

## Accepted Manuscript

Title: Evaluation of mRNA markers for estimating blood deposition time: towards alibi testing from human forensic stains with rhythmic biomarkers

Author: Karolina Lech Fan Liu Katrin Ackermann Victoria L. Revell Oscar Lao Debra J. Skene Manfred Kayser



PII: S1872-4973(15)30105-8  
DOI: <http://dx.doi.org/doi:10.1016/j.fsigen.2015.12.008>  
Reference: FSIGEN 1449

To appear in: *Forensic Science International: Genetics*

Received date: 16-9-2015  
Revised date: 13-11-2015  
Accepted date: 16-12-2015

Please cite this article as: Karolina Lech, Fan Liu, Katrin Ackermann, Victoria L.Revell, Oscar Lao, Debra J.Skene, Manfred Kayser, Evaluation of mRNA markers for estimating blood deposition time: towards alibi testing from human forensic stains with rhythmic biomarkers, *Forensic Science International: Genetics* <http://dx.doi.org/10.1016/j.fsigen.2015.12.008>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Evaluation of mRNA markers for estimating blood deposition time: towards alibi testing from human forensic stains with rhythmic biomarkers**

Karolina Lech<sup>1</sup>, Fan Liu<sup>1</sup>, Katrin Ackermann<sup>1,2</sup>, Victoria L. Revell<sup>3</sup>, Oscar Lao<sup>1,4</sup>, Debra J. Skene<sup>3</sup>, and Manfred Kayser<sup>1\*</sup>

<sup>1</sup>-Department of Genetic Identification, Erasmus MC University Medical Centre Rotterdam, Rotterdam, The Netherlands

<sup>2</sup>-EaStCHEM School of Chemistry, Biomedical Sciences Research Complex and Centre of Magnetic Resonance, University of St Andrews, St Andrews, United Kingdom

<sup>3</sup>-Centre for Chronobiology, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

<sup>4</sup>- CNAG - CRG; Centro Nacional de Análisis Genómico, Parc Científic de Barcelona - Torre I, Barcelona, Spain

\* - Corresponding author:

Prof. Dr. Manfred Kayser, Department of Genetic Identification, Erasmus MC University Medical Centre Rotterdam, Rotterdam, The Netherlands, Email:[m.kayser@erasmusmc.nl](mailto:m.kayser@erasmusmc.nl); Tel: +31 10 703 80 73; Fax: +31 10 704 45 75

## Highlights

- **Knowing the time a human trace was left at a crime scene is crucial for solving the case**
- **We recently showed that timing trace deposition with circadian hormones is feasible**
- **Here we assess the suitability of 21 mRNA markers for blood deposition timing**
- **We present 11 rhythmic mRNAs from analysing blood collected every 2 h for 36 h**
- **Three mRNAs, melatonin, cortisol allow estimating 3 time categories with AUC 0.88-0.95**

## Abstract

Determining the time a biological trace was left at a scene of crime reflects a crucial aspect of forensic investigations as - if possible - it would permit testing the sample donor's alibi directly from the trace evidence, helping to link (or not) the DNA-identified sample donor with the crime event. However, reliable and robust methodology is lacking thus far. In this study, we assessed the suitability of mRNA for the purpose of estimating blood deposition time, and its added value relative to melatonin and cortisol, two circadian hormones we previously introduced for this purpose. By analysing 21 candidate mRNA markers in blood samples from 12 individuals collected around the clock at 2 h intervals for 36 h under real-life, controlled conditions, we identified 11 mRNAs with statistically significant expression rhythms. We then used these 11 significantly rhythmic mRNA markers, with and without melatonin and cortisol also analysed in these samples, to establish statistical models for predicting day/night time categories. We found that although in general mRNA-based estimation of time categories was less accurate than hormone-based estimation, the use of three mRNA markers HSPA1B, MKNK2 and PER3 together with melatonin and cortisol generally enhanced the time prediction accuracy relative to the use of the two hormones alone. Our data best support a model that by using these five molecular biomarkers estimates three time categories, i.e., night/early morning, morning/noon, and afternoon/evening with prediction accuracies expressed as AUC values of 0.88, 0.88, and 0.95, respectively. For the first time, we demonstrate the value of mRNA for blood deposition timing and introduce a statistical model for estimating day/night time categories based on molecular biomarkers, which shall be further validated with additional samples in the future. Moreover, our work provides new leads for molecular approaches on time of death estimation using the significantly rhythmic mRNA markers established here.

**Keywords:** forensic time estimation; blood deposition time; circadian hormones; rhythmic gene expression; mRNA

## Introduction

The crucial question regarding the time when a human trace was left at a scene of crime is frequently encountered during criminal investigations, alongside individual identification of the trace donor. While the principle, the markers, the methodology, and the interpretation of DNA-based individual identification are all well established and routinely applied to forensic casework around the world, estimation of trace deposition time lacks reliable and robust techniques as of yet. If possible however, it would permit testing the sample donor's alibi directly from analysing the trace evidence, which would allow linking (or not) the DNA-identified sample donor with the crime event, and thus is of importance for solving a forensic case. Furthermore, molecular alibi testing from crime scene traces could provide information useful for finding unknown sample donors.

