A novel dataset for human skin circadian rhythms: Learnings about sources of variation and internal time

Marta del Olmo[1], Sandra Korge[2], Karsten Jürchott[affil?],
Florian Spörl[2,3 – new affil?], Beiersdorf people?[3],
Astrid Grudziecki[2], ..., Hanspeter Herzel[1], Achim Kramer[2],
Bharath Ananthasubramaniam[1]

marta.del-olmo@charite.de
bharath.ananthasubramaniam@hu-berlin.de

Abstract

Skin is the largest organ in the human body and serves as an important protective barrier. With the direct exposure to strong day-dependent changes in the environment, recent work has unraveled an important role of the circadian clock in regulating skin functions. We present in this paper a novel human dataset, obtained through whole-genome microarray analysis of skin punch biopsies, in which we performed a comprehensive analysis of the circadian transcriptome in human dermis and epidermis with respect to internal time. We found that a fifth of of the circadian skin transcriptome is shared between both layers, while a large part is layer-specific. Despite such close physical proximity, amplitude and phase seem to be different in both layers, with epidermis exhibiting higher amplitude rhythms and earlier circadian phases. We quantified the sources of variation and assessed how layers, subjects, time, mid sleep time (our proxy for chronotype and thus internal time) and sex contribute to variability in mean expression (magnitude) or to circadian-specific differences. We found that magnitude of clock-controlled genes varies largely between subjects (and layers), while almost no magnitude- or circadian-specific variance could be attributed to differences in chronotype. Lastly, we identified a set of time-telling genes that can assess circadian phase with high accuracy in human dermis and epidermis. What, in our opinion, distinguishes our approach from previous skin transcriptomic studies in human skin is the fact that we have used mid sleep time to correct the sampling time to internal time, which should be the aim of any study that aims to report about circadian (internal) time.

I. Introduction

The skin is the largest organ of the body and one of its main functions is protection against bacteria, radiation or temperature from the exterior, as well as against water loss from the interior. Morphologically, the skin is complex, populated by many cell types, including epidermal keratinocytes, melanocytes, dermal fibroblasts, adipocytes, neurons, immune cells, and vascular cells, as well as many appendage-specific cell types, including hair follicle keratinocytes, sebocytes, and eccrine gland cells [1]. This heterogeneous population is organized into several structures and compartments to fulfill a range of tasks such as water loss prevention, sensation and hormone synthesis, among others [2, 3]. In humans and mice, the skin consists of three main layers: epidermis, dermis and hypodermis. With the outside world changing throughout the 24 h day and given the direct environmental exposure of the skin, it does not come as a surprise that circadian clocks have evolved to allow the skin to anticipate the environmental shifts and adjust its physiology accordingly. In fact, diurnal rhythms are observed in multiple (if not all) cell types across all layers of skin that regulate a variety of physiological responses.

A full description of the hierarchical or molecular architecture of the mammalian circadian clock is beyond the scope of this paper, but the reader is referred to reviews [4, 5]. In summary, the molecular circadian clockwork consists of a number of auto-regulatory interlocked transcription-translation negative feedback loops. In mammals, the transcription factors CLOCK and BMAL1 induce the expression of their own inhibitors, *PER* and *CRY* genes. When translated, PER and CRY proteins form large complexes that travel back to the nucleus to repress CLOCK and BMAL1, thus repressing their own transcription and thereby creating self-sustained 24 h rhythms in gene expression. A key feature required to generate oscillations is a lag between the transcriptional activation of *PER* and *CRY* genes and the nuclear translocation of the repressor proteins they encode. The nuclear receptors RORs and REVERBs constitute additional transcriptional loops that regulate the expression of *BMAL1*: ROR induces the activation of *BMAL1* expression, while REVERB proteins repress it. By acting at genomic regulatory sequences, this core clock network generates rhythmic oscillations in the expression of a large number of output genes (almost 10% of all genes!) in a cell-autonomous and tissue-specific manner [6–12].

At least 1400 genes involved in different functions show circadian expression changes in mouse skin [1], suggesting that the circadian clock may, indeed, influence various aspects of skin physiology, including susceptibility to UV-induced DNA damage [13, 14], barrier recovery [15], trans-epidermal water loss [16], sebum secretion, skin temperature or skin pH [17]. While it is known that the central clock influences circadian rhythms within skin [18], evidence in the last years has shown that clock regulation in skin is not just an output of the central suprachiasmatic nucleus (SCN), but rather, skin itself, like most organs, harbors robust extrinsic clocks. Already more than 10 years ago, circadian oscillations were found to be present in several skin cell types, including epidermal and hair follicle keratinocytes, dermal fibroblasts and melanocytes [19–25]. Nevertheless, on a molecular level, it is still unclear what are the differences between clocks in different skin layers and how such clocks might contribute to rhythmic skin function.

But skin, besides providing additional knowledge into its circadian biology, also represents a potential source for circadian biomarker discovery. Around 50% of all drugs target the product of a circadian gene [26]. Moreover, therapeutic outcomes such as survival after surgical procedures [27], efficacy and tolerance of chemotherapy [28] or antibody response to vaccination [29] all vary diurnally. For this reason, a practical measure for circadian phase is needed if we want circadian medicine to influence health in any way. A key limitation in the implementation of chronotherapeutic approaches is the fact that humans are heterogeneous with respect to the timing of their internal clocks [30]. Humans exhibit different phases of entrainment or chronotypes. This is, the alignment phase angle of one's physiological and behavioral rhythms with respect to the environmental changes from individual to individual and it shows a normal distribution ranging from very early chronotypes (larks) to very late chronotypes (owls) [30]. The human chronotype is usually assessed by questionnaires that examine sleeping habits [31, 32], which are normally not objective, or with strategies that require multiple measurements under controlled conditions. The current cold standard tool for assessing human circadian phase is the dim-light melatonin onset (DLMO) assay, which requires a subject to sit in a dim room for repeated saliva sample collection, a difficult practice to standardize and perform at large scales and burdensome for clinical practice. DLMO is considered the marker of SCN phase and locomotor activity [33, 34]. This raises several important and open questions: Is the DLMO phase aligned with peripheral clocks? Are there better sources of circadian biomarkers? What should a good biomarker measure?

An individual's (or a tissue's) circadian time refers to the phase of its *internal* biological clocks.

This inner phase of entrainment (or chronotype) depends on many factors. It has a genetic basis [24, 35], it is age- and sex- dependent [30], depends on the light exposure level [36, 37], on the season [36, 38] and on the time-zone where that subject is located [30]. Thus, any algorithm that aims to assess the phase of circadian biomarkers should take these factors into account in order to correct external time (i.e. time of sampling) to *internal* circadian time [39]. For this reason, when aiming at determining circadian time in a specific human tissue, it is important to obtain a dataset in which not only the time of sampling is recorded, but also as much information from the subjects as possible. This way we might be able to more accurately control and correct wall (external) time to *internal* time.

Here, we present a novel high-resolution human dataset obtained through whole-genome microarray analysis of suction-blister skin, where 5 females and 6 males were taken a skin biopsy every 4 h for 24 h. In order to control for as many external factors as possible, we gathered meta-data about the subjects including time at which they go to bed during weekdays and weekends. We calculated their mid sleep time (MST) and used this value to correct wall time to internal time. Thus, our results report about clock-controlled gene expression in human dermis and epidermis with respect to internal time, with differences in chronotypes being taken into consideration. We found ∼1400 rhythmic transcripts in at least one of the layers and that ∼280 transcripts are shared between dermis and epidermis, although with some layer-specific differences despite their physical proximity: epidermis displays higher amplitude rhythms and earlier phases compared to dermis. Because of the meta-data availability, we quantified the sources of variation and assessed how layers, subjects, time, mid sleep time (our proxy for chronotype) and sex contribute to variability in mean expression or to circadian-specific differences. We found that magnitude of clock-controlled genes varies largely between subjects (and layers), while almost no magnitude- or circadian-specific variance could be attributed to differences in chronotype. Lastly, we identified a set of time-telling genes that can assess circadian phase with high accuracy in human dermis and epidermis. The novelty of our approach, in our opinion, is the fact that we corrected sampling time to internal time. It is this what should be taken into account if any evaluation of circadian phase of biomarkers is desired, as circadian time refers to the phase of *internal* biological clocks.

II. Results

Inner rhythms in human skin layers

To explore molecular rhythms in human dermis and epidermis, 3 mm biopsies were obtained from the upper back of 11 healthy volunteers every 4 h across a 24 h sampling period as described in Materials and Methods. The skin transcriptome was assessed through whole-genome microarray analysis (Figure 1A). In order to control for chronotype differences, we calculated the mid sleep time (MST) of each individual from their sleeping schedules (available in Supplementary Table S1) to correct external (wall) time to *internal* time. The MST was adjusted for the sleep-debt accumulated during the workweek as suggested by [40]. For each subject, internal time was determined by subtracting wall time minus the difference of his/her MST to a reference subject (the individual with median MST).

To identify the clock-regulated transcriptome in both human skin layers, we performed cosinor analysis and fitted the expression values versus internal time to 24 h sine and cosine curves. We identified 523 circadian (FDR < 0.05 and relative amplitude > 0.26) genes in dermis and 1191 in epidermis, from which 283 genes appeared rhythmic in both skin layers (Figure 1C, inset).

Moreover, we observed a relatively stable number of rhythmic genes independently of the exact FDR cutoff (Figure 1C). Of note, controlling for MST did not affect determination of rhythmic genes: the same cosinor analysis done against wall time did not produce major differences in the number of rhythmic genes, their phases or amplitudes (Supplementary Figure S1). This might be due to the range of MST, which is in the order of the sampling frequency, or because of additional sources of inter-individual variation.

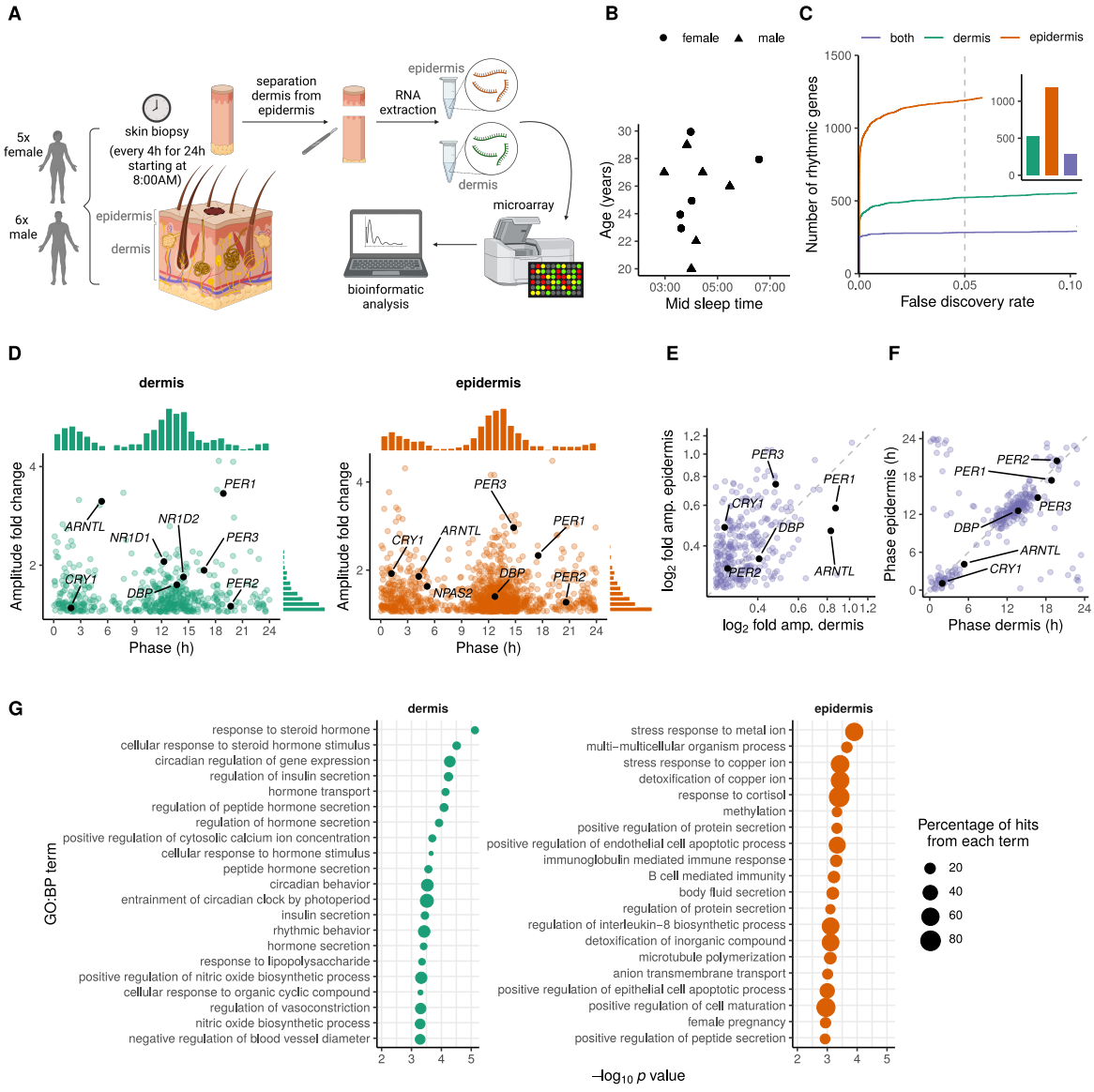


Figure 1: Functional clocks in human dermis and epidermis. A. Experimental setup: the dataset includes dermal and epidermal samples collected from the back of 11 healthy subjects (5 females, 6 males). Punch biopsies were collected every 4 h for 24 h starting at 8 AM. Dermis and epidermis were separated and gene expression was analyzed using microarrays. **B.** Composition of the study cohort by sex, age and mid sleep time. **C.** Number of circadian transcripts as a function of the false discovery rate (FDR). Rhythmic transcripts with respect to *internal* time in dermis, epidermis or in both layers were determined by cosinor analysis (relative amplitude > 0.26). For FDR= 0.05 (inset), 523 transcripts were found to oscillate with a circadian period in dermis, 1191 in epidermis and 283 were common in both layers. **D.** Acrophase and amplitude distributions of the 24 h cycling transcripts in human dermis (in green, left panel) and epidermis (orange, right panel) (FDR< 0.05, relative amplitude > 0.26). Each transcript is represented by a dot; clock genes are highlighted in black. **E.** Amplitude correlation of cycling transcripts in dermis *and* epidermis. **F.** Phase correlation of cycling transcripts in dermis *and* epidermis. **G.** Circadian GO enrichment analysis of the rhythmic genes in dermis (green) and epidermis (orange). The top 20 enriched biological processes (with a minimum gene set of 5 terms from each category) in each layer are shown.

