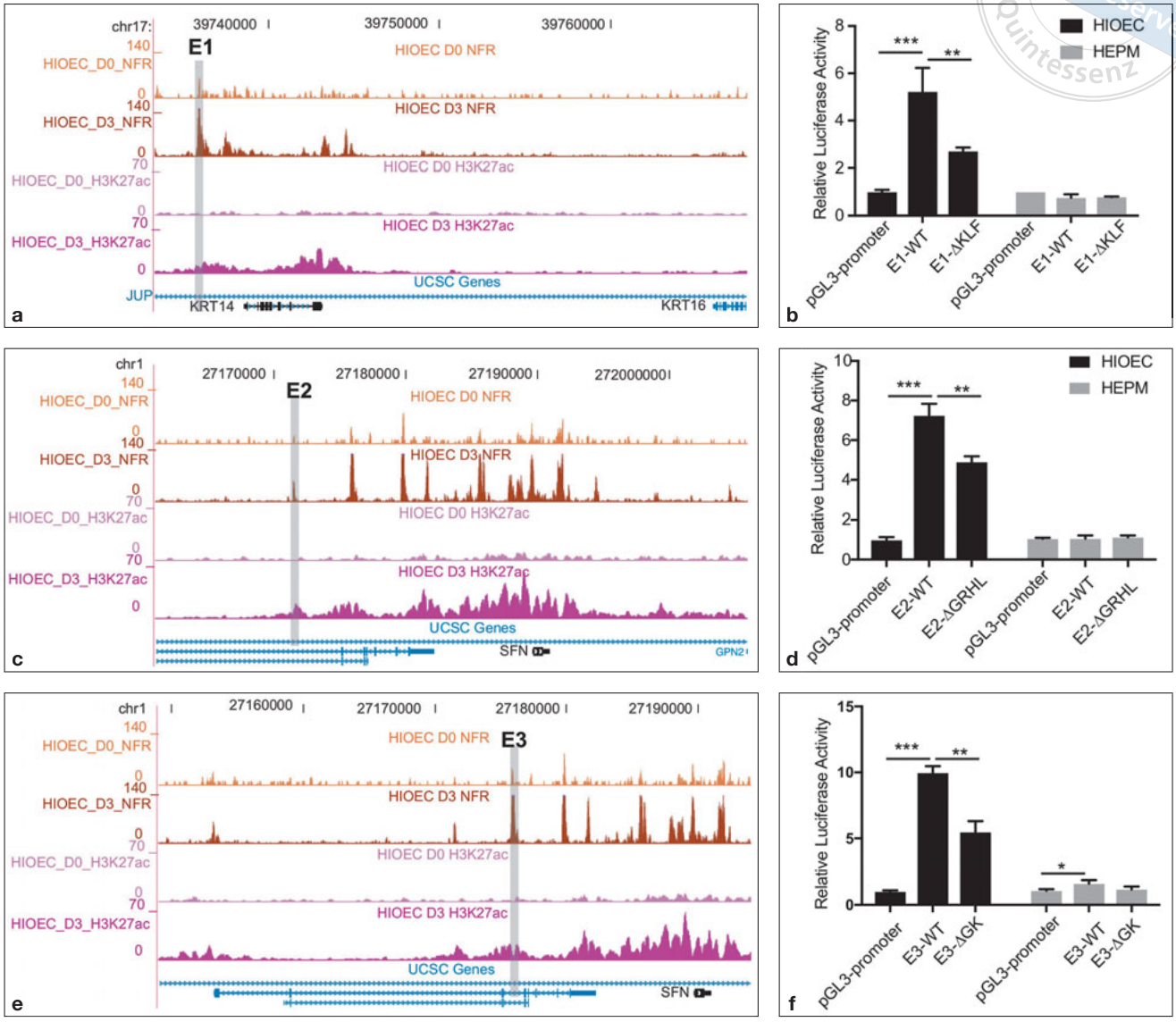
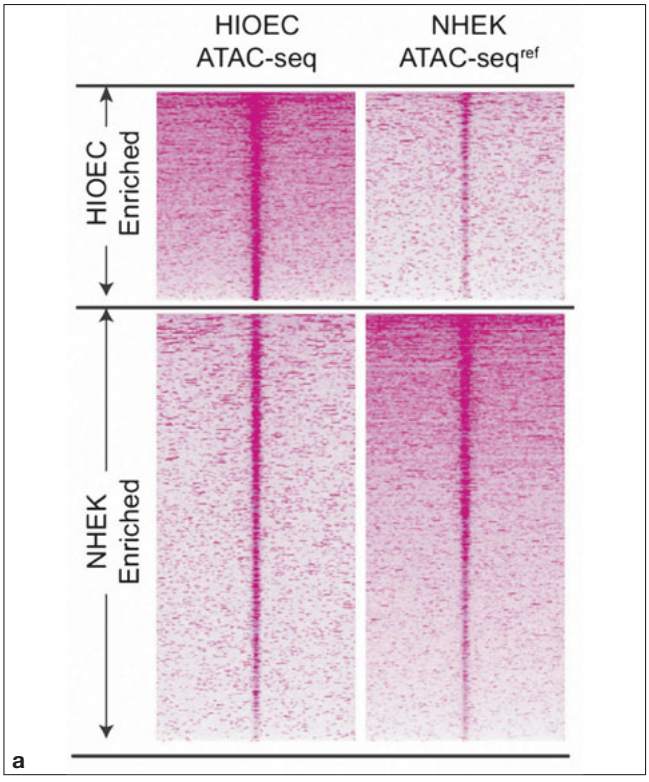