Currently, research on estimating trace deposition time is mainly focussed on determining the age of a human forensic stain i.e., how much time has passed since the human material was left at the scene of a crime. Reflectance spectroscopy and biochemical methods have been suggested for estimating the age of blood stains [1,2,3], as well as the principle of differential, time-dependent RNA degradation [4,5]. A second aspect of trace timing i.e., determining when during the 24 h day/night cycle the human material was left at the crime scene, was introduced in the last years by employing knowledge of circadian biology [6].

Circadian rhythms are endogenous oscillations with an approximately 24 h period that govern the daily lives of most organisms, humans included. These intrinsic rhythms are generated by an autoregulatory negative-feedback loop that is formed by a set of core clock genes, such as Circadian Locomotor Output Cycles Kaput (*CLOCK*), Brain and Muscle ARNT-Like 1 (*BMAL1*), Cryptochrome (*CRY*) and Period (*PER*), together with their respective proteins [7-10]. These oscillations, through regulation of expression of other genes, so called clock-

controlled genes (CCGs), are reflected in various biological processes like hormone secretion, metabolic reactions, behaviour, and many others [10-13]. It was shown that approximately 3-10% of all mRNAs in a respective tissue exhibit diurnal variations in expression, and therefore can be considered as CCGs [14-16]. Thus, the ubiquitous nature of these rhythms presents a vast amount of potential molecular biomarkers, which in principle are suitable for application in forensic trace deposition timing as well as time of death estimation under certain prerequisites [17].

The first study describing the application of the circadian hormone melatonin for forensic time estimation i.e., in determining the time of death was published in 1994 [18], and over the next years more studies focusing on time estimations with rhythmic biomarkers were reported [6,19,20,21]. Estimating time of death is similar to estimating trace deposition time, and assessment of both can in principle be done with rhythmic biomarkers, provided that they are stable under post-mortem/post-deposition conditions. In 2010, we reported a proof-of-concept study on the use of two circadian hormones, melatonin and cortisol, for determining the deposition time of blood and saliva [6], describing for the first time trace deposition timing from the chronobiological perspective. The proposed approach and methodology demonstrated the feasibility of reproducing circadian profiles of two hormones in blood and saliva samples, and highlighted the advantages for forensic applications. For instance, this method requires a small sample volume typically encountered in forensic casework, and the assays, as well as the laboratory equipment, are commercially available. Furthermore, no or only limited signs of in-vitro time-wise degradation, as prerequisite for using these markers for trace deposition timing, were observed for these hormones. However, the day/night time range that can be estimated with melatonin and cortisol alone is limited [6]. Moreover, the effect of external and internal factors

on both of the circadian hormones is recognized and, if disregarded, can be a cause of difficulties in result interpretation under particular circumstances [6]. For example, melatonin secretion is inhibited by exposure to light, in a dose-dependent manner [22,23]. In normal subjects melatonin suppression starts between 200 – 400 lux (equivalent to ordinary fluorescent light), however upon light removal, melatonin concentration returns to normal night time levels [22]. In another study, it has been shown that acute suppression of melatonin secretion occurs after exposure to intensive light (600 lux or higher) for an hour [22,23]. Furthermore, a disruption in melatonin's circadian pattern has been noted in subjects suffering from mental disorders such as major depressive disorder [24]. Cortisol levels have been shown to be disrupted in individuals suffering from addiction, chronic stress, or posttraumatic stress disorder (PTSD) [25-28]. Because of such factors influencing these two circadian hormones, and due to the limited time resolution they provided when being applied to trace deposition timing, additional rhythmic molecular biomarkers are required for increasing the accuracy as well as the reliability of molecular means for trace deposition timing.

In recent years, RNA profiling for forensic purposes has become more enthusiastically explored, and many studies describe its utility, especially for forensic tissue and body fluid identification [29-34] and, less so, for post-mortem interval determination [35,36]. In 2011 and 2012, applications of rhythmic mRNA and microRNA markers for the time of death determination were reported [19,37]. Kimura et al. [191] analysed the expression levels of three circadian genes – *BMAL1*, *PER2* and *REV-ERBA* (also known as *NR1D1* – Nuclear Receptor Subfamily 1, Group D, Member 1) – in kidney, liver and heart samples obtained from forensic autopsy material. Based on gene expression values, the authors constructed a range of ratios used for time of death estimation. Odriozola et al. [37] reported two microRNAs, miR-541 and miR-

142-5p, with diurnal variations in their abundance in vitreous humour samples from deceased individuals, further proposed as suitable candidate markers for time of death estimation. Recently, we demonstrated that these two microRNA markers are not suitable for blood trace deposition timing [38]. In this study, miR-541 was shown to be present in very low levels in blood, not allowing for meaningful conclusions, whereas miR-142-5p was not rhythmic in the tested blood samples [38]. Among other reasons, these findings may be explained by possible tissue specificity of miR-541 and miR-142-5p expression, which should be further explored.