We observed a bimodal distribution of phases of all rhythmic dermal and epidermal transcripts, with peaks clustering at 1-2 AM and 1-2 PM (Figure 1D and Supplementary Figure S2A, in contrast to previous studies that have reported phases clustered at 8-9 AM and 8-9 PM [41, 42]). Despite these similarities, we found that clock amplitude and phase varies between skin layers: out of the rhythmic transcripts in both layers, epidermal transcripts oscillate with a higher amplitude than those in dermis (Figure 1E, also evident in Figure 1D), although *PER1*, *ARNTL* and *DBP* show a larger amplitude in dermis (see also Supplementary Figure S2B). Moreover, from the common rhythmic genes, we found that dermis is delayed 1-2 h with respect to epidermis (Figure 1F). These results suggest an interesting amplitude and phase specificity among genes that are rhythmic in both layers, despite being physically so closely located in skin.

Dermal rhythmic clock-controlled genes were enriched in rhythmic processes, hormone-related terms and blood vessel related terms, whereas the epidermis circadian transcriptome was enriched in terms related to the response to metals, immunity and body fluid and protein secretion (Figure 1G). We also performed KEGG pathway enrichment analysis and found that viral-, parasite- and bacterial-related pathways appeared in dermis; on the other hand, the epidermal circadian transcriptome was enriched in pathways associated to absorption and phototransduction (Supplementary Figure S2C). Morning-time genes were enriched for immune-related pathways together with phosphorylation and metabolic signaling, whereas the evening was marked by genes involved in carbohydrate, nitrogen and phenol metabolism (Supplementary Figure S2D and E, as determined by phase set enrichment analysis (PSEA) [43]).

Sources of variation in skin clock-controlled genes

In order to translate insights from controlled and rather small-sized cohorts to large populations, it is of vital importance to know what the magnitudes of the different sources of variation are and how they compare across skin layers, subjects, etc. To study the different drivers of variability in the rhythmic skin transcriptome we used the open-source `variancePartition` software, which fits a linear mixed model for each gene and partitions the total variance into fractions correspondent to each category of the experimental design, plus a residual variation [44].

We wanted to understand what the sources of variation in the rhythmic skin transcriptome might be, for which we considered the 1410 genes that we found rhythmic in at least one of the two skin layers from the 11 individuals. We reasoned that, for a single gene, the total variance can be partitioned into the individual contributions of subject, sex, skin layer, time and chronotype plus a residual variance. But this assumption relies on the fact that the effect of each component of variation on the gene expression does not depend on other variables in the model. To relax this strict assumption, we modeled also interaction effects whereby the effect of time depended also on subject and skin layer. Thus we ended up dividing the variability of the data in seven sources:

- inter-sex mean variation (variability in mean expression, i.e., magnitude only due to differences between sexes),
- inter-chronotype mean variation (variability in magnitude due to differences in chronotypes)
- inter-layer mean variation (variability in magnitude only due to differences in skin layers),
- inter-subject mean variation (variance in mean expression only due to differences between subjects),
- common circadian variation (variation in magnitude attributed solely to the different time points but not specific to layers or subjects),
- inter-layer circadian variation (genes that show different time (circadian) variation across the two layers) and
- inter-subject circadian variation (rhythmic genes that show differences between subjects).

Of note, the discrete *external* time was used in this analysis since (i) chronotype was taken into account as an additional variable and (ii) correcting wall time to internal time produced a continuous variable, which cannot be modeled as a random effect with variancePartition. Is this really a reason?

Applying variancePartition to this data illustrates how the method can decouple biological variation into multiple components, including a residual variation which remains uncharacterized. Results from representative genes show how variancePartition identifies genes where the majority of variation in mean expression is explained by a single variable such as skin layer (e.g., *SLIT3*, Figure 2A, left panel), while variation in other genes is driven by multiple variables (e.g. *ZNF436*, Figure 2A, right panel). These results, illustrated in a rhythmic skin transcriptome-wide manner, show that variation across skin layers represents the major source of magnitude variability and explains a median of 36.4% of the variance from the rhythmic skin transcriptome (Figure 2B). The median variance explained by subject (9.1%), common circadian variation (6.1%), inter-layer circadian variation (1.4%) and inter-subject circadian variation (0.9%) are smaller, but with a high unidentified residual variation (27.3%). Moreover, we found that neither chronotype nor sex contribute to variability in mean expression of rhythmic genes, as `variancePartition` attributed very little variance to these components (< 0.5%, data not shown).

Of particular interest here were two things: firstly, the fact that subjects differ more strongly in mean expression than in circadian rhythms (compare blue violin plot to green violin plot from Figure 2B). This means that although rhythms (phases?) of clock controlled genes are not very subject-dependent, their magnitudes are. Secondly, the observation that time variation does not seem to be layer- or subject-specific, since the inter-subject and inter-layer circadian variation appeared smaller than the common circadian variability. In other words, time alone can explain variation without being dependent on layer or subject in our cohort. This result, together with the almost nonexistent contribution of chronotype to variability and the findings from Supplementary Figure S1, speaks for population sampling from skin providing a good estimation of circadian time,

regardless of wall time or internal time. As negative and positive control of the variancePartition analyses, we checked that the time variation across 1000 non-rhythmic genes (FDR > 0.1) was almost 0 (Supplementary Figure S3A), while it represented the largest source of variation for clock genes (Supplementary Figure S3B).

We performed GO analysis in top 200 most variable genes from each category and found that wound healing processes were enriched in those with high variation in mean expression across layers (Supplementary Figure S3C). On the other hand, genes with highest circadian variation across skin layers (orange category from Figure 2A) were enriched in pyruvate metabolism, phos-

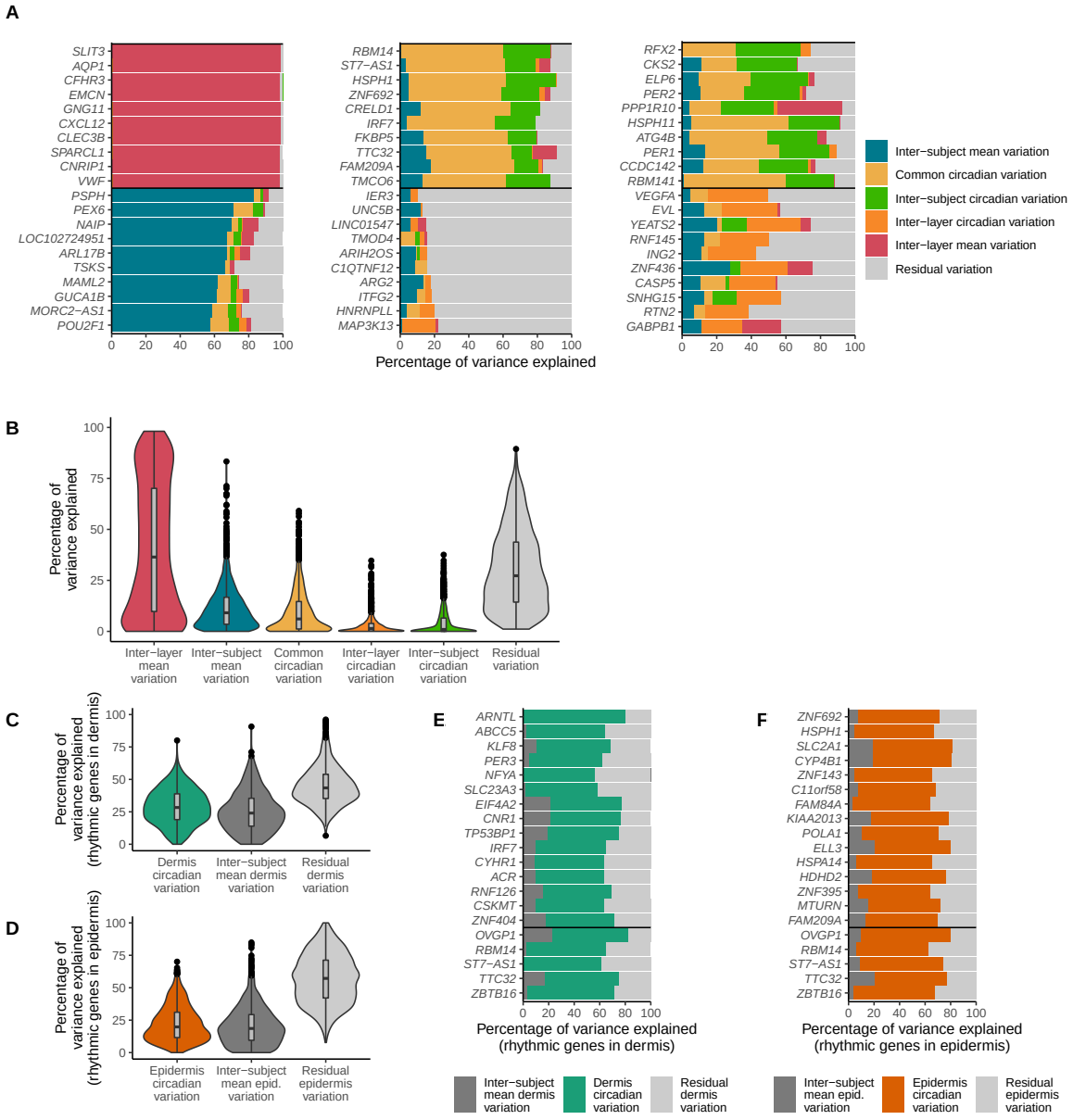


Figure 2: Drivers of variation in the human circadian skin transcriptome identified with variancePartition. A. Top 10 genes with highest variance in mean expression across layers (top left panel), across subjects (bottom left panel), across time (top middle panel) or with highest residual (unidentified) variation (bottom middle panel). The panels in the right show the top 10 genes with highest differences in circadian expression across subjects (i.e., rhythmic transcripts where rhythms seem to be subject-specific, top panel) or with highest differences in circadian expression across skin layers (i.e., rhythmic transcripts where rhythms seem to be layer-specific, bottom panel). **B.** Quantification of the contribution of each meta-data variable to the variation in expression of each gene in a circadian transcriptome-wide trend. To plot panels A and B, variancePartition was run with the ~ 1400 genes that are rhythmic in *at least* one layer and with external time as a meta-data variable. **C** and **D.** Contribution of each variable to variation in mean expression in the rhythmic genes in dermis (C) and epidermis (D). **E** and **F.** 20 genes with highest variation across time in dermis (E) and epidermis (F). Common time-varying genes across layers are depicted below the black line in panels E and F. variancePartition was run with the 523 rhythmic genes in dermis to plot panels C and E and with the 1191 rhythmic genes in epidermis to plot panels D and F.

phorylation and ATP generation processes (Supplementary Figure S3D). As for the genes with highest (unidentified) residual variation, we found that they were enriched in processes related to cytoskeleton and stimuli detection (Supplementary Figure S3E).

Performing `variancePartition` on the epidermal and dermal rhythmic genes separately hinted to what the drivers of variation in each layer separately are. Common circadian variation exceeded the inter-subject mean variation in gene expression in both skin layers (although the residual variation, in either case, was found to be larger than the other sources of variation, Figure 2C, D). The clock genes *ARNTL*, *PER3* were found among the top 20 common circadian varying genes in dermis, but not in epidermis (Figure 2E). Some genes appeared to have a high variability across time in both layers: *OVGP1*, *RBM13*, *TTC32* or *ZBTB16* (shown in the bottom of Figure 2E, below the black line).

Of note, the variance in mean expression explained by skin layers represents also the highest source of variation in previously published skin circadian transcriptomic studies [45, 46], and these also show high residual variability (data not shown). This suggests that the learnings from our dataset can be translated to larger populations. Interestingly, the quantification of variation in our data is robust to sampling frequency. This is, we found similar contributions of each biological source when the time series of rhythmic genes was made sparser by removing time points (data not shown).

Predictive biomarkers of internal time in human dermis and epidermis

In addition to analyzing the skin biology controlled by the circadian clock or quantifying the sources of variation, this dataset also provides a good opportunity to discover (and compare to previously published) robust markers of circadian phase. From the different bioinformatic approaches that have been proposed to obtain predictors that associate gene signatures to time, we applied `ZeitZeiger` [47] to our microarray gene expression data. We tested 3 sets of predictors that differed on the dataset used (the whole set of expressed genes in at least one layer versus expressed genes in dermis and epidermis separately). The predicted variable was, in all cases, internal time. Details

on the ZeitZeiger implementation and the main parameters it uses are provided in Materials and Methods.

To measure the accuracy of the prediction, `ZeitZeiger` uses the median absolute error (MAE): the lower this value, the better the prediction. Running `ZeitZeiger` in the whole set of ~ 11000 skin-expressed genes (in either dermis *or* epidermis) resulted in a minimum MAE of ~ 0.08 for the software parameters `nSPC=2` and `sumabsv=3` (Supplementary Figure S4A), with 38 genes needed for prediction (Supplementary Figure S4B). On the other hand, running `ZeitZeiger` in each skin layer separately performed better, as seen by the lower MAE and the lower number of genes needed to predict internal time: 2 SPCs and `sumabsv=2` resulted in a MAE of 0.04 in both layers (Figure 3A) and a total of 15-25 genes needed for prediction (Figure 3B, time series in Supplementary Figure S5).