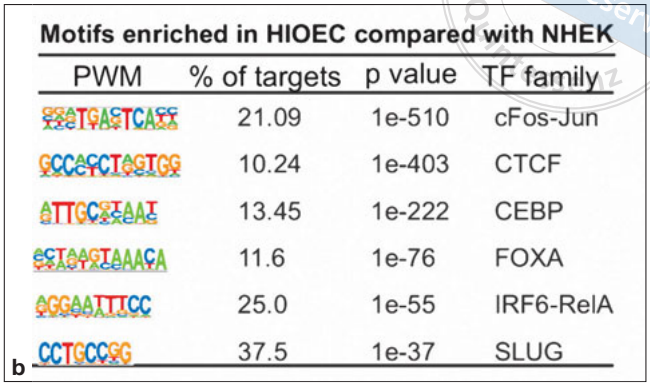
Fig 3 Reporter assay showing the role of KLF and GRHL motifs in maintaining HIOEC-specific enhancer activity. **(a)** UCSC genome browser tracks showing the genomic location of E1 in the KRT14 locus. **(b)** Dual luciferase assay showing the relative enhancer activity of E1 and E1 without KLF motif (E1-ΔKLF) in HIOECs and HEPMs. **(c)** UCSC genome browser tracks showing the genomic location of E2 in the SFN locus. **(d)** Dual luciferase assay showing the relative enhancer activity of E2 and E2 without GRHL motif (E2-ΔGRHL) in HIOECs and HEPMs. **(e)** UCSC genome browser tracks showing the genomic location of E3 in the SFN locus. **(f)** Dual luciferase assay showing the relative enhancer activity of E3 and E3 without KLF and GRHL motif (E3-ΔGK) in HIOECs and HEPMs. Student t tests were performed to evaluate the statistical significance of the difference. *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$.

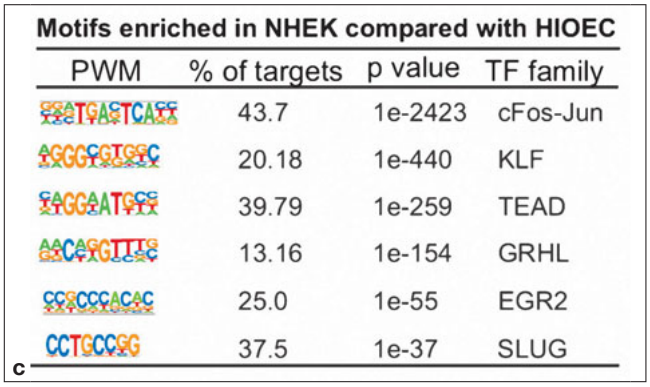
the presence of GRHL motifs; and the last one in the upstream of SFN (E3, chr1:27175645-27176211) (Fig 3e), with the presence of both GRHL and KLF motifs. We compared the enhancer activity of these enhancers in HIOECs and HEPMs (mesenchymal cell line) and found all of these enhancers exhibited significantly higher activity. Deletion of KLF motifs in E1 (Fig 3b), of GRHL motifs in E2 (Fig 3d) and of GRHL+KLF motifs in E3 (Fig 3f) also significantly reduced enhancer activity in HIOECs but not in HEPMs. Collectively, these assays indicated that GRHL and KLF motifs were essential for enhancer activity in the differentiated HIOEC cells.