Because in all those previous studies only a small number of biomarkers were tested, this ultimately limited the precision and significance of the obtained time estimates [6,19,37]. Expectedly, a larger set of biomarkers is needed, in order to achieve reliable and narrow time predictions, suitable for forensic applications. In the present study, we provide the first attempt for assessing the suitability of mRNA markers for estimating blood deposition time. Recently, we analysed the expression of 12 well-known clock and clock-related genes [39] and of 9 candidate clock-controlled genes [40] and measured the concentration of melatonin and cortisol, in blood samples drawn from 12 individuals around the clock at 2 h intervals for 48 h under controlled conditions of a sleep/sleep deprivation (S/SD) study protocol, and under a separate constant routine study protocol (CR) [39-41]. Data analysis in these previous biologically-motivated studies focused on identifying diurnal and circadian genes, understanding their biological function, and assessing the influence of sleep and no-sleep on gene expression.

In the present forensically-motivated study, we used the raw expression data of the 21 genes as well as melatonin and cortisol from the 2 hourly collected samples during the first 36 hours of the S/SD study. Here, we did not consider data from the samples collected during the one night of sleep deprivation in the S/SD study and did not use any data from the samples

collected in the CR study. Both sampling scenarios represent non-natural conditions not relevant for the present study, where we like to simulate real-life conditions as much as possible for selecting time predictive mRNA markers for future forensic applications. Based on these data, we selected the statistically significant rhythmic mRNA markers, and performed multinomial logistic regression modelling with and without melatonin and cortisol for predicting day/night time categories.

## **Materials and Methods**

### *Gene expression data*

The gene expression data used in this study is part of a larger data set we previously generated from the blood samples collected during the Sleep/Sleep Deprivation Study (S/SD) conducted at Surrey Clinical Research Centre (CRC) at the University of Surrey, UK. Full details of the study protocol, eligibility criteria, and the data acquisition procedure were reported elsewhere [39-41]. However, for the present analysis we used 18 two-hourly blood samples per each of 12 male participants (mean age  $\pm$  SD = 23  $\pm$  5 years) i.e., 216 samples in total. These samples spanned the first 36 h of the S/SD study (from 12:00 h Day 2 to 22:00 h Day 3), excluding the sleep deprivation condition (00:00 h Day 3 to 12:00 h Day 4). The reason for selecting this sample set was that while conditions such as lighting intensity, food intake, posture and physical activity etc., were controlled throughout the study, the in-laboratory day/night layout still resembled that of real life [39-41]. We also did not consider the data previously generated from the samples collected under the constant routine (CR) protocol [40], as they do not represent real-life conditions. For blood sample collection procedure, RNA extraction method, cDNA synthesis, qPCR data and subsequent analyses we refer to the Method sections of the two previous articles



[40,41]. Expression data from the following mRNA markers were used: Brain and Muscle ARNT-Like 1 (*BMALI*) [39,42,44], Circadian Locomotor Output Cycles Kaput (*CLOCK*) [39,45], Cryptochrome 1 (*CRY1*) [39,45], Cryptochrome 2 (*CRY2*) [39,45], D site of albumin promoter (albumin D-box) binding protein (*DBP*) [39], Deleted In Esophageal Cancer 1 (*DEC1*) [39,43,45], Heat shock 70kDa protein 1B (*HSPA1B*) [39], Period 1 (*PER1*) [39,43-46], Period 2 (*PER2*) [39,42-45], Period 3 (*PER3*) [39,42,43,45], Nuclear receptor subfamily 1, group D, member 1 (*REV-ERBA*) [39], RAR-related orphan receptor alpha (*RORA*) [39] and of Cell Cycle Associated Protein 1 (*CAPRIN1*) [14,40], MAP Kinase Interacting Serine/Threonine Kinase 2 (*MKNK2*) [40,46], Rho-Associated, Coiled-Coil Containing Protein Kinase 2 (*ROCK2*) [40,47], Sirtuin 1 (*SIRT1*) [40,48], Sterol Regulatory Element-Binding Transcription Factor 1 (*SREBF1*) [40,49], Signal Transducer and Activator of Transcription 3 (*STAT3*) [9,40,49], Thyroid Hormone Receptor Alpha (*THRA1*) [40,50], Tribbles Homolog 1 (*TRIB1*) [40,51], Upstream Transcription Factor 1 (*USF1*) [40,52], and Actin Beta (*ACTB*), used here as the reference gene. Melatonin and cortisol concentration measurements, obtained in the same samples [39,41] were used as well.

In short, gene expression data were analysed with the delta-delta-cycle-threshold ( $\Delta\Delta$  CT) method [53], and afterwards with the single cosinor and nonlinear curve fitting (nlcf) and nonlinear mixed model (nlm) methods, to determine the presence of 24 h rhythmicity, as described previously [39-41]. Selection of genes for time category prediction was based on the statistically significant outcomes from the nlcf, nlm, and single cosinor methods. The requirements to be met were statistically significant outcomes of either the nlcf (for 2 out of 3 conditions: sleep, sleep deprivation and collapsed) or nlm methods, and presence of statistically significant rhythms in at least 25% of tested individuals (single cosinor method). Afterwards,

based on the mean peak time estimates (obtained via either nlcf or nlm method) of the selected genes were used for establishing the most suitable time categories that were subsequently used in the prediction modelling.