We focused on the internal time predictors from dermis and epidermis separately due to their better performance in terms of accuracy of the prediction. In both layers, we observed a high proportion of the time-telling genes overlapping with the top 20 time-varying genes from the `variancePartition` analysis (highlighted in yellow in Figure3B). (*OVGP1* was not only a highly time-varying gene, but also one of the top 20 genes that have different rhythms in epidermis compared to dermis (orange category from Figure 2), reason why it is marked in orange.) In fact, when doing `variancePartition` only on the `ZeitZeiger` genes, we found circadian variance to represent the major driver of variation in both skin layers, while variation attributed to subjects is very minor (Supplementary Figure S5B, also evident from the time series in Supplementary Figure S5A). This illustrates that the biomarkers of circadian phase show little differences in mean expression (i.e., magnitude) across subjects, as would be expected from a good biomarker. Interestingly, when time-telling dermal and epidermal samples were each graphically represented in SPC space (SPC2 versus SPC1), the points described a cycle for which the progression of internal time followed an counter-clockwise trajectory in both skin layers (Figure 3C). These results imply that the circadian clock can be reasonably well approximated by a two-dimensional oscillator. Moreover, this cyclical behavior in SPC space was observed for each individual subject (Supplementary Figure S5C), suggesting that such an approach may be used to detect perturbations of the skin clock in humans.

Our results indicate that dermis and epidermis time series data are well suited to extract time-telling genes, and that internal cross-validation performance is similar to that of previous studies [41, 42] (I'm not sure if sfig7 and sfig8 really support this...). Moreover, integrated with prior published circadian skin transcriptomic studies, these results provide robustness, as some of the biomarker genes that we have found have already been described as skin phase-telling genes (*ZBTB16*, *FKBP5*, *TRIM35*, *PER3*, *ARNTL*) [41, 42]. Moreover, although the common circadian variability exceeds the inter-subject and inter-layer circadian variation (Figure 2A), and thus one could expect to find good time-telling gene sets even across layers, our results show that predicting internal phase in dermis and epidermis separately performs better than predicting internal time in skin as a whole.

Of note, ZeitZeiger was also used to predict external (wall) time. Although the accuracy of wall time predictors was comparable to those of internal time, the value of `sumabsv` was higher, thus resulting in a larger number of genes needed to predict wall time in both, skin as a whole, and in dermis versus epidermis separately (data not shown).

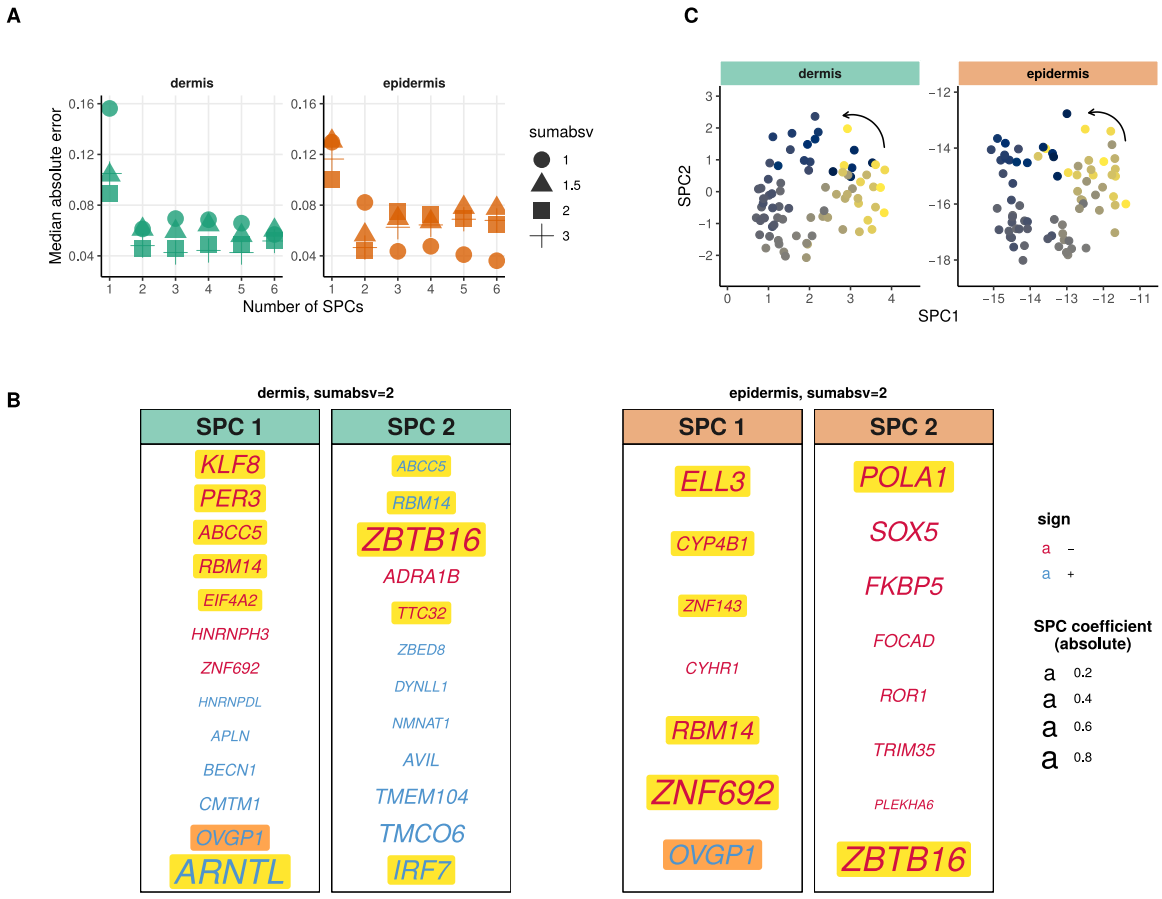


Figure 3: Identification of internal time-telling genes in human dermis and epidermis with `ZeitZeiger`. **A.** Median absolute error of the internal-time prediction on cross-validation (see Materials and Methods for details) as a function of the two main parameters of ZeitZeiger, `sumabsv` and nSPC. **B.** Internal time predictors from human dermis (left panels) and epidermis (right panels) for `sumabsv`=2 and nSPC=2. Genes assigned to SPC1 or SPC2 as well as their coefficients are shown. Highlighted in yellow are genes that appeared in the top 20 most common circadian varying genes in the variancePartition analysis done in dermis and epidermis separately; in orange, genes that showed differential rhythms across layers (i.e., genes with high inter-layer circadian variation from the variancePartition analysis). **C.** Expression profiles of our cohort in dermis (left) and epidermis (right) represented in SPC space. Colors indicate the internal time of the subjects. ZeitZeiger was run with all ~ 11000 expressed genes and separately for dermis and epidermis.

III. Discussion

Here we report a novel human dataset in which we examined circadian gene expression of skin in 11 subjects that were sampled longitudinally every 4 h for 24 h. Although skin circadian rhythms have been analyzed in previous studies in a genome-wide manner [41, 42], this dataset has a higher sampling frequency and contains information about the subjects, what allows to control for external factors and to assess what the sources of variation in such a human study might be. We identified the circadian transcriptome of human dermis and epidermis with respect to *internal* time

by correcting external time to each individual's mid sleep time. We found that roughly a fifth of the circadian skin transcriptome (~280 out of 1400 rhythmic genes) is shared between both layers, while a large part is layer-specific. Interestingly, and despite their physical proximity, we found amplitude and phase differences in the core clock and clock-controlled genes in both layers. In general, epidermis is earlier in phase and displays higher amplitude rhythms (Figures 1D-F), with some exceptions like *ARNTL*, *DBP* or *PER1* that oscillate with higher amplitudes in dermis.

Circadian timing mechanisms are sensitive to day length and temperature, and therefore the clock-controlled transcriptome represents a candidate for the regulation of seasonal phenomena within the skin. Interestingly, several skin diseases have been shown to exhibit seasonal change in severity [48]. Nevertheless, given the cellular complexity and heterogeneity of human skin, it is probably simplistic to talk about a singular "skin clock". Our results show that epidermis presents, overall, larger amplitude rhythms compared to dermis. We can only speculate that this might be because of the direct environmental exposure of the epidermis to external Zeitgebers. Could it be that epidermal clocks entrain more efficiently and display higher amplitude resonance, resulting in the observed higher amplitude rhythms? Do amplitude rhythms in skin change with seasons? Skin cancer progression has been shown to be under clock control [49] and possibly affected by feeding schedules, as suggested by a study from 2017, where the authors proposed that time restricted feeding influences sensitivity to UV-induced DNA damage [14]. Moreover, the amount of caloric intake affects gene expression and function of the skin [50]. Could it be that skin clocks contain potential information about circadian rhythmicity in additional organs, or vice versa? All these remain important and open questions.

A fundamental challenge in the analysis of high-throughput datasets is to quantify and interpret the contribution of different sources of variation. How does the population or how do the different skin layers affect the genetic regulation of rhythmic gene expression? What are the major drivers of variability? Are there rhythm-specific differences across subjects or skin layers? This set, in which meta-data was available (sex, mid sleep time, age, etc.), allowed to study these kind of questions. We used `variancePartition`, a publicly available software that leverages the power of linear mixed models, to partition and quantify the contribution of each meta-data variable in the experimental design, plus a residual variance. In our case, we took the rhythmic skin transcripts and analyzed variability in mean expression *across* time, *across* individuals and *across* skin layers as well as the contribution of variance *within* layers and individuals. We defined *across*-layer (or -subject or -time) variability as the variance in mean expression (i.e., magnitude) between dermis and epidermis (inter-layer mean variation in Figure 2). *Within*-layer (or -subject) variability was defined as the variability in rhythms between layers, i.e., the inter-layer circadian variation.

We found that differences in magnitude across layers (inter-layer mean variation in Figure 2) represent the strongest source of variance in our cohort, followed by differences in mean expression across subjects and across time. Moreover, when analyzing previously published circadian skin transcriptomic datasets (GSE139300 [46] and GSE112660 [45]) we observed that the design variables that contribute most to magnitude variability are also differences between layers, followed by subjects, with large residual variance (not shown maybe remove from results?). Taken together, these insights suggest (i) that the magnitude of clock-controlled genes varies largely between skin layers and subjects, a novel observation which has, to our knowledge, not been reported previously; and (ii) that such observations are also the case in other population studies and thus might be translatable to larger human cohorts.

Although not included in these results, we also assessed whether MST (proxy for chronotype) and sex contributed to magnitude variations in our data, but found almost no contribution. In principle, this speaks for population sampling from skin providing a good estimation of circadian time, regardless of wall time or internal time. Nevertheless, the fact that we observed almost no variance due to chronotype differences might have to do with the fact that the range of MST in our cohort (~4 h, Figure 1B, consistent with previous studies [30]) is in the order of our sampling time. Thus, this experimental design might not be enough to capture subtle chronotype differences and could be a reason explaining why the analyses of internal and external time yield similar results (Suppplementary Figure S1). To discern whether the similarity in the internal versus external time analyses of the skin circadian transcriptome is a general property in any cohort or something particular from this dataset, an ideal population study should be sampled at a higher frequency but this poses a complication for obvious practical reasons.

Over the last years, a number of novel approaches have been introduced to assess circadian parameters and, in particular, circadian phase in humans (see [51] for a nice review). Some of these methodologies use machine-learning approaches to extract features that predict circadian parameters from high-dimensional -omics data, whereas others are based on collection of data from wearable devices. In this work, we propose a viable set of biomarkers that can accurately predict molecular clock phase in human skin. But, what makes a biomarker a good biomarker? First, a biomarker should be stable and show little variation in magnitude (mean levels) across subjects. Second, it should show a strong temporal variation (that might be specific to the tissue of interest, but not necessarily). It should have relatively strong amplitude rhythms as well as the correct phase. We have shown that a few transcriptome samples contain enough information to predict the phase of *internal* skin rhythms with a small set of candidate genes, whose mean levels are roughly invariant across subjects while showing strong temporal variation (Supplementary Figure S5B). We have shown that internal time is predicted more accurately if it is done in both layers separately, as seen by the lower prediction error in Figure 3A compared to Supplementary Figure S4A. We have achieved an accuracy (MAE~0.04 h) similar to that of the current gold-standard, DLMO (0.5-1 h [52,53]). Nevertheless, the fact that circadian phase is predicted more accurately in human dermis and epidermis separately unfortunately represents an experimental limitation, since it means that layers must be separated prior to analysis. It should be noted, however, that despite the better performance when done separately, none of the time-telling genes were dermis- or epidermis-specific and just one, *OVGP*, was among the top genes with layer-specific rhythms.

We have argued that predicting *internal* time is more appropriate if any evaluation of circadian phase of biomarkers is desired; nevertheless, there are little differences between the rhythmic genes (and their phase and amplitude) obtained when analyzing rhythmicity with respect to internal or external time (Suppplementary Figure S1). Thus, we in principle expect the same biomarker set to properly predict internal time in any cohort of individuals even if their chronotype information remains unknown. In this regard, firstly, we observed that the set of biomarkers and their coefficients did not change significantly when ZeitZeiger was run with external instead of internal time (data not shown). Secondly, among our time-telling genes, we found some that are in agreement with previous transcriptomic studies done in human skin with no information about chronotype (*ZBTB16*, *FKBP5*, *TRIM35*, *PER3* and *ARNTL*) [41,42]. These studies used a hybrid experimental design in which they combined data from human subjects that were sampled throughout the 24 h cycle and data from larger population that were sampled just once. The way in which the authors ordered in time and assigned a phase to the samples that were taken just once takes into account internal time. Nevertheless, no chronotype information was available for the longitudinal data,

and thus whether the time-telling biomarkers predict internal or wall time was not clear. The authors also validated their biomarker set against the clock time at which the sample was taken, which is not a marker of internal phase. The silver lining of this confusion between internal and external time is that our results suggest that chronotype does not seem to affect the identification or performance of the biomarker set.