C/EBP TF family can distinguish open chromatin regions in oral keratinocytes from those in skin

Although the oral mucosa and skin are both epidermis, their structure and physiological characteristics are not the same. Recently, SOX2 and PITX2 have been report-







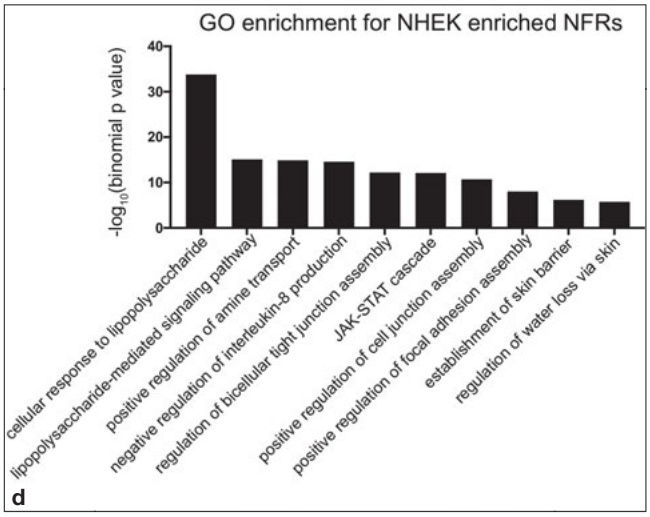


Fig 4 Comparison of ATAC-seq between HIOEC day 3 and NHEK day 4 conditions[18]. **(a)** NFR summit-centred heatmap of ATAC-seq signal in HIOEC day 3 and NHEK day 4 conditions. **(b)** Motifs enriched in NFRs from HIOEC conditions compared with NHEK. **(c)** Enriched motifs in NFRs from NHEK conditions compared with HIOEC. **(d)** Dot plot showing the GO enrichment assay for the nearby genes (within 100 bp) of NHEK-enriched NFRs.

ed to be two transcriptional regulators that uniquely distinguish oral keratinocytes from skin and have the potential to improve wound resolution[8]. We asked whether SOX2 and PITX2 reprogram skin through binding to some specific open chromatin regions. Thus, we compared our HIOEC NFRs (day 3) with published NHEK NFRs (day 4), and in both tissues, keratin genes were significantly induced. Between the NFRs in HIOECs and NHEKs, we identified 16161 HIOEC-enriched NFRs and 31696 NHEK-enriched NFRs (Fig 4a; more data available on request). As expected, a GO enrichment assay for the genes near NHEK-enriched NFRs showed these NFRs were related with epidermal cell functions such as cellular response to lipopolysaccharide and establishment of skin barrier (Fig 4d). However, for the motifs enriched in HIOEC or NHEK NFRs, neither SOX nor PITX motifs were overrepresented. Instead, IRF and C/EBP motifs were more enriched in the NFRs of HIOECs than those of NHEKs.

Wondering whether IRF or C/EBP acted upstream of SOX2 or PITX2 through binding to tissue-specific active enhancers, we searched for all the HIOEC-