### *Model building and time predictions*

Prediction models were constructed based on multinomial logistic regression, where the *ACTB*-normalized expression levels of the genes and the concentration values of the hormones were considered as the predictors, and the multinomial time categories as the response variable, similar as described elsewhere [54]. Note that for model building the expression levels of gene markers and hormone measurements were not z-scored, because the z-scoring will not be possible for evidentiary samples in future forensic application. Multinomial logistic regression was used to predict the probabilities of different possible outcomes of a categorically distributed dependent variable, given a set of independent variables, as previously applied for prediction of eye and hair colour categories based on SNP genotypes [54,55]. Besides logistic regression, there is an array of well-established statistics or machine-learning techniques for prediction modeling, such as linear discriminant analysis [56] and support vector machines [57]. It has been shown that different methods often perform similarly in the work of eye colour prediction [54]. We chose multinomial logistic regression here for its simplicity (only regression betas needed), portability (compatible to all statistical platforms), and robustness (without the fundamental assumption on normality of explanatory variables). The most suitable time categories used in the prediction modelling were selected by considering the mean peak time estimates of the selected genes and hormones. We then tested different combinations of molecular predictors: hormones alone, genes alone, and hormones and genes together. Selection of the final set of molecular

predictors was based on their contribution to the prediction accuracy, and mRNA markers with an insignificant effect were removed from the models, which were then rebuilt.

Due to the relatively small sample size, the performances of the prediction models were evaluated using the leaving-one-out cross-validation (LOOCV) approach [58] i.e., building the prediction model from 215 observations and predicting the day/night category for the remaining observation. This procedure was then repeated 216 times (one time for each observation). The area under the receiver operating characteristic (ROC) curves, the AUC, was measured. Its values range from 0.5, representing random prediction to 1.0, representing perfect prediction. The concordance between the predicted and observed categories was categorized into 4 groups: true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). We derived 4 accuracy parameters: sensitivity =  $TP/(TP+FN)*100$ , specificity =  $TN/(TN+FP)*100$ , positive predictive value (PPV) =  $TP/(TP+FP)*100$ , and negative predictive value (NPV) =  $TN/(TN+FN)*100$ . Note that, although the 216 observations were not completely independent from each other, the prediction results are unlikely to be biased, since all prediction results were cross validated. Because the variability (such as confidence intervals) cannot be directly derived from LOOCV results, we conducted a permutation analysis to estimate the variability of our accuracy estimates under the null hypothesis, i.e., assuming no relationship between predictors and time categories, by randomly reshuffling the dependent variable (time categories,  $k=1000$ ). For each shuffling, we derived the accuracy estimates based on the above described multinomial logistic regression and LOOCV. The resultant average values and 5%-95% quantiles (Table S1) provide unbiased estimates about the variability of our accuracy estimates under the null that is specific to our data set.

## Results and Discussion

### *Choosing mRNA predictors of blood deposition time*

From the 21 genes tested for day/night rhythms in expression (Table 1), we identified 11 genes i.e., *BMAL1*, *CAPRIN1*, *HSPA1B*, *MKNK2*, *PER1*, *PER3*, *ROCK2*, *SIRT1*, *STAT3*, *THRA1* and *TRIB1* that showed statistically significant results according to either the nlcf or nlm method, and at the same time had statistically significant rhythms in at least 25% of the individuals according to the cosinor method (Table 1). Their expression patterns are presented in Figure 1.

Next, we assigned the 11 significantly rhythmic mRNA markers, as well as melatonin and cortisol, to day/night time categories based on their mean expression/concentration peak time estimates. We found that peak times of several of the 13 molecular biomarkers were overlapping. In particular, expression of *PER1* and cortisol concentration were highest in the same time range i.e., between 08:00 h and 10:00 h, whereas expression of *ROCK2*, *SIRT1*, and *THRA1* as well as melatonin concentration were highest between 01:00 h and 03:00 h. The expression of *PER3* was highest around 04:00 h, and the expression of *BMAL1*, *HSPA1B*, *TRIB1*, *MKNK2*, and *STAT3* were highest in the afternoon between 15:30 h and 17:00 h. Consequently, we established three time categories of peak expression/concentration of these 13 biomarkers, i.e., night/early morning (23:00 – 06:59), morning/noon (07:00 – 14:59) and afternoon/evening (15:00 – 22:59), together comprising one complete day/night cycle.

### *Modelling blood deposition time using molecular biomarkers*

We then performed statistical prediction modelling of these three day/night time categories using the 11 mRNA markers (*BMAL1*, *CAPRIN1*, *HSPA1B*, *MKNK2*, *PER1*, *PER3*, *ROCK2*, *SIRT1*, *STAT3*, *THRA1* and *TRIB1*) and the 2 hormone markers (melatonin and cortisol)

in different combinations (see Methods) to establish the most accurate and robust model, while minimizing the number of biomarkers involved. The latter was done by keeping the final forensic application in mind, where typically the amount of biological material is limited, and thus is the number of molecular tests possible before the evidence material is exhausted. We tested the genes with and without the hormones, as well as the hormones alone to work out the relationship between these two types of molecular biomarkers in the time prediction modelling.