Although chronotype is not relevant to predict skin phase in our cohort, we cannot extrapolate and apply the same rule for other tissues, or even in skin from extreme chronotypes. Archer et al. previously showed in [54] that when sleep was scheduled out of circadian phase, the blood transcriptome was affected resulting in lower amplitude circadian, or even arrhythmic transcripts. This means that if a biomarker set includes such genes, it may fail to predict circadian phase in individuals that sleep out of their phase such as shift workers, the elderly and/or sick patients. Whether our proposed biomarker set is able to properly predict internal time in patients or cohorts with skin diseases, with known low circadian amplitudes, in extreme chronotypes and/or in the presence of internal circadian desynchronization remains to be elucidated. Additional unanswered questions, although out of the scope of this paper, are the elucidation of methods that might help to accurately and unobtrusively assess not just phase, but also amplitude, robustness and circadian disruption in order to understand the role of circadian rhythms in physical and mental health and disease. But unfortunately theconcepts of circadian amplitude or robustness are not well defined [51], and a simple gold standard measure for circadian amplitude and robustness has not yet been agreed upon.

Any clinical use of circadian biomarkers needs fast, cost-effective and noninvasive methods such as hair follicle- or oral mucosa-sampling techniques, standardization across platforms (qPCR, NanoString, a limitation in this study) and ideally samples taken just once. Moreover, any algorithm assessing circadian phase should be validated against gold-standard markers of internal phase. For example, the phase of melatonin rhythm is considered the gold-standard marker for the phase of the SCN [33, 34]. In order to unlock the future of circadian medicine and chronotherapy, larger studies should compare their findings with other techniques such as DLMO, actigraphy, etc. and confirm their results across a range of disease states and pathologies. Solving questions like these will undoubtedly help in defining good sources of biomarkers for application in circadian medicine.

IV. Conclusions

In summary, we have identified the circadian transcriptome in human dermis and epidermis and reported in differences between these two closely located layers. We have quantified the variation in mean expression of rhythmic genes, and the circadian-specific variation across subjects and skin layers, and found that clock-controlled gene magnitude varies largely between subjects. Lastly, we have identified a set of 15 and 25 genes that can assess circadian phase with high accuracy in human epidermis and dermis, respectively. What, in our opinion, distinguishes our approach from previous clock-controlled gene expression studies in human skin is the fact that we have used MST to correct the sampling time, which should be the goal of human studies that aim to predict circadian (internal) time.

V. Materials and Methods

Experimental design and collection of human skin samples

The study to obtain skin punch biopsies was approved by the Beiersdorf AG Legal Review Board. Tissue samples were collected according to the recommendations of the Declaration of Helsinki and according to applicable laws for a non-drug study. All donors provided written and informed consent. The study was performed at the study center of Beiersdorf AG. Eleven healthy volunteers (six males, five females, aged 20-30 years) participated in the study. Individual chronotypes were assessed using the Munich Chronotype Questionnaire (MCTQ) by calculating the *corrected* mid sleep time, this is, the mid sleep time on free days adjusted for the sleep-debt accumulated during the workweek [40]. meals? physical activity? All information about the subjects is provided in Supplementary Table S1.

3 mm punch biopsies were obtained from the upper back for seven time points over a period of 24 h (8 AM, 12 PM, 4 PM, 8 PM, 12 AM, 4 AM and 8 AM the following day). Skin biopsies were subsequently incubated in PBS at 55 °C for 3 min to separate epidermis and dermis. Tissue samples were then frozen in liquid nitrogen and stored at -80 °C. RNA extraction and quality control from punch-biopsies was performed by Miltenyi Biotec using the TRIzol method. Linear amplification and labeling of RNA and hybridization of Agilent Whole Genome Oligo Microarrays 4x44k (Agilent Technologies) using 1.2–1.65 µg of Cy3-labeled cRNA was performed by Miltenyi Biotec, essentially as reported in [55].

Gene expression analysis

The microarray gene expression analysis was conducted in R. The RMA (Robust Multichip Average) algorithm was used to pre-process and extract expression profiles from the raw CEL files. Transcripts were annotated with ENSEMBL and ENTREZ IDs using Agilent "Human Genome, Whole" annotation data (hgug4112a.db, v3.2). The raw gene expression data has been deposited in the Gene Expression Omnibus (GEO) under the accession number GSEblabla.

Raw data of the hybridized microarrays were normalized using the Bioconductor R-Project package Linear Models for Microarray Data (`limma`). Background correction was performed using the `normexp` function; normalization between the different arrays was done with the `quantile` function. Non-annotated probes as well as control probes were filtered out. Genes were considered to be expressed if the signal across the 7 time points was well above the background signal in at least 50% of the microarrays. Lowly-expressed genes were removed from further analysis. The expression of genes annotated by multiple ENSEMBL IDs was averaged across replicates using the `avereps` function from `limma`, resulting in a filtered dataset containing data for 11578 transcripts.

Principal Component Analysis (PCA) was performed in order to remove outliers. The expressed genes nicely organized in two clusters in PCA space, separated by skin layer (data not shown). Nevertheless, the epidermis sample from subject 109 taken at 8 AM the following day did not cluster with the rest of epidermal samples and for this reason was removed from further analyses.

Rhythmicity analysis and functional annotation of circadian gene lists

To detect transcripts exhibiting rhythmic behavior with a 24 h period in their expression, successive filtering steps were applied. For detection of circadian rhythmicity, we applied cosinor analysis and

we tested only *one* null hypothesis (in the lines of [56]), namely that rhythms in dermis and epidermis are the same. Each microarray was assigned to its corresponding skin layer and time-group (i.e., 8 AM, 12 PM, 4 PM, 8 PM, 12 AM, 4 AM or 8 AM the following day). In the analyses where *internal* time was used, sampling (wall) time was corrected to internal time in each subject by calculating the difference in mid sleep time (MST) to a reference subject (that with median MST), and subtracting this value to wall time. A linear model was then fitted to the expression data for each probe using the `lmFit` function from the `limma` package. Subsequently, a moderated F-test of the empirical Bayes statistics for differential expression (eBayes function from the `limma` package) was applied to test differential gene expression between the respective groups for statistical significance. The `topTable` function from `limma` was used to summarize the linear model fit object produced by `lmFit` and processed by eBayes.

Acrophases and relative amplitudes were estimated from the cosinor analysis. In order to identify significant circadian transcripts, p values were adjusted for multiple testing by the Benjamini-Hochberg (BH) method, thus controlling the expected false discovery rate (FDR). An FDR cutoff below 0.05 and a relative amplitude threshold of 0.26 were used to filter for expressed trancripts with significant rhythms in dermis or epidermis.

Genes showing significant diurnal expression patterns were tested for over-representation in Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways against the background of expressed transcripts in each skin layer (unless otherwise stated) by applying the `enrichGO` and `enrichKEGG` functions included in the `clusterProfiler` package.

Circadian GO terms and KEGG pathways were determined by Phase Set Enrichment Analysis [43] based on the sets of circadian transcripts in each layer. Gene sets were downloaded from the Molecular Signatures database (MSigDB) C2 (KEGG gene sets) and C5 (GO:BP terms) [57]. Sets containing fewer than five circadian transcripts were excluded from the analysis. The Kuiper test was used to identify circadian gene sets by comparing the acrophases of all circadian transcripts (rounded to the full hour) belonging to each gene set to a uniform background distribution and by testing for nonuniformity ($q < 0.05$ for GO terms, $q < 0.25$ for KEGG pathways).

A full description of all versions and references of R packages used is provided in Supplementary Table S2.

Assessment of variation in human epidermal and dermal rhythmic genes using the `variancePartition` pipeline

To study the different sources of variation in mean and circadian expression of the rhythmic genes in dermis and epidermis and to quantify their contribution, we used the open-source `variancePartition` package (available in Bioconductor) [44]. `variancePartition` provides a general framework for understanding drivers of variation in gene expression experiments with complex designs. It uses a linear mixed model to partition the variation attributable to multiple variables in the data. When providing an R formula indicating which meta-data variables to consider, `variancePartition` assesses the contribution of each meta-data variable to variation in gene expression by using a multiple regression model. This way, the effect of each meta-data variable is analyzed but while jointly accounting for all the others. Importantly, *external* and not internal time was used in the `variancePartition` analyses in order to be able to model time as a random effect (continuous variables like *internal* time cannot be modeled as random effects in `variancePartition`).

We assessed the drivers of variation in different sets of transcripts: in the total number of rhythmic genes in at least one layer (∼ 1400 transcripts) in Figure 2A, B; in the total number of rhythmic genes only in dermis (∼ 500 transcripts, Figure 2C, E) and only in epidermis (∼ 1200 transcripts, Figure 2D, F); in 1000 non-rhythmic genes (Figure S3A); in the rhythmic core-clock genes (Figure S3B) and in the 25 and 15 time-telling genes that we identified using `ZeitZeiger` in dermis and epidermis, respectively (Figure S5B).

Identification of potential biomarkers of molecular skin phase using `ZeitZeiger`

From the multiple methods to estimate internal circadian time, we used ZeitZeiger [47] (available in Bioconductor) to identify skin biomarkers of circadian phase. We tested three sets of predictors that differed on the dataset used (the whole set of expressed genes in at least one layer versus expressed genes in dermis and epidermis separately). The predicted variable was, in both cases, internal time. To evaluate the performance of the different predictors, we followed a leave-one-subject-out cross-validation approach in the lines of [47, 58]. To do this, predictors are trained with data from all subjects except one and internal time from the subject who is left-out is predicted. The process is iterated along all subjects and for different values of the two main parameters of ZeitZeiger, `sumabsv` and `nSPC`. The first parameter `sumabsv` controls how many genes form each sparse principal component (SPC) and the second parameter, `nSPC`, controls how many SPCs are used for prediction. Large values of either parameter result in more genes being needed for prediction. For each set of values of `sumabsv` and `nSPC` from the leave-one-subject-out cross-validation, we calculated the median absolute difference between the predicted and the observed internal time stamp across all subjects. We refer to this parameter as median absolute error (MAE), and it serves as a measure of accuracy of the prediction: the lower the error, the better the prediction.

Abbreviations

DLMO: dim-light melatonin onset; SCN: suprachiasmatic nucleus; MST: mid sleep time; FDR: false discovery rate; MAE: median absolute error; PSEA: phase set enrichment analysis; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; GEO: Gene Expression Omnibus; BH: Benjamini-Hochberg; SPC: sparse principal component.

Acknowledgments

AG Herzel, AG Kramer

Conflict of interest

Is any author working for Beiersdorf?, which markets skin care products. The remaining authors declare that they have no competing interests.

Data accessibility

The gene expression microarray data of dermis and epidermis generated in this study are available at the Gene Expression Omnibus (GEO) database and can be accessed under the accession number GSEblabla – ref – nicely documented!!. The gene expression data of longitudinal dermis and epidermis were publicly available at GEO (GSE139300 [46] and GSE112660 [45]). The source code is available through GitHub: link github. The software package versions used and their available links are listed in Supplement Table S2.

Funding

Beiersdorf paid for 100% of the costs of clinical work reported in this paper. This work is also supported by ...

Author contributions

Study design and conceptualization: BA, MdO, AK and HH. Experimental Methodology: SK, FS and AG. Bioinformatic Methodology: MdO, BA and KJ. Investigation: BA, MdO, KJ, AK and HH. Writing (original draft and editing): MdO and BA. Writing (review): MdO, BA, AK and HH. Funding acquisition: BA, AK and HH.

Author affiliations

[1]Institute for Theoretical Biology – Laboratory of Theoretical Chronobiology, Humboldt Universität zu Berlin and Charité Universitätsmedizin Berlin, Philippstraße 13, House 4, 10115 Berlin, Germany. [2]Institute for Medical Immunology – Laboratory of Chronobiology, Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. [3]Beiersdorf? [4]New affiliations from people?.