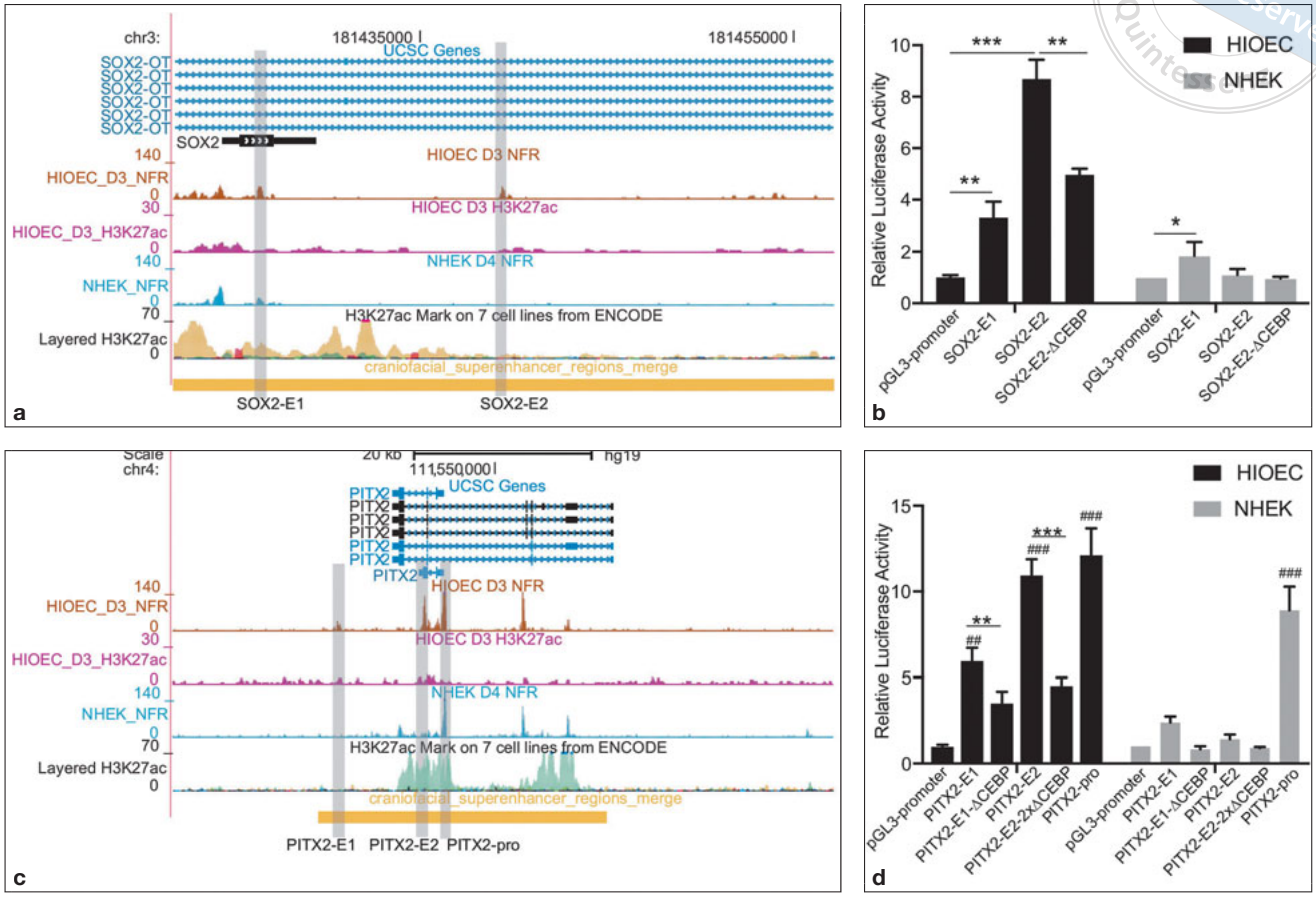


Fig 5 The C/EBP motif is essential for the HIOEC-specific activity of enhancers near SOX2 and PITX2. **(a)** UCSC genome browser tracks showing the genomic location of SOX2-E1 and SOX2-E2 in the SOX2 locus. **(b)** Dual luciferase assay showing the relative enhancer activity of SOX2-E1, SOX2-E2 and SOX2-E2 without the C/EBP motif (SOX2-E2ΔC/EBP) in HIOECs and NHEKs. **(c)** UCSC genome browser tracks showing the genomic location of PITX2-E1, PITX2-E2 and PITX2-pro (promoter of PITX2) in the PITX2 locus. **(d)** Dual luciferase assay showing the relative enhancer activity of PITX2-E1, PITX2-E1 without the C/EBP motif (PITX2-E1ΔC/EBP), PITX2-E2, PITX2-E2 without the C/EBP motif (PITX2-E2xΔC/EBP) and PITX2-pro in HIOECs and NHEKs. Student t tests were performed to evaluate the statistical significance of the difference. *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$. ##: $P < 0.01$ compared with pGL3-promoter group, ###: $P < 0.001$ compared with pGL3-promoter group.