We noticed that some of the mRNA markers used for model building did not contribute significantly to the prediction accuracy obtained with the models. From this, we concluded that their effect was “masked” by other, more robust markers included in the respective model. The markers with negligible independent input into the prediction accuracy were consequently excluded from the respective models to keep the number of markers at the minimum, which were then rebuilt with only considering the independently contributing molecular predictors.

In the model considering all 11 significantly rhythmic mRNA markers, 5 were excluded due to redundant predictive effects, while 6 mRNA markers with independent predictive effects were identified i.e., *HSPA1B*, *MKNK2*, *PER1*, *PER3*, *THRA1*, and *TRIB1*. The model based on these 6 independent mRNA predictors achieved AUC values of 0.75 for morning/noon, 0.80 for afternoon/evening, and 0.93 for night/early morning (Table 2). A model considering only melatonin and cortisol provided lower AUC for the night/early morning (0.85) but somewhat higher AUCs for the afternoon/evening (0.83) and morning/noon (0.85) time categories compared with the model based on 6 mRNA markers (Table 2). These results indicate that at least for the night/early morning category the use of mRNA markers is largely beneficial. The results for the hormone-based model were not surprising, as the two hormones are characterized by robust, truly circadian patterns of secretion with sharp, distinct peaks and high amplitude

[39,59,60], and melatonin is an established marker of circadian phase, also used in sleep-related studies [39,61]. It is clear that all AUC values from all models are far beyond the corresponding 95% upper boundaries (max  $AUC_{NULL}=0.58$ ) from a permutation analysis under the null hypothesis assuming no relationship between predictors and time category (Table S1). Additionally, regression parameters from the final models are provided in Table S2 in the Supplementary Material.

Because of the higher AUCs for one but the lower AUCs for the other time categories as achieved with the genes and the hormones separately, we then combined all 11 significant mRNA markers and the 2 hormone markers in a prediction model and tested for redundant biomarker effects. As a result, 8 (*BMAL1*, *CAPRINI*, *PER1*, *ROCK2*, *SIRT1*, *STAT3*, *THRA1* and *TRIB1*) out of the 11 mRNA markers were removed from the model.

The final model comprising of five biomarkers (*MKNK2*, *HSPA1B*, *PER3*, melatonin and cortisol) predicted the night/early morning category with particularly high accuracy AUC of 0.95, while the other two time categories were both predicted with an AUC of 0.88 (Table 2). Compared with the model based solely on the two circadian hormones, and with the model based solely on mRNA markers, the combined hormone/mRNA model achieved higher accuracies for all three predicted time categories (Table 2). Hence, the use of mRNA markers, particularly *HSPA1B*, *MKNK2* and *PER3*, did overall increase the day/night time prediction accuracy, compared to those achieved with hormones only, underlining the beneficial use of mRNA markers for blood deposition timing.

*Future forensic application considerations*

We chose to conduct our trace timing analyses in blood mainly because bloodstains are amongst the most commonly found biological evidence on scenes of violent crimes. It is important to keep in mind that without additional testing in other forensically relevant tissues, the results we present here are to be regarded as specific for blood. Even though blood has been proposed as sort of a “gate to access and analyse the transcriptome of various organs” [62], it should be noted that our findings would need to be properly revalidated in other forensically relevant tissue types, such as saliva, semen, skin, vaginal secretion, and menstrual blood, before applying them to such forensic traces. Besides expanding molecular trace deposition timing to other forensically relevant tissue types than blood, additional testing in various human tissues would also be advantageous in determining whether the mRNA markers proposed here are also informative for time of death estimations, as previously suggested for melatonin [18].

Another crucial aspect that shall be tested carefully in future studies is the time-wise stability of the proposed mRNA markers. Since blood stains are found at various crime scenes, they are exposed to a multitude of variable conditions, i.e., high/low humidity, drought, temperature changes, but are also located on different types of surfaces. All these factors can possibly influence the stability of the mRNA, which should be tested for the specific mRNA markers proposed here. Many mRNA markers previously suggested for forensic tissue identification purposes have shown strong time-wise stability [29,32], which has to be demonstrated for the candidate mRNA markers we suggest here for blood timing before they can be introduced to molecular alibi testing in forensic case-work.

Finally, our results were established from 12 male individuals, and additional samples shall be analysed in the future to further validate our prediction model.

## Conclusions

In this study, we investigated whether mRNA provides a suitable resource for establishing biomarkers to estimate blood deposition time. We demonstrated that particular mRNA markers have added value for blood deposition timing over the previously established two circadian hormones melatonin and cortisol. We introduced a prediction model comprising three mRNA markers, *HSPA1B*, *MKNK2* and *PER3*, together with two circadian hormones melatonin and cortisol, provides improved prediction accuracy of three day/night categories compared to those achieved with a model based on the two hormones alone or mRNA markers alone. To our knowledge this is the first study assessing the suitability of mRNA markers for trace deposition timing and the first time a statistical model for estimating blood deposition time with molecular biomarkers is presented. To achieve a more detailed level of time category prediction than revealed here, additional rhythmic biomarkers with different peak times will be needed. Future studies should focus on identifying them, and eventually incorporating them together with the markers used here to develop a final prediction model. Moreover, our work provides new leads for future studies on time of death investigation using the significantly rhythmic mRNA markers established here, which represents a second aspect of forensic time estimation in need of improved biomarkers, methodology, and technology.