Supplementary Material

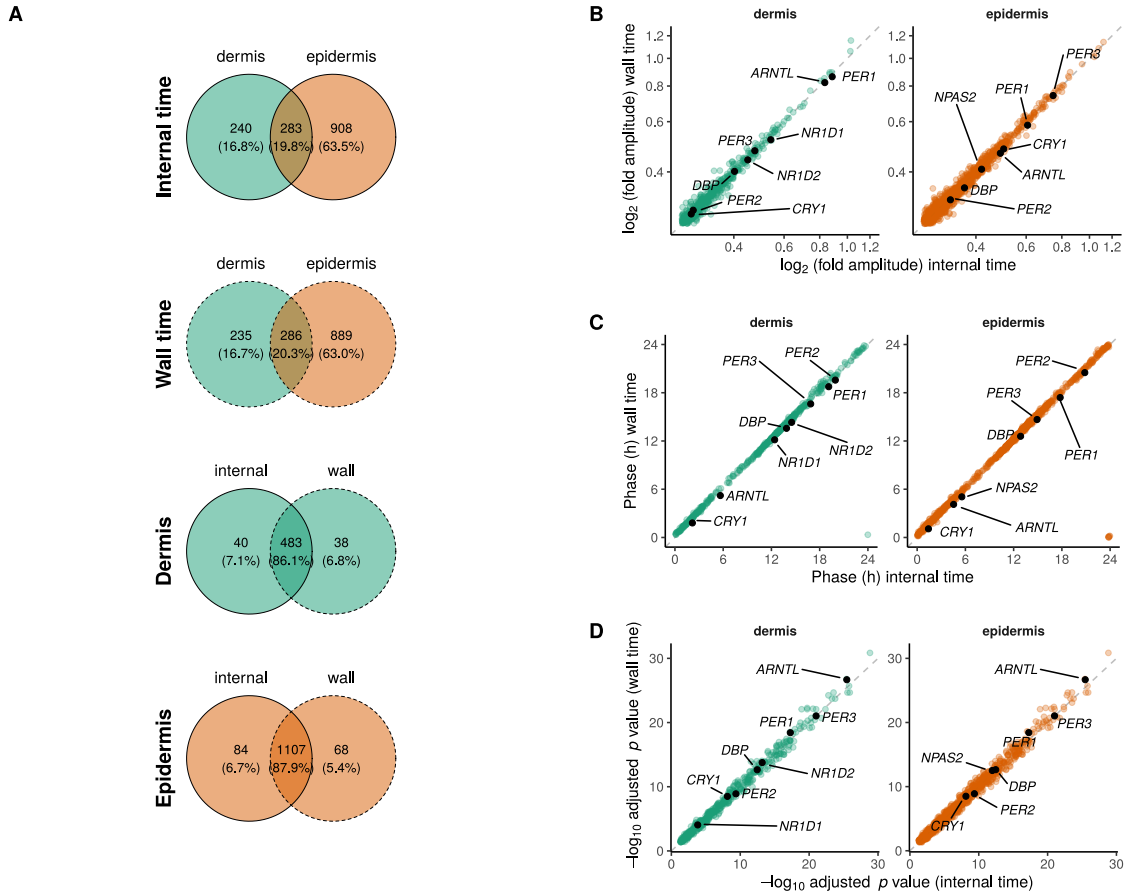


Figure S1: Analysis of the circadian transcriptome in human dermis and epiderms done with external (instead of internal) time. A. Venn diagram comparing the number of genes identified as circadian in dermis (green) vs. epidermis (orange) and in the analysis using internal time (solid line) or wall time (dashed line). **B.** Amplitude correlation of genes identified as rhythmic with the internal time analysis compared to external time analysis. **C.** Acrophase correlation of genes identified as rhythmic with the internal time analysis compared to external time analysis. **D.** BH-adjusted *p* value correlation of rhythmic genes identified with internal vs. external time analysis.

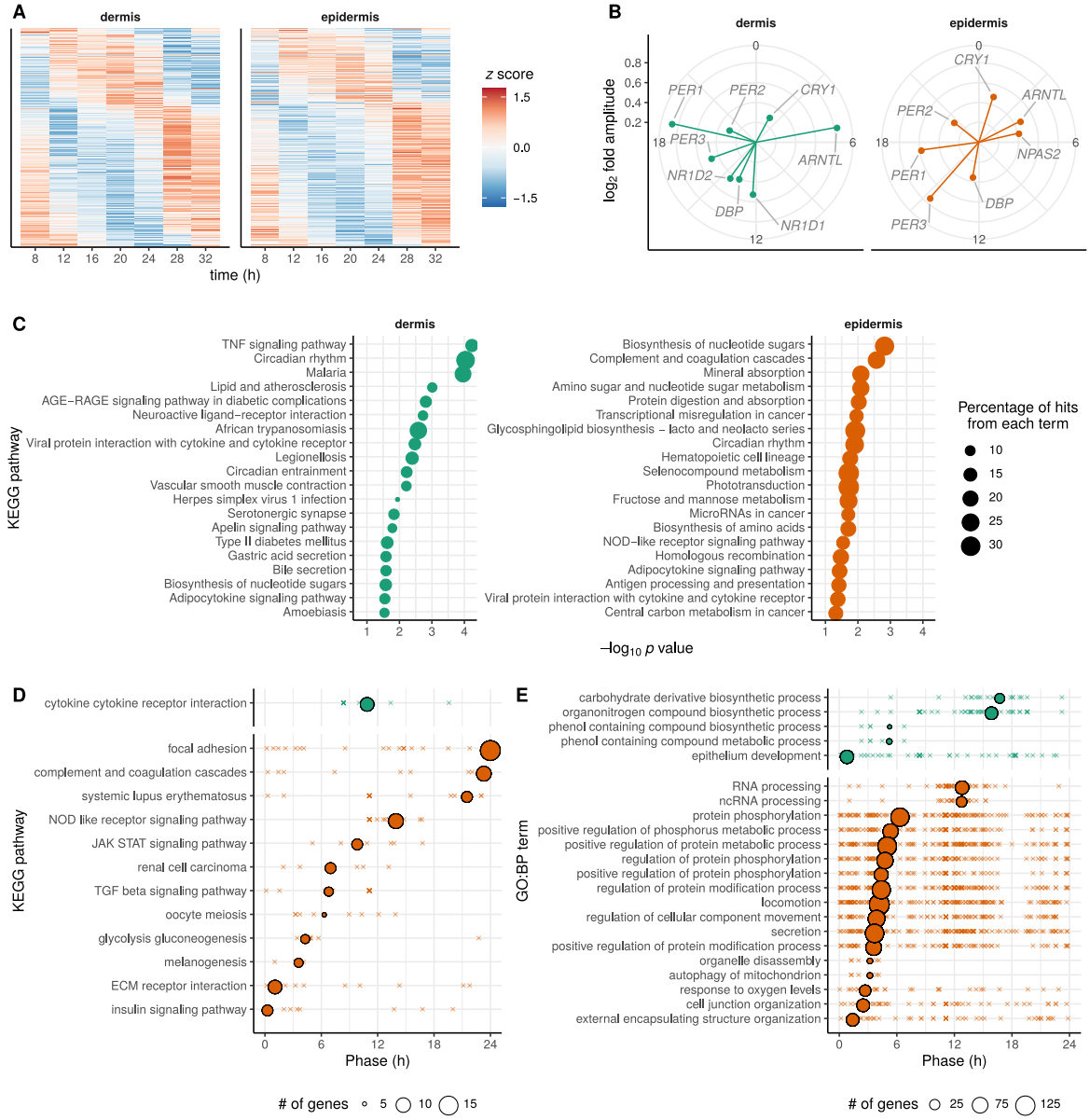


Figure S2: Analysis of the human circadian transcriptome in dermis and epiderms with respect to internal time. Acrophases and amplitudes of rhythmic genes were estimated from the cosinor analysis at FDR< 0.05 and relative ampitude > 0.26. **A.** *z* score-normalized, acrophase-ordered expression heatmap of the circadian transcriptome components from human dermis (left) and epidermis (right). **B.** Expression profiles of circadian core clock genes in human dermis (left) and epidermis (right). Arrow direction represents phase (expressed in units of circadian *internal* time) and arrow length depicts amplitude. **C.** Circadian KEGG pathway enrichment analysis of the rhythmic genes in dermis (green) and epidermis (orange). The top 20 enriched pathways (with a minimum gene set of 5 terms per category) in each layer are shown. **D.** Summary of significantly phase-clustered circadian GO terms ($q < 0.05$) and KEGG pathways ($q < 0.25$) in dermis (green) and epidermis (orange) as determined by PSEA. Sets containing fewer than five circadian transcripts were excluded from the analysis.

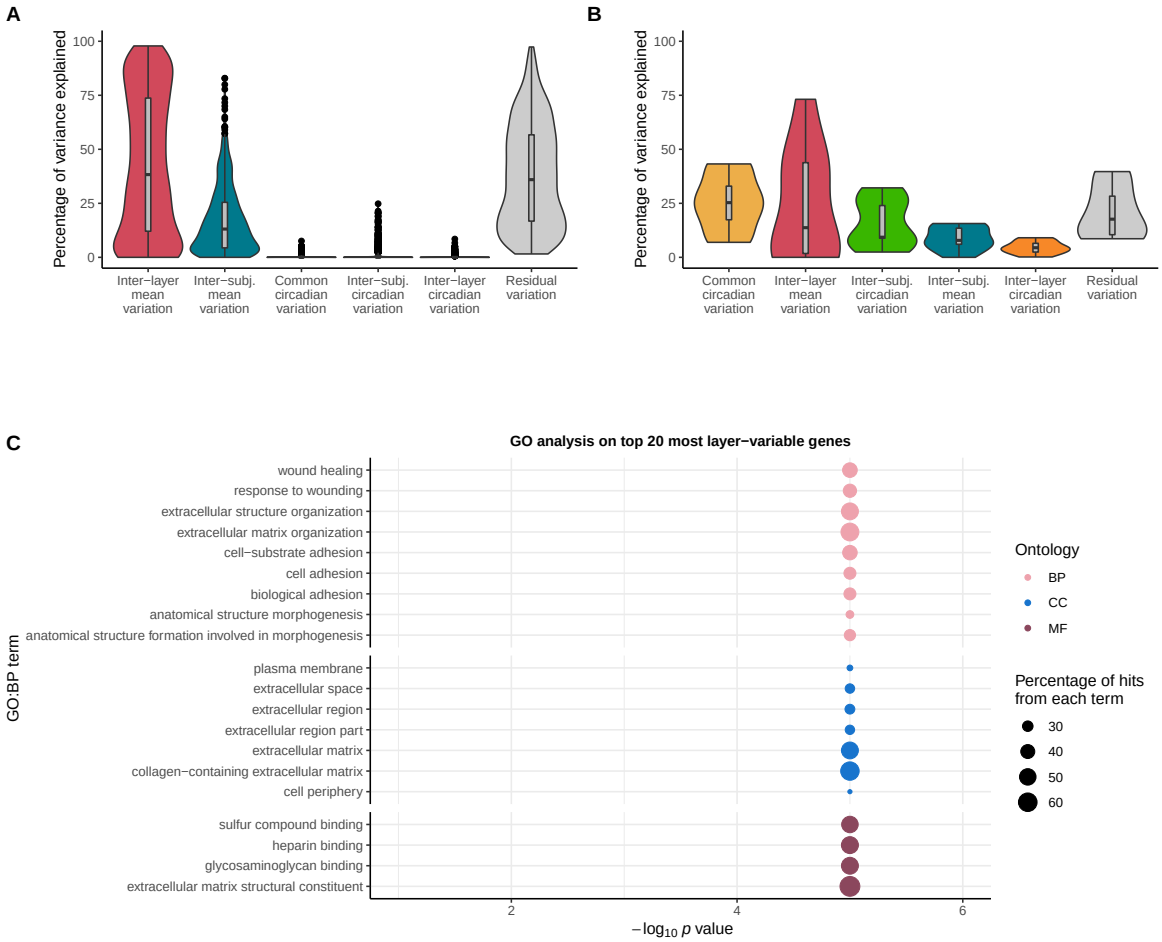
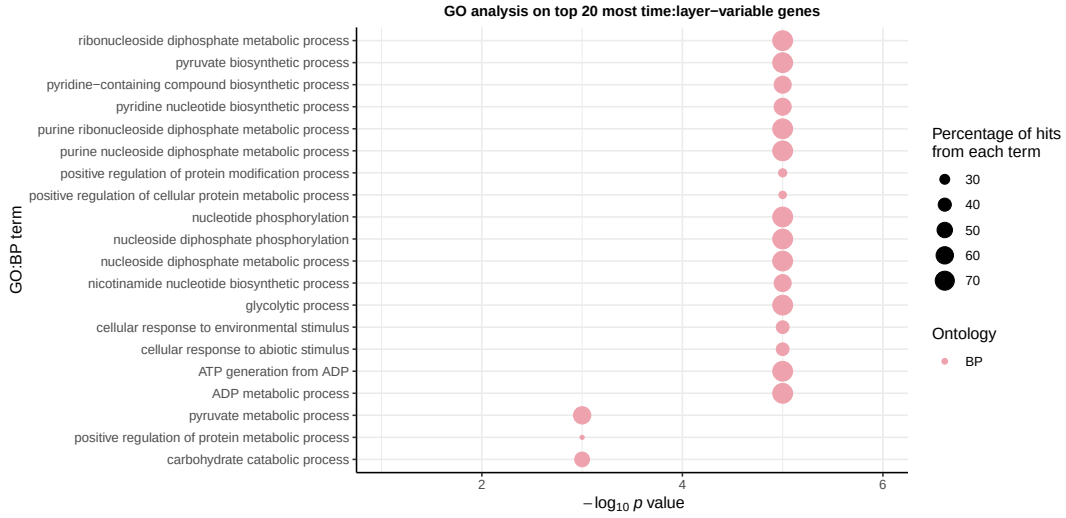
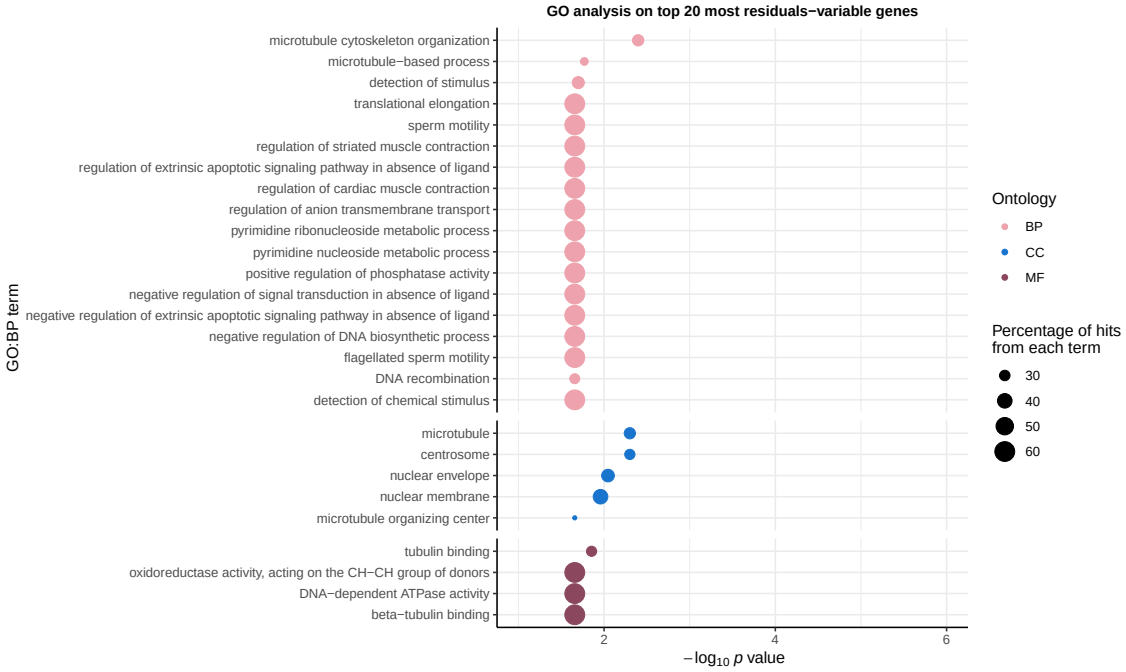


Figure S3: Drivers of variation in the human skin circadian transcriptome. **A.** Negative control of the `variancePartition` analysis: the pipeline was run in 1000 non-rhythmic genes (FDR>0.1) to show that external time does not represent a major source of variation among these transcripts. **B.** Positive control: `variancePartition` was run in the clock genes to show that time represents a major (in fact, the largest) source of variation. **C.** GO enrichment analysis in the 200 circadian transcripts with largest variability in mean expression across layers. **D.** GO enrichment analysis in the 200 rhythmic transcripts with largest rhythm variability across layers, i.e., transcripts that show differential rhythms in human dermis versus epidermis. **E.** GO enrichment analysis in the 200 circadian transcripts with largest residual variability in mean expression. For C-E: the top 20 enriched terms (with a minimum gene set of 5 terms from each category) are shown; pink represents biological processes; blue, cellular compartment; purple, molecular function; enrichment was performed testing against the background of ∼1400 rhythmic genes in at least one layer.