enriched NFRs in SOX2 and PITX2 genes and scanned for IRF and C/EBP motifs using the JASPAR 2016 database. As expected, we found one HIOEC-enriched NFR downstream of SOX2 (SOX2-E2, chr3:181436988-181437368) with the presence of C/EBP motifs with a score of over 9 in the JASPAR 2016 database (Fig 5a). We also chose another region assessable in both NHEKs and HIOECs (SOX2-E1, chr3:181430597-181430850) (Fig 5a). A reporter assay showed that SOX2-E1 were active in both NHEKs and HIOECs, but more so in HIOECs. SOX2-E2 was particularly active in HIOECs, and deletion of C/EBP motifs resulted in significant downregulation of enhancer activity (Fig 5b). For the NFRs near PITX2, we chose three regions near PITX2 (PITX2-E1, -E2 and -pro), among which PITX2-E1 (chr4:111532516-111532960) and PITX2-E2 (chr4:111541723-111542499) were specifically accessible in HIOECs (Fig 5c). PITX2-pro (chr4:111544164-111544829) is the promoter region of PITX2, and no significantly different accessibility was found in this region. Motif scanning showed there was one potential C/EBP motif in PITX2-E1 and two C/EBP motifs in PITX2-E2. A reporter assay showed that, consistent with the ATAC-seq result, PITX2-E1 and PITX2-E2 were only active in HIOECs, while PITX2-pro showed similar activity in HIOECs and NHEKs. Deletion of C/EBP motifs in PITX2-E1 and PITX2-E2 resulted in a significant decrease in enhancer activity in HIOECs (Fig 5d). Notably, all the SOX2 and PITX2 enhancers overlapped with human craniofacial super-enhancers[24]. The expression levels of most of the members of the C/EBP family were compared.

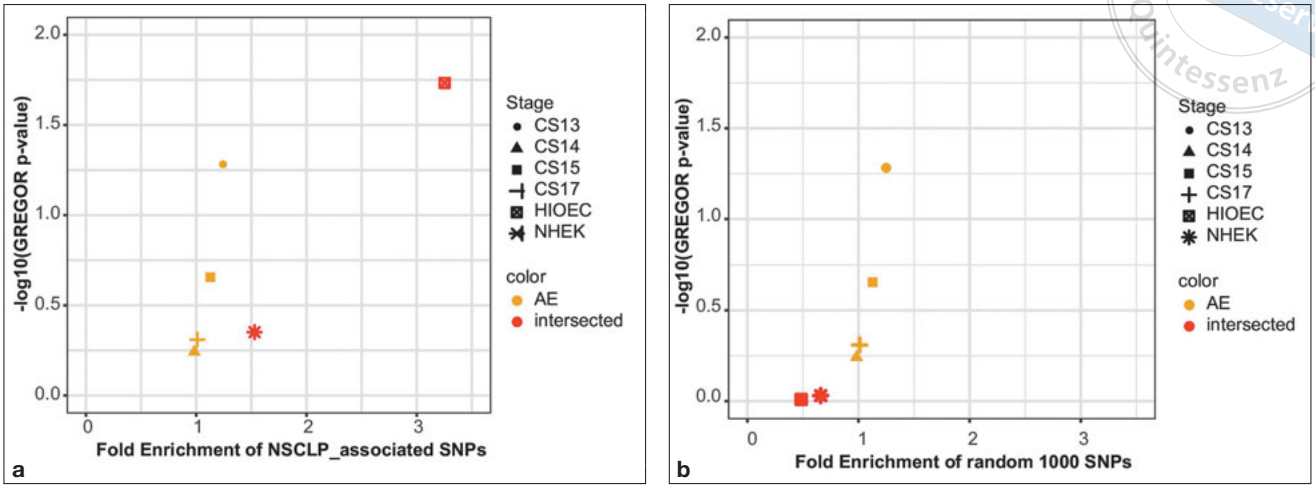


Fig 6 Enrichment of orofacial cleft common variants in HIOEC-overlapped craniofacial super-enhancers. **(a)** Enrichment of orofacial cleft GWAS tag SNPs[26] in previously published human craniofacial active enhancers (AEs) in different embryonic stages and HIOEC-/NHEK-overlapped craniofacial super-enhancers was assessed using GREGOR. **(b)** Enrichment of 1000 randomly selected common SNPs in human craniofacial AEs and HIOEC-/NHEK-overlapped human super-enhancers. The orange colour indicates the AE while the red colour indicates the HIOEC- and NHEK-enriched NFRs overlapped with human craniofacial super-enhancers.

qRT-PCR revealed that except for CEBPE, other C/EBP members shared similar basal RNA expression levels in HIOECs and NHEKs. However, CEBPA and CEBPB were upregulated more dramatically during Ca^{2+} induction of HIOECs compared with NHEKs (data available on request). To sum up, comparison of the NFRs of HIOECs and NHEKs did not reveal the enrichment of SOX2 and PITX2 in HIOECs, but C/EBP, an overrepresented motif in HIOEC, was shown to regulate the expression of SOX2 and PITX2 through binding to the specific active enhancers in HIOECs (oral keratinocytes).