Conflicts of interest: none

### **Acknowledgements**

The authors thank the Surrey CRC medical, clinical and research teams for their help with the study conduct and sample collection. This study was supported by grant 27.011.001 by the Netherlands Organization for Scientific Research (NWO) Forensic Science Program, Erasmus MC University Medical Center Rotterdam, by the EU 6th Framework project EUCLOCK (018741), by a UK Biotechnology and Biological Sciences Research Council (BBSRC) Grant BB/I019405/1, and by a previous grant from the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands (FGCN). D.J.S. is a Royal Society Wolfson Research Merit Award holder.

## References

- [1] Bremmer RH, de Bruin KG, van Gemert MJC, van Leeuwen TG, Aalders MCG. Forensic quest for age determination of bloodstains. *Forensic Sci Int.* 2012;216:1-11.
- [2] Arany S, Ohtani S. Age estimation of bloodstains: a preliminary report based on aspartic acid racemization rate. *Forensic Sci Int.* 2011;212(1-3):e36-9.
- [3] Bremmer RH, Nadort A, van Leeuwen TG, van Gemert MJ, Aalders MC. Age estimation of blood stains by hemoglobin derivative determination using reflectance spectroscopy. *Forensic Sci Int.* 2011;206(1-3):166-71.
- [4] Bauer M, Polzin S, Patzelt D. Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: a possible indicator of the age of bloodstains? *Forensic Sci. Int.* 2003;138:94–103.
- [5] Hampson C, Louhelainen J, McColl S. An RNA expression method for aging forensic hair samples. *J Forensic Sci.* 2011;56(2):359-65.
- [6] Ackermann K, Ballantyne KN, Kayser M. Estimating trace deposition time with circadian biomarkers: a prospective and versatile tool for crime scene reconstruction. *Int J Legal Med.* 2010;124:387-95.
- [7] Yagita K, Tamanini F, van Der Horst GT, Okamura H. Molecular mechanisms of the biological clock in cultured fibroblasts. *Science.* 2001;292(5515):278-81.
- [8] Okamura H. Clock genes in cell clocks: roles, actions, and mysteries. *J Biol Rhythms.* 2004;19(5):388-99.
- [9] Hughes ME, DiTacchio L, Hayes KR, Vollmers C, Pulivarthy S, Baggs JE et al. Harmonics of circadian gene transcription in mammals. *PLoS Genet.* 2009;5(4):e1000442.

- [10] Mohawk JA, Green CB, Takahashi JS. Central and Peripheral Circadian Clocks in Mammals. *Annu Rev Neurosci.* 2012;35:445-62.
- [11] Kalsbeek A, la Fleur S, Fliers E. Circadian control of glucose metabolism. *Mol Metab.* 2014;3(4):372-83.
- [12] Eckel-Mahan K, Sassone-Corsi P. Metabolism and the circadian clock converge. *Physiol Rev.* 2013;93(1):107-35.
- [13] Keller M, Mazuch J, Abraham U, Eom GD, Herzog ED, Volk HD et al. A circadian clock in macrophages controls inflammatory immune responses. *Proc Natl Acad Sci USA.* 2009;106(50):21407-12.
- [14] Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M et al. Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell.* 2002;109(3):307-20.
- [15] Storch KF, Lipan O, Leykin I, Viswanathan N, Davis FC, Wong WH, et al. Extensive and divergent circadian gene expression in liver and heart. *Nature* 2002;417: 78–83.
- [16] Cermakian N, Boivin DB. The regulation of central and peripheral circadian clocks in humans. *Obes Rev.* 2009;10:25-36.
- [17] Kayser M, de Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet.* 2011;12:179-92.
- [18] Mikami H, Terazawa K, Takatori T, Tokudome S, Tsukamoto T, Haga K. Estimation of Time of Death by Quantification of Melatonin in Corpses. *Int J Legal Med.* 1994;107:42-51.
- [19] Kimura A, Ishida Y, Hayashi T, Nosaka M, Kondo T. Estimating time of death based on the biological clock. *Int J Legal Med.* 2011;125:385-91.
- [20] Minami Y, Kasukawa T, Kakazu Y, Iigo M, Sugimoto M, Ikeda S, et al. Measurement of internal body time by blood metabolomics. *Proc Natl Acad Sci USA.* 2009;106:9890-5.