D


GO analysis on top 20 most time:layer−variable genes
ribonucleoside diphosphate metabolic process
pyruvate biosynthetic process
pyridine−containing compound biosynthetic process
pyridine nucleotide biosynthetic process
purine ribonucleoside diphosphate metabolic process
purine nucleoside diphosphate metabolic process
positive regulation of protein modification process
positive regulation of cellular protein metabolic process
nucleotide phosphorylation
nucleoside diphosphate phosphorylation
nucleoside diphosphate metabolic process
nicotinamide nucleotide biosynthetic process
glycolytic process
cellular response to environmental stimulus
cellular response to abiotic stimulus
ATP generation from ADP
ADP metabolic process
pyruvate metabolic process
positive regulation of protein metabolic process
carbohydrate catabolic process
GO:BP term
2
4
6
−log10 p value
Percentage of hits from each term
30
40
50
60
70
Ontology
BP

E


GO analysis on top 20 most residuals−variable genes
microtubule cytoskeleton organization
microtubule−based process
detection of stimulus
translational elongation
sperm motility
regulation of striated muscle contraction
regulation of extrinsic apoptotic signaling pathway in absence of ligand
regulation of cardiac muscle contraction
regulation of anion transmembrane transport
pyrimidine ribonucleoside metabolic process
pyrimidine nucleoside metabolic process
positive regulation of phosphatase activity
negative regulation of signal transduction in absence of ligand
negative regulation of extrinsic apoptotic signaling pathway in absence of ligand
negative regulation of DNA biosynthetic process
flagellated sperm motility
DNA recombination
detection of chemical stimulus
microtubule
centrosome
nuclear envelope
nuclear membrane
microtubule organizing center
tubulin binding
oxidoreductase activity, acting on the CH−CH group of donors
DNA−dependent ATPase activity
beta−tubulin binding
GO:BP term
2
4
6
−log10 p value
Ontology
BP
CC
MF
Percentage of hits from each term
30
40
50
60

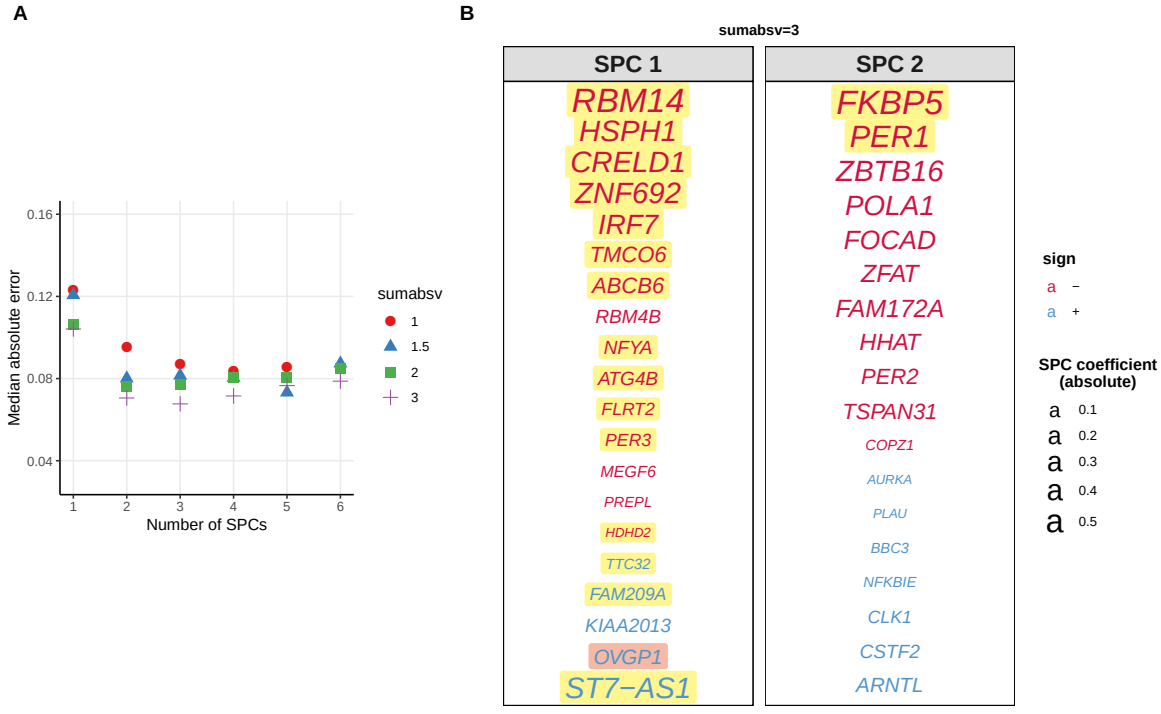


Figure S4: Identification of internal time-telling genes in human skin with `ZeitZeiger` [47]. A. Median absolute error of the internal-time prediction on cross-validation (see Materials and Methods for details) as a function of the two main parameters of `ZeitZeiger`, `sumabsv` and nSPC. **B.** Internal time predictors from human skin for `sumabsv`=3 and nSPC=2. Genes assigned to SPC1 or SPC2 as well as their coefficients are shown. Highlighted in yellow are genes that appeared in the top 20 most common circadian varying genes in the `variancePartition` analysis done in the rhythmic genes in human skin (i.e., in *at least* one layer); in orange, genes that showed differential rhythms across layers (i.e., genes with high inter-layer circadian variation from the `variancePartition` analysis). Note, that the difference to Figure 3 is that here, `ZeitZeiger` was run once using the whole ~11000 expressed genes in at least one layer, and not separately in each layer.

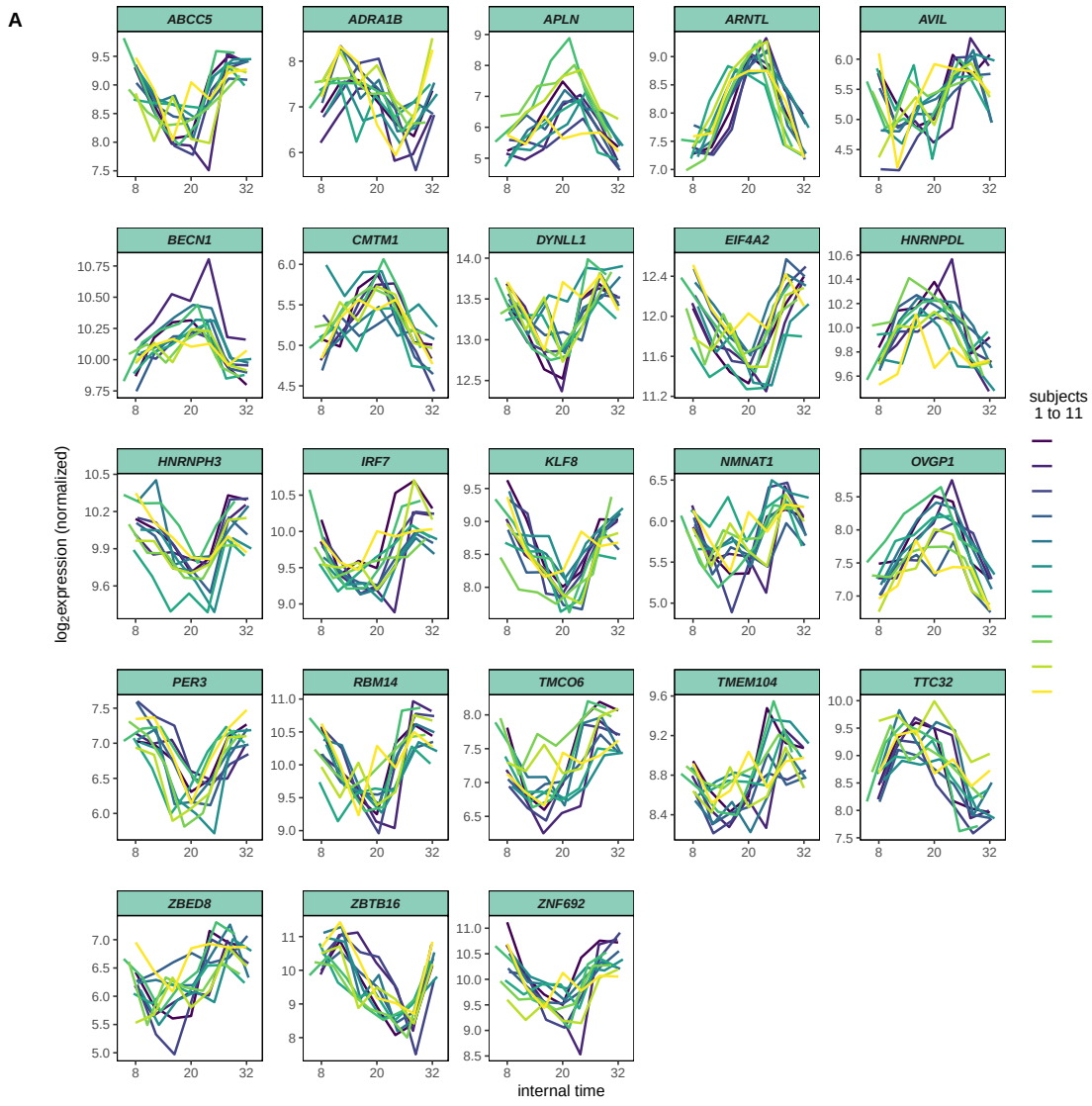
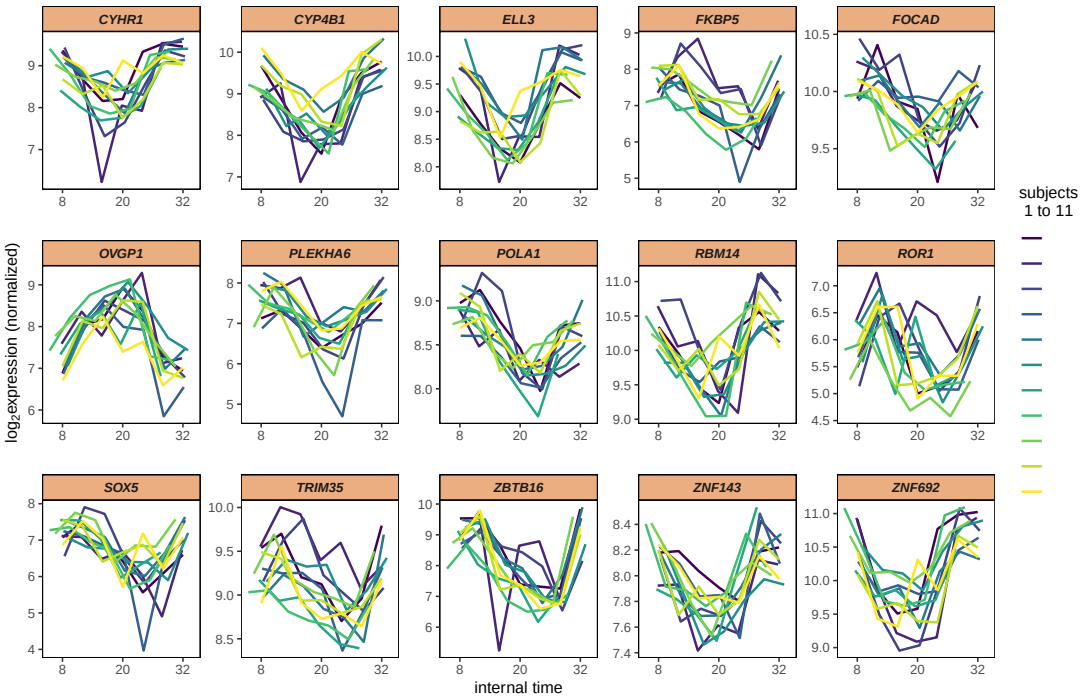
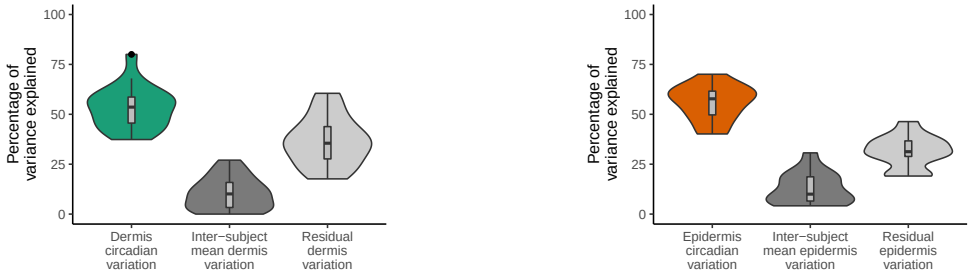


Figure S5: Identification of internal time-telling genes in human dermis and epidermis with `ZeitZeiger` [47]. **A.** Time series of expression of the time-telling genes in dermis (green) and epidermis (orange). Colored lines represent the expression profiles in different subjects. **B.** `variancePartition` analysis on the `ZeitZeiger` genes in dermis (left) or epidermis (right) shows high temporal variation and minor variability in mean expression across subjects. **C.** Expression profiles of the time-telling genes from our cohort in dermis (left) and epidermis (right) represented in SPC space and faceted by subject. Colors indicate internal time. `ZeitZeiger` was run with all ~11000 expressed genes and separately for dermis and epidermis.