NFRs in HIOECs could be used for the annotation of common variants associated with nonsyndromic cleft lip and palate

In the previous work, we used the epigenome status of NHEKs to annotate the common variants associated with orofacial cleft. However, given the different tissue origin of HIOECs and NHEKs, we asked if NFRs in HIOECs could be more suitable for the annotation of common variants associated with nonsyndromic cleft lip and palate. We first compared the overlapping rates of the NFRs enriched in HIOECs and NHEKs with craniofacial super-enhancers. It was found that, of the 16161 HIOEC NFRs, 6695 were located in the craniofacial super-enhancers (41%), while of the 31696 NHEK NFRs, 8421 overlapped with craniofacial super-enhancers (26%). A GREGOR algorithm[25] was used to assess the enrichment of published orofacial cleft–associated tag single nucleotide polymorphisms (SNPs)[26] using craniofacial active enhancers along with HIOEC and NHEK NFRs intersected with craniofacial super-enhancers. It was also found that, compared with the craniofacial active enhancers in different embryonic stages, orofacial cleft–associated SNPs were more significantly enriched in the HIOEC-intersected craniofacial super-enhancers (Fig 6a). In contrast, we randomly selected 1000 SNPs from the human genome, but did not find such enrichment of these SNPs in the HIOEC-intersected craniofacial super-enhancers (Fig 6b). These results collectively showed that NFRs in HIOECs outperformed the reported epigenome data from other tissues/cells for the annotation of common variants associated with orofacial cleft.

Discussion

Here, we used NFRs as representative open chromatin regions, and identified a set of open chromatin regions active during oral keratinocyte differentiation and made several deductions. One important finding was its use in annotation for craniofacial-related genetic findings. Human embryonic craniofacial cis-regulatory elements[24] could be used to annotate orofacial cleft–associated SNPs, the majority of which are located in the intergenic regions of human genomes. However, due to the mixture of tissues during human embryonic craniofacial development, this database cannot pinpoint in which tissue

these SNPs functionally matter. Here, with the open chromatin regions specifically enriched in HIOEC cells, we revealed a set of craniofacial-derived epithelial enhancers, which can better annotate orofacial cleft–associated SNPs than the published NHEK21 and human craniofacial enhancer datasets.

Combining with H3K27ac ChIP-seq, we found this set of open chromatin regions exhibited cell type–specific enhancer activity. Based on our assessment of the overrepresented TF motifs in different sets of NFRs along with dual luciferase assay, we found CTCF, KLF and GRHL were the three overrepresented TF motifs in the differentiated HIOECs. Interestingly, by comparing the H3K27ac ChIP-seq signals in the CTCF-occupied open chromatin regions and those in KLF- or GRHL-occupied regions, we found that CTCF-occupied regions exhibited less enhancer activity. Since CTCF has long been regarded as an insulator protein[23], blocking "communications" between enhancers and promoters, our comparison indicated that the CTCF-occupied open chromatin regions are likely to be the insulators. However, this has yet to be proved by chromosome conformation capture results in HIOECs.

Previous work using RNA-seq revealed major differentially expressed genes between oral and epidermal keratinocytes. However, the function of these genes has remained uncertain; in particular, the differential cytokine production and lower inflammatory response in oral keratinocytes are fairly descriptive without any mechanism for verification[27]. Given the dominant role of TFs and open chromatin regions in determining cell fate, we sought to uncover the different TFs from the differential cis-regulatory elements, which are the direct targets of TFs. When comparing the open chromatin regions between HIOECs and NHEKs, we did not find that SOX2 or PITX2 were enriched in the HIOEC-enriched open chromatin regions as expected following previous studies. However, the C/EBP TF family appeared to be the motif that distinguished the open chromatin regions in HIOECs from those in NHEKs. Interestingly, several HIOEC-enriched active enhancers near SOX2 and PITX2 were maintained by C/EBP motifs and exhibited HIOEC-specific enhancer activity. These results indicate that C/EBP is a driving TF for SOX2 and PITX2 expression in HIOEC oral keratinocytes.