- [21] Kasukawa T, Sugimoto M, Hida A, Minami Y, Mori M, Honma S, et al. Human blood metabolite timetable indicates internal body time. *Proc Natl Acad Sci USA*. 2012;109:15036-41.
- [22] Lewy AJ, Wehr TA, Goodwin FK, Newsome DA, Markey SP. Light suppresses melatonin secretion in humans. *Science*. 1980;210(4475):1267-9.
- [23] McIntyre IM, Norman TR, Burrows GD, Armstrong SM. Melatonin, cortisol and prolactin response to acute nocturnal light exposure in healthy volunteers. *Psychoneuroendocrinology*. 1992;17:243–24.
- [24] Frazer A, Brown R, Kocsis J, Caroff S, Amsterdam J, Winokur A, Sweeney J, Stokes P. Patterns of melatonin rhythms in depression. *J Neural Transm Suppl*. 1986;21:269-90.
- [25] Ockenfels MC, Porter L, Smyth J, Kirschbaum C, Hellhammer DH, Stone AA. Effect of chronic stress associated with unemployment on salivary cortisol: overall cortisol levels, diurnal rhythm, and acute stress reactivity. *Psychosom Med*. 1995;57(5):460-7.
- [26] Chrousos GP. Stress and disorders of the stress system. *Nat Rev Endocrinol*. 2009;5(7):374-81.
- [27] Fries E, Dettenborn L, Kirschbaum C. The cortisol awakening response (CAR): facts and future directions. *Int J Psychophysiol*. 2009;72(1):67-73.
- [28] Lovallo WR. Cortisol secretion patterns in addiction and addiction risk. *Int J Psychophysiol*. 2006;59(3):195-202.
- [29] Zubakov D, Hanekamp E, Kokshoorn M, van IJcken W, Kayser M. Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of time-wise degraded samples. *Int J Legal Med*. 2008;122:135-42.

- [30] Lindenbergh A, de Pagter M, Ramdayal G, Visser M, Zubakov D, Kayser M, Sijen T. A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces. *Forensic Sci Int Genet.* 2012;6(5):565-77.
- [31] Sijen T. Molecular approaches for forensic cell type identification: On mRNA, miRNA, DNA methylation and microbial markers. *Forensic Sci Int Genet.* 2014;S1872-4973(14)00263-4.
- [32] Zubakov D, Kokshoorn M, Kloosterman A, Kayser M. New markers for old stains: stable mRNA markers for blood and saliva identification from up to 16-year-old stains. *Int J Legal Med.* 2009;123:71-4.
- [33] Juusola J, Ballantyne J. Messenger RNA profiling: a prototype method to supplant conventional methods for body fluid identification. *Forensic Sci Int.* 2003;135(2):85-96.
- [34] Juusola J, Ballantyne J. mRNA profiling for body fluid identification by multiplex quantitative RT-PCR. *J Forensic Sci.* 2007;52(6):1252-62.
- [35] Bauer M, Gramlich I, Polzin S, Patzelt D. Quantification of mRNA degradation as possible indicator of postmortem interval - a pilot study. *Leg Med (Tokyo).* 2003;5(4):220-7.
- [36] Sampaio-Silva F, Magalhães T, Carvalho F, Dinis-Oliveira RJ, Silvestre R. Profiling of RNA degradation for estimation of *post mortem* interval. *PLoS ONE* 2013;8(2): e56507.
- [37] Odriozola A, Riancho JA, de la Vega R, Agudo G, Garcia-Blanco A, de Cos E, et al. MiRNA analysis in vitreous humor to determine the time of death: a proof-of-concept pilot study. *Int J Legal Med.* 2013;127:573-8.
- [38] Lech K, Ackermann K, Wollstein A, Revell VL, Skene DJ, Kayser M. Assessing the suitability of miRNA-142-5p and miRNA-541 for bloodstain deposition timing. *Forensic Sci Int Genet.* 2014;12:181-184.

- [39] Ackermann K, Plomp R, Lao O, Middleton B, Revell VL, Skene DJ, et al. Effect of sleep deprivation on rhythms of clock gene expression and melatonin in humans. *Chronobiol Int.* 2013;30:901-9.
- [40] Lech K, Ackermann K, Revel VL, Skene DJ, Kayser M. Dissecting diurnal and circadian rhythms of clock-controlled genes in human blood. *J Biol Rhythms.* 2015; in press
- [41] Ackermann K, Revell VL, Lao O, Rombouts EJ, Skene DJ, Kayser M. Diurnal rhythms in blood cell populations and the effect of acute sleep deprivation in healthy young men. *Sleep.* 2012;35:933-40.
- [42] Archer SN, Viola AU, Kyriakopoulou V, von Schantz M, Dijk DJ. Inter- individual differences in habitual sleep timing and entrained phase of endogenous circadian rhythms of BMAL1, PER2 and PER3 mRNA in human leukocytes. *Sleep* 2008;31:608–17
- [43] Boivin DB, James FO, Wu A, Cho-Park PF, Xiong H, Sun ZS. Circadian clock genes oscillate in human peripheral blood mononuclear cells. *Blood* 2003;102:4143–5.
- [44] Fukuya H, Emoto N, Nonaka H, Yagita K, Okamura H, Yokoyama M. Circadian expression of clock genes in human peripheral leukocytes. *Biochem Biophys Res Commun.* 2007;354:924–8.
- [45] Kusanagi H, Hida A, Satoh K, Echizenya M, Shimizu T, Pendergast JS, Yamazaki S, Mishima K. Expression profiles of 10 circadian clock genes in human peripheral blood mononuclear cells. *Neurosci Res.* 2008;61:136–42.
- [46] Chudova D, Ihler A, Lin KK, Andersen B, Smyth P. Bayesian detection of non-sinusoidal periodic patterns in circadian expression data. *Bioinformatics.* 2009;1;25(23):3114-20.