CYHR1
CYP4B1
ELL3
FKBP5
FOCAD
OVGP1
PLEKHA6
POLA1
RBM14
ROR1
SOX5
TRIM35
ZBTB16
ZNF143
ZNF692
log2expression (normalized)
internal time
subjects
1 to 11

B


Percentage of variance explained
Dermis circadian variation
Inter−subject mean dermis variation
Residual dermis variation
Epidermis circadian variation
Inter−subject mean epidermis variation
Residual epidermis variation

C

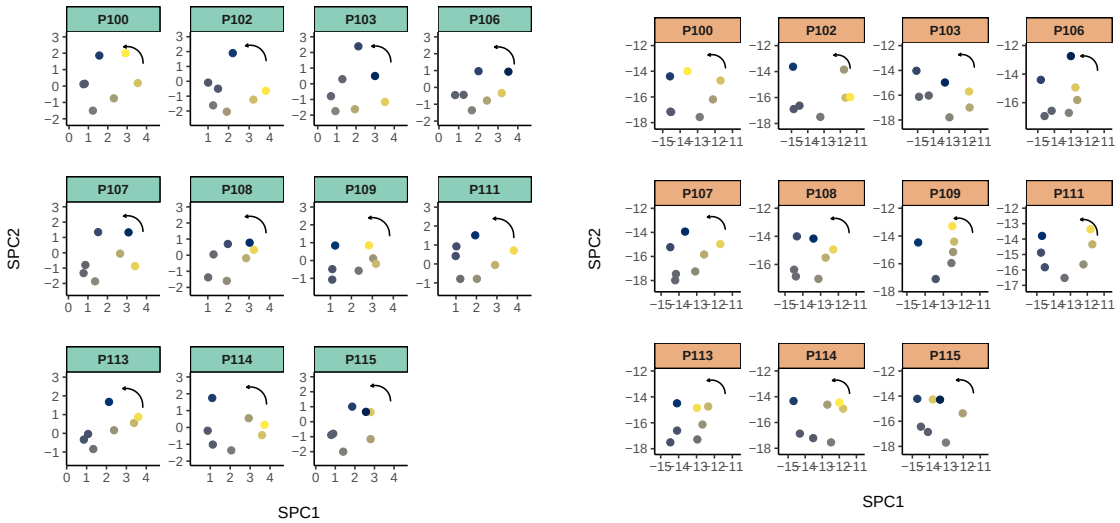

P100
P102
P103
P106
P107
P108
P109
P111
P113
P114
P115
SPC1
SPC2

Table S1: Meta-data about the subjects collected for this study. Mid sleep time was calculated and was corrected for the sleep-debt accumulated during the working days as described in [30, 40].

Subject	Sex	Birth year	Bed time work days	Sleep time work days	Min fall asleep work days	Wake up time work days	Min wake up work days	Alarm work days?
P108	male	1984	22:45	23:00	20	6:30	15	Y
P100	male	1991	23:00	23:00	7.5	7:00	5	Y
P113	male	1985	23:30	0:00	30	8:30	10	Y
P106	male	1982	23:00	23:15	15	6:52	7.5	Y
P102	male	1989	23:00	23:00	10	7:30	0	Y
P109	male	1984	23:00	23:00	5	8:00	0	Y
P103	female	1988	0:00	0:25	25	7:30	9	Y
P107	female	1987	23:00	23:00	15	6:00	5	Y
P111	female	1983	0:00	0:00	5	8:00	15	Y
P114	female	1986	23:30	23:30	5	8:30	5	Y
P115	female	1981	22:00	22:15	5	5:20	5	Y

Subject	Wake up before alarm?	Bed time free days	Sleep time free days	Min fall asleep free days	Wake up time free days	Wake up time free days	Alarm free days?
P108	Y	23:00	23:15	15	6:30	30	N
P100	N	0:00	0:00	5	8:00	15	Y
P113	N	1:00	1:15	20	9:30	15	Y
P106	Y	23:30	23:45	15	9:15	60	Y
P102	N	0:00	0:00	10	8:00	10	N
P109	Y	0:00	0:00	5	8:30	30	N
P103	N	0:00	0:00	25	7:30	5	Y
P107	Y	0:00	0:00	15	7:30	5	N
P111	Y	2:30	2:30	5	11:00	30	N
P114		23:30	23:30	5	8:30	5	N
P115	N	0:00	0:15	5	8:30	30	N

Subject	Corrected mid sleep time
P108	02:58
P100	04:00
P113	05:28
P106	03:50
P102	04:11
P109	04:26
P103	03:36
P107	03:34
P111	06:34
P114	04:00
P115	03:58

Table S2: List of software packages used in this study. Versions and references are included.

Software package	Version	Ref.
Biobase	2.46.0	[59]
clusterProfiler	3.14.3	[60]
cowplot	1.1.1	[61]
doParallel	1.0.16	[62]
dplyr	1.0.5	[63]
GEOquery	2.54.1	[64]
ggforce	0.3.3	[65]
ggplot2	3.3.3	[66]
ggrepel	0.9.1	[67]
ggthemes	4.2.4	[68]

Software package	Version	Ref.
ggvenn	0.1.8	[69]
hgug4112a.db	3.2.3	[70]
hms	1.0.0	[71]
limma	3.42.2	[72]
lubridate	1.7.10	[73]
magrittr	2.0.1	[74]
msigdbr	7.4.1	[75]
oligo??		**which?**
PSEA	1.1	[43]
R	3.6.3	[76]

Software package	Version	Ref.
stringr	1.4.0	[77]
tibble	3.1.1	[78]
tidyr	1.1.3	[79]
tidytext	0.3.2	[80]
tidyverse	1.3.1	[81]
variancePartition	1.16.1	[44]
viridis	0.6.1	[82]
zeitzeiger	2.0.2	[47]

References

[1] Plikus, M. V., Van Spyk, E. N., Pham, K., Geyfman, M., Kumar, V., Takahashi, J. S., and Andersen, B., "The circadian clock in skin: implications for adult stem cells, tissue regeneration, cancer, aging, and immunity," *Journal of biological rhythms*, Vol. 30, No. 3, 2015, pp. 163–182.

[2] Wong, R., Geyer, S., Weninger, W., Guimberteau, J.-C., and Wong, J. K., "The dynamic anatomy and patterning of skin," *Experimental dermatology*, Vol. 25, No. 2, 2016, pp. 92–98.

[3] Zouboulis, C. C., "The skin as an endocrine organ," *Dermato-endocrinology*, Vol. 1, No. 5, 2009, pp. 250–252.

[4] Takahashi, J. S., "Transcriptional architecture of the mammalian circadian clock," *Nature Reviews Genetics*, Vol. 18, No. 3, 2017, pp. 164–179.

[5] Dibner, C., Schibler, U., and Albrecht, U., "The mammalian circadian timing system: organization and coordination of central and peripheral clocks," *Annual review of physiology*, Vol. 72, 2010, pp. 517–549.

[6] Akhtar, R. A., Reddy, A. B., Maywood, E. S., Clayton, J. D., King, V. M., Smith, A. G., Gant, T. W., Hastings, M. H., and Kyriacou, C. P., "Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven by the suprachiasmatic nucleus," *Current biology*, Vol. 12, No. 7, 2002, pp. 540–550.

[7] Duffield, G. E., Best, J. D., Meurers, B. H., Bittner, A., Loros, J. J., and Dunlap, J. C., "Circadian programs of transcriptional activation, signaling, and protein turnover revealed by microarray analysis of mammalian cells," *Current Biology*, Vol. 12, No. 7, 2002, pp. 551–557.

[8] Miller, B. H., McDearmon, E. L., Panda, S., Hayes, K. R., Zhang, J., Andrews, J. L., Antoch, M. P., Walker, J. R., Esser, K. A., Hogenesch, J. B., et al., "Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation," *Proceedings of the National Academy of Sciences*, Vol. 104, No. 9, 2007, pp. 3342–3347.

[9] Panda, S., Antoch, M. P., Miller, B. H., Su, A. I., Schook, A. B., Straume, M., Schultz, P. G., Kay, S. A., Takahashi, J. S., and Hogenesch, J. B., "Coordinated transcription of key pathways in the mouse by the circadian clock," *Cell*, Vol. 109, No. 3, 2002, pp. 307–320.

[10] Keller, M., Mazuch, J., Abraham, U., Eom, G. D., Herzog, E. D., Volk, H.-D., Kramer, A., and Maier, B., "A circadian clock in macrophages controls inflammatory immune responses," *Proceedings of the National Academy of Sciences*, Vol. 106, No. 50, 2009, pp. 21407–21412.

[11] Storch, K.-F., Lipan, O., Leykin, I., Viswanathan, N., Davis, F. C., Wong, W. H., and Weitz, C. J., "Extensive and divergent circadian gene expression in liver and heart," *Nature*, Vol. 417, No. 6884, 2002, pp. 78–83.

[12] Mohawk, J. A., Green, C. B., and Takahashi, J. S., "Central and peripheral circadian clocks in mammals," *Annual review of neuroscience*, Vol. 35, 2012, pp. 445–462.

[13] Geyfman, M., Kumar, V., Liu, Q., Ruiz, R., Gordon, W., Espitia, F., Cam, E., Millar, S. E., Smyth, P., Ihler, A., et al., "Brain and muscle Arnt-like protein-1 (BMAL1) controls circadian cell proliferation and susceptibility to UVB-induced DNA damage in the epidermis," *Proceedings of the National Academy of Sciences*, Vol. 109, No. 29, 2012, pp. 11758–11763.

[14] Wang, H., van Spyk, E., Liu, Q., Geyfman, M., Salmans, M. L., Kumar, V., Ihler, A., Li, N., Takahashi, J. S., and Andersen, B., "Time-restricted feeding shifts the skin circadian clock and alters UVB-induced DNA damage," *Cell reports*, Vol. 20, No. 5, 2017, pp. 1061–1072.

[15] Yosipovitch, G., Sackett-Lundeen, L., Goon, A., Huak, C. Y., Goh, C. L., and Haus, E., "Circadian and ultradian (12 h) variations of skin blood flow and barrier function in non-irritated and irritated skin—effect of topical corticosteroids," *Journal of investigative dermatology*, Vol. 122, No. 3, 2004, pp. 824–829.

[16] Yosipovitch, G., Xiong, G. L., Haus, E., Sackett-Lundeen, L., Ashkenazi, I., and Maibach, H. I., "Time-dependent variations of the skin barrier function in humans: transepidermal water loss, stratum corneum hydration, skin surface pH, and skin temperature," *Journal of investigative dermatology*, Vol. 110, No. 1, 1998, pp. 20–23.

[17] Le Fur, I., Reinberg, A., Lopez, S., Morizot, F., Mechkouri, M., and Tschachler, E., "Analysis of circadian and ultradian rhythms of skin surface properties of face and forearm of healthy women," *Journal of investigative dermatology*, Vol. 117, No. 3, 2001, pp. 718–724.

[18] Tanioka, M., Yamada, H., Doi, M., Bando, H., Yamaguchi, Y., Nishigori, C., and Okamura, H., "Molecular clocks in mouse skin," *Journal of Investigative Dermatology*, Vol. 129, No. 5, 2009, pp. 1225–1231.

[19] Zanello, S. B., Jackson, D. M., and Holick, M. F., "Expression of the circadian clock genes clock and period1 in human skin," *Journal of investigative dermatology*, Vol. 115, No. 4, 2000, pp. 757–760.

[20] Bjarnason, G. A., Jordan, R. C., Wood, P. A., Li, Q., Lincoln, D. W., Sothern, R. B., Hrushesky, W. J., and Ben-David, Y., "Circadian expression of clock genes in human oral mucosa and skin: association with specific cell-cycle phases," *The American journal of pathology*, Vol. 158, No. 5, 2001, pp. 1793–1801.

[21] Kawara, S., Mydlarski, R., Shivji, G., Tavadia, S. K., Suzuki, H., Mamelak, A. J., Freed, I., Wang, B., Watanabe, H., Bjarnason, G. A., et al., "Low-dose ultraviolet B rays alter the mRNA expression of the circadian clock genes in cultured human keratinocytes," *Journal of investigative dermatology*, Vol. 119, No. 6, 2002, pp. 1220–1223.

[22] Oishi, K., Fukui, H., Sakamoto, K., Miyazaki, K., Kobayashi, H., and Ishida, N., "Differential expressions of mPer1 and mPer2 mRNAs under a skeleton photoperiod and a complete light–dark cycle," *Molecular brain research*, Vol. 109, No. 1-2, 2002, pp. 11–17.

[23] Brown, S. A., Fleury-Olela, F., Nagoshi, E., Hauser, C., Juge, C., Meier, C. A., Chicheportiche, R., Dayer, J.-M., Albrecht, U., and Schibler, U., "The period length of fibroblast circadian gene expression varies widely among human individuals," *PLoS biology*, Vol. 3, No. 10, 2005, pp. e338.

[24] Brown, S. A., Kunz, D., Dumas, A., Westermark, P. O., Vanselow, K., Tilmann-Wahnschaffe, A., Herzel, H., and Kramer, A., "Molecular insights into human daily behavior," *Proceedings of the National Academy of Sciences*, Vol. 105, No. 5, 2008, pp. 1602–1607.