We notably used HIOEC, an immortalised human oral epithelium cell line, as a standard in vitro model for oral keratinocytes. The results from this standard cell line are highly reproducible but may differ from the real situation in the primary oral keratinocytes collected from individual human samples.

Conclusion

Using high-throughput sequencing, this study identified a set of open chromatin regions and active enhancers that are highly related with oral keratinocyte differentiation and revealed the enriched TF motifs maintaining the cell type–specificity. Moreover, the identification of major TF families in the open chromatin regions between oral and epidermal keratinocytes provide more candidates for improving the skin's wound healing ability. Ultimately, the HIOEC-derived open chromatin regions served as a better craniofacial tissue–specific functional annotation for orofacial cleft common variants.

Conflicts of interest

The authors declare no conflicts of interest related to this study.

Author contribution

Drs Yao XIAO and Huan LIU designed the study and collected and analysed the data for the manuscript; Dr Huan LIU drafted and revised the manuscript.

(Received Sep 17, 2019; accepted Jan 07, 2020)

References

1. Singer AJ, Clark RA. Cutaneous wound healing. N Engl J Med 1999;341:738–746.
2. Martin P. Wound healing--aiming for perfect skin regeneration. Science 1997;276:75–81.
3. Chen L, Arbieva ZH, Guo S, Marucha PT, Mustoe TA, DiPietro LA. Positional differences in the wound transcriptome of skin and oral mucosa. BMC Genomics 2010;11:471.
4. Szpaderska AM, Zuckerman JD, DiPietro LA. Differential injury responses in oral mucosal and cutaneous wounds. J Dent Res 2003;82:621–626.
5. Schrementi ME, Ferreira AM, Zender C, DiPietro LA. Site-specific production of TGF-beta in oral mucosal and cutaneous wounds. Wound Repair Regen 2008;16:80–86.
6. Schmidt R, Plath K. The roles of the reprogramming factors Oct4, Sox2 and Klf4 in resetting the somatic cell epigenome during induced pluripotent stem cell generation. Genome Biol 2012;13:251.
7. Klein RH, Andersen B. Dynamic networking for epidermal differentiation. Dev Cell 2015;32:661–662.
8. Iglesias-Bartolome R, Uchiyama A, Molinolo AA, et al. Transcriptional signature primes human oral mucosa for rapid wound healing. Sci Transl Med 2018;10. pii: eaap8798.
9. Stergachis AB, Neph S, Reynolds A, et al. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell 2013;154:888–903.
10. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317–330.

11. Sdek P, Zhang ZY, Cao J, Pan HY, Chen WT, Zheng JW. Alteration of cell-cycle regulatory proteins in human oral epithelial cells immortalized by HPV16 E6 and E7. Int J Oral Maxillofac Surg 2006;35:653–657.
12. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013;10:1213–1218.
13. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–2120.
14. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–359.
15. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–2079.
16. Ramírez F, Ryan DP, Grüning B, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 2016;44:W160–W165.
17. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9:R137.
18. Bao X, Rubin AJ, Qu K, et al. A novel ATAC-seq approach reveals lineage-specific reinforcement of the open chromatin landscape via cooperation between BAF and p63. Genome Biol 2015;16:284.
19. Ross-Innes CS, Stark R, Teschendorff AE, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 2012;481:389–393.
20. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010;38:576–589.
21. Liu H, Leslie EJ, Carlson JC, et al. Identification of common non-coding variants at 1p22 that are functional for non-syndromic orofacial clefting. Nat Commun 2017;8:14759.
22. Iwafuchi-Doi M, Donahue G, Kakumanu A, et al. The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. Mol Cell 2016;62:79–91.
23. Kim S, Yu NK, Kaang BK. CTCF as a multifunctional protein in genome regulation and gene expression. Exp Mol Med 2015;47:e166.
24. Wilderman A, VanOudenhove J, Kron J, Noonan JP, Cotney J. High-resolution epigenomic atlas of human embryonic craniofacial development. Cell Rep 2018;23:1581–1597.
25. Schmidt EM, Zhang J, Zhou W, et al. GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. Bioinformatics 2015;31:2601–2606.
26. Yu Y, Zuo X, He M, et al. Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. Nat Commun 2017;8:14364.
27. Turabelidze A, Guo S, Chung AY, et al. Intrinsic differences between oral and skin keratinocytes. PLoS One 2014;9:e101480.