[25] Spörl, F., Schellenberg, K., Blatt, T., Wenck, H., Wittern, K.-P., Schrader, A., and Kramer, A., "A circadian clock in HaCaT keratinocytes," *Journal of Investigative Dermatology*, Vol. 131, No. 2, 2011, pp. 338–348.

[26] Anafi, R. C., Francey, L. J., Hogenesch, J. B., and Kim, J., "CYCLOPS reveals human transcriptional rhythms in health and disease," *Proceedings of the National Academy of Sciences*, Vol. 114, No. 20, 2017, pp. 5312–5317.

[27] Montaigne, D., Marechal, X., Modine, T., Coisne, A., Mouton, S., Fayad, G., Ninni, S., Klein, C., Ortmans, S., Seunes, C., et al., "Daytime variation of perioperative myocardial injury in cardiac surgery and its prevention by Rev-Erbα antagonism: a single-centre propensity-matched cohort study and a randomised study," *The Lancet*, Vol. 391, No. 10115, 2018, pp. 59–69.

[28] Dallmann, R., Okyar, A., and Lévi, F., "Dosing-time makes the poison: circadian regulation and pharmacotherapy," *Trends in molecular medicine*, Vol. 22, No. 5, 2016, pp. 430–445.

[29] Long, J. E., Drayson, M. T., Taylor, A. E., Toellner, K. M., Lord, J. M., and Phillips, A. C., "Morning vaccination enhances antibody response over afternoon vaccination: a cluster-randomised trial," *Vaccine*, Vol. 34, No. 24, 2016, pp. 2679–2685.

[30] Roenneberg, T., Kuehnle, T., Juda, M., Kantermann, T., Allebrandt, K., Gordijn, M., and Merrow, M., "Epidemiology of the human circadian clock," *Sleep medicine reviews*, Vol. 11, No. 6, 2007, pp. 429–438.

[31] Horne, J. A. and Östberg, O., "A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms." *International journal of chronobiology*, 1976.

[32] Roenneberg, T., Wirz-Justice, A., and Merrow, M., "Life between clocks: daily temporal patterns of human chronotypes," *Journal of biological rhythms*, Vol. 18, No. 1, 2003, pp. 80–90.

[33] Pandi-Perumal, S. R., Smits, M., Spence, W., Srinivasan, V., Cardinali, D. P., Lowe, A. D., and Kayumov, L., "Dim light melatonin onset (DLMO): a tool for the analysis of circadian phase in human sleep and chronobiological disorders," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, Vol. 31, No. 1, 2007, pp. 1–11.

[34] Laing, E. E., Möller-Levet, C. S., Poh, N., Santhi, N., Archer, S. N., and Dijk, D.-J., "Blood transcriptome based biomarkers for human circadian phase," *Elife*, Vol. 6, 2017, pp. e20214.

[35] Hsu, P.-K., Ptáček, L. J., and Fu, Y.-H., "Genetics of human sleep behavioral phenotypes," *Methods in enzymology*, Vol. 552, 2015, pp. 309–324.

[36] Stothard, E. R., McHill, A. W., Depner, C. M., Birks, B. R., Moehlman, T. M., Ritchie, H. K., Guzzetti, J. R., Chinoy, E. D., LeBourgeois, M. K., Axelsson, J., et al., "Circadian entrainment to the natural light-dark cycle across seasons and the weekend," *Current Biology*, Vol. 27, No. 4, 2017, pp. 508–513.

[37] Wright Jr, K. P., McHill, A. W., Birks, B. R., Griffin, B. R., Rusterholz, T., and Chinoy, E. D., "Entrainment of the human circadian clock to the natural light-dark cycle," *Current Biology*, Vol. 23, No. 16, 2013, pp. 1554–1558.

[38] Allebrandt, K. V., Teder-Laving, M., Kantermann, T., Peters, A., Campbell, H., Rudan, I., Wilson, J. F., Metspalu, A., and Roenneberg, T., "Chronotype and sleep duration: the influence of season of assessment," *Chronobiology international*, Vol. 31, No. 5, 2014, pp. 731–740.

[39] Laing, E. E., Möller-Levet, C. S., Archer, S. N., and Dijk, D.-J., "Universal and robust assessment of circadian time?" *Proceedings of the National Academy of Sciences*, Vol. 116, No. 12, 2019, pp. 5205–5205.

[40] Vetter, C., Winnebeck, E. C., and Roenneberg, T., "Asking the Clock: How to Use Information from Questionnaires for Circadian Phenotyping," *Circadian Clocks*, 2021, pp. 79–85.

[41] Wu, G., Ruben, M. D., Schmidt, R. E., Francey, L. J., Smith, D. F., Anafi, R. C., Hughey, J. J., Tasseff, R., Sherrill, J. D., Oblong, J. E., et al., "Population-level rhythms in human skin with implications for circadian medicine," *Proceedings of the National Academy of Sciences*, Vol. 115, No. 48, 2018, pp. 12313–12318.

[42] Wu, G., Ruben, M. D., Francey, L. J., Smith, D. F., Sherrill, J. D., Oblong, J. E., Mills, K. J., and Hogenesch, J. B., "A population-based gene expression signature of molecular clock phase from a single epidermal sample," *Genome medicine*, Vol. 12, No. 1, 2020, pp. 1–12.

[43] Zhang, R., Podtelezhnikov, A. A., Hogenesch, J. B., and Anafi, R. C., "Discovering biology in periodic data through phase set enrichment analysis (PSEA)," *Journal of biological rhythms*, Vol. 31, No. 3, 2016, pp. 244–257.

[44] Hoffman, G. E. and Schadt, E. E., "variancePartition: interpreting drivers of variation in complex gene expression studies," *BMC bioinformatics*, Vol. 17, No. 1, 2016, pp. 1–13.

[45] Mills, K. J. and Oblong, J. E., "The effect of circadian rhythm on gene expression in human skin." *Datasets. Gene Expression Omnibus.*, `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112660` (Accessed 18 Oct 2021).

[46] Mills, K. J. and Oblong, J. E., "The effect of circadian rhythm on gene expression in human skin II." *Datasets. Gene Expression Omnibus.*, `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139300` (Accessed 18 Oct 2021).

[47] Hughey, J. J., Hastie, T., and Butte, A. J., "ZeitZeiger: supervised learning for high-dimensional data from an oscillatory system," *Nucleic Acids Research*, Vol. 44, No. 8, 2016, pp. e80–e80.

[48] Weiss, S. C., Rowell, R., and Krochmal, L., "Impact of seasonality on conducting clinical studies in dermatology," *Clinics in dermatology*, Vol. 26, No. 5, 2008, pp. 565–569.

[49] Gaddameedhi, S., Selby, C. P., Kaufmann, W. K., Smart, R. C., and Sancar, A., "Control of skin cancer by the circadian rhythm," *Proceedings of the National Academy of Sciences*, Vol. 108, No. 46, 2011, pp. 18790–18795.

[50] Forni, M. F., Peloggia, J., Braga, T. T., Chinchilla, J. E. O., Shinohara, J., Navas, C. A., Camara, N. O. S., and Kowaltowski, A. J., "Caloric restriction promotes structural and metabolic changes in the skin," *Cell reports*, Vol. 20, No. 11, 2017, pp. 2678–2692.

[51] Dijk, D.-J. and Duffy, J. F., "Novel approaches for assessing circadian rhythmicity in humans: a review," *Journal of Biological Rhythms*, Vol. 35, No. 5, 2020, pp. 421–438.

[52] Klerman, E. B., Gershengorn, H. B., Duffy, J. F., and Kronauer, R. E., "Comparisons of the variability of three markers of the human circadian pacemaker," *Journal of biological rhythms*, Vol. 17, No. 2, 2002, pp. 181–193.

[53] Danilenko, K. V., Verevkin, E. G., Antyufeev, V. S., Wirz-Justice, A., and Cajochen, C., "The hockey-stick method to estimate evening dim light melatonin onset (DLMO) in humans," *Chronobiology International*, Vol. 31, No. 3, 2014, pp. 349–355.

[54] Archer, S. N., Laing, E. E., Möller-Levet, C. S., van der Veen, D. R., Bucca, G., Lazar, A. S., Santhi, N., Slak, A., Kabiljo, R., von Schantz, M., et al., "Mistimed sleep disrupts circadian regulation of the human transcriptome," *Proceedings of the National Academy of Sciences*, Vol. 111, No. 6, 2014, pp. E682–E691.

[55] Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P., and Trent, J. M., "Expression profiling using cDNA microarrays," *Nature genetics*, Vol. 21, No. 1, 1999, pp. 10–14.

[56] Pelikan, A., Herzel, H., Kramer, A., and Ananthasubramaniam, B., "Venn diagram analysis overestimates the extent of circadian rhythm reprogramming," *The FEBS Journal*, 2021.

[57] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., et al., "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences*, Vol. 102, No. 43, 2005, pp. 15545–15550.

[58] Wittenbrink, N., Ananthasubramaniam, B., Münch, M., Koller, B., Maier, B., Weschke, C., Bes, F., de Zeeuw, J., Nowozin, C., Wahnschaffe, A., et al., "High-accuracy determination of internal circadian time from a single blood sample," *The Journal of clinical investigation*, Vol. 128, No. 9, 2018, pp. 3826–3839.

[59] Gentleman, R., Carey, V., Morgan, M., and Falcon, S., "Biobase: base functions for Bioconductor," *R package version*, Vol. 1, 2005, pp. 12.

[60] Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y., "clusterProfiler: an R package for comparing biological themes among gene clusters," *Omics: a journal of integrative biology*, Vol. 16, No. 5, 2012, pp. 284–287.

[61] Wilke, C. O., "cowplot: streamlined plot theme and plot annotations for 'ggplot2'. R package version 1.0.0," 2019, see `https://cran.r-project.org/web/packages/cowplot/index.html`.

[62] Analytics, R. and Weston, S., "doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.10," 2015, see `https://cran.r-project.org/web/packages/doParallel/index.html`.

[63] Wickham, H., Francois, R., Henry, L., and Müller, K., "dplyr: A Grammar of Data Manipulation. R package version 0.4.3," *R Found. Stat. Comput., Vienna*, 2015, see `https://cran.r-project.org/web/packages/dplyr/index.html`.

[64] Davis, S. and Meltzer, P. S., "GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor," *Bioinformatics*, Vol. 23, No. 14, 2007, pp. 1846–1847.

[65] Pedersen, T. L., "ggforce: accelerating 'ggplot2'. R package version 0.3.1," 2019, see https://cran.r-project.org/web/packages/ggforce/index.html.

[66] Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., and Yutani, H., "ggplot2: Create elegant data visualisations using the grammar of graphics. R package version 2.2.1," *Stata Software Package: College Station, TX, USA*, 2016, see https://cran.r-project.org/web/packages/ggplot2/index.html.

[67] Slowikowski, K., "ggrepel: Automatically position non-overlapping text labels with 'ggplot2'. R package version 0.8.0," 2018, see https://cran.r-project.org/web/packages/ggrepel/index.html.

[68] Arnold, J. B., "ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. R package version 4.2.0," *Retrieved September*, Vol. 12, 2019, pp. 2020, see https://cran.r-project.org/web/packages/ggthemes/index.html.

[69] Yan, L., "ggvenn: Draw Venn Diagram by 'ggplot2'. R package version 0.1.8," 2021, see https://cran.r-project.org/web/packages/ggvenn/index.html.

[70] Carlson, M., Falcon, S., et al., "hgug4112a.db: Agilent "Human Genome, Whole" annotation data (chip hgug4112a)," *R package version*, Vol. 3, No. 2, 2016, see https://bioconductor.org/packages/release/data/annotation/html/hgug4112a.db.html.

[71] Müller, K., "hms: Pretty Time of Day. R package version 0.5.0," 2019, see https://cran.r-project.org/web/packages/hms/index.html.

[72] Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K., "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic acids research*, Vol. 43, No. 7, 2015, pp. e47–e47.

[73] Grolemund, G. and Wickham, H., "Package lubridate (Version 1.2.0)," 2012, see https://cran.r-project.org/web/packages/lubridate/index.html.

[74] Bache, S. M. and Wickham, H., "MAGRITTR: A forwardpipe operator for R. R package version, 1.5 (1)," 2016, see https://cran.r-project.org/web/packages/magrittr/index.html.

[75] Dolgalev, I., "MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format [R package msigdbr version 7.0.1]," 2020, see https://cran.r-project.org/web/packages/msigdbr/index.html.

[76] Ripley, B. D. et al., "The R project in statistical computing," *MSOR Connections. The newsletter of the LTSN Maths, Stats & OR Network*, Vol. 1, No. 1, 2001, pp. 23–25, see https://www.r-project.org/.

[77] Wickham, H., "stringr: Simple, consistent wrappers for common string operations (R package Version 1.4.0) [Computer software]," 2018, see https://cran.r-project.org/web/packages/stringr/index.html.

[78] Müller, K. and Wickham, H., "tibble: Simple Data Frames. R package version 2.1.3," 2019, see https://cran.r-project.org/web/packages/tibble/index.html.

[79] Wickham, H. and Henry, L., "tidyr: Tidy Messy Data. R package version 1.0.2," 2020, see https://cran.r-project.org/web/packages/tidyr/index.html.

[80] De Queiroz, G., Keyes, O., Robinson, D., and Silge, J., "TidyText: text mining using 'dplyr', 'ggplot2', and other tidy tools [R package tidytext version 0.1.5]," 2017, see https://cran.r-project.org/web/packages/tidytext/index.html.

[81] Wickham, H., "tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1," 2017, see https://cran.r-project.org/web/packages/tidyverse/index.html.

[82] "viridis: Colorblind-Friendly Color Maps for R [R package version 0.6.2]," see https://cran.r-project.org/web/packages/viridis/index.html.