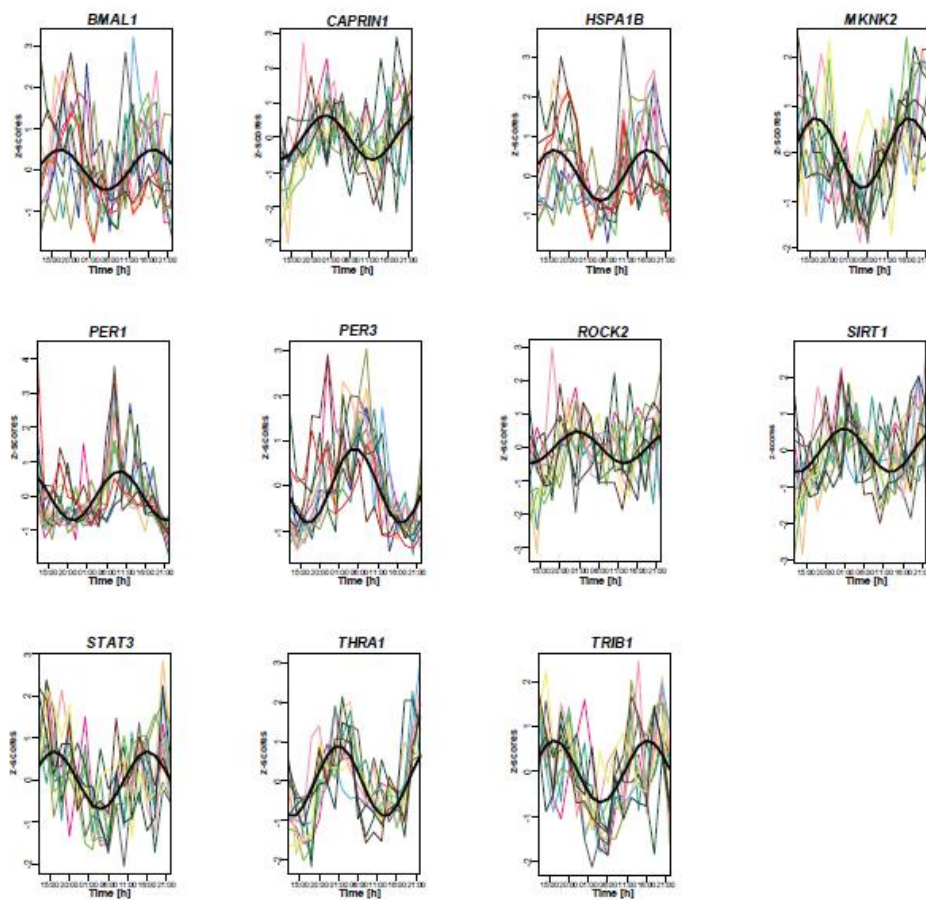
- [47] Saito T, Hirano M, Ide T, Ichiki T, Koibuchi N, Sunagawa K, et al. Pivotal role of Rho-associated kinase 2 in generating the intrinsic circadian rhythm of vascular contractility. *Circulation*. 2013;127(1):104-14.
- [48] Asher G, Gatfield D, Stratmann M, Reinke H, Dibner C, Kreppel Fet al. SIRT1 regulates circadian clock gene expression through PER2 deacetylation. *Cell*. 2008;25;134(2):317-28.
- [49] Bozek K, Relógio A, Kielbasa SM, Heine M, Dame C, Kramer A, et al. Regulation of clock-controlled genes in mammals. *PLoS One*. 2009;4(3):e4882.
- [50] Zandieh Doulabi B, Platvoet-Ter Schiphorst M, Kalsbeek A, Fliers E, Bakker O, Wiersinga WM. Diurnal variation in rat liver thyroid hormone receptor (TR)-alpha Messenger ribonucleic acid (mRNA) is dependent on the biological clock in the suprachiasmatic nucleus, whereas diurnal variation of TR beta 1 mRNA is modified by food intake. *Endocrinology*. 2004;145(3):1284-9.
- [51] Ollila HM, Utge S, Kronholm E, Aho V, Van Leeuwen W, Silander K, et al. TRIB1 constitutes a molecular link between regulation of sleep and lipid metabolism in humans. *Transl Psychiatry*. 2012;20;2:e97.
- [52] Shimomura K, Kumar V, Koike N, Kim TK, Chong J, Buhr ED, et al. Usf1, a suppressor of the circadian Clock mutant, reveals the nature of the DNA-binding of the CLOCK:BMAL1 complex in mice. *Elife*. 2013;2:e00426.
- [53] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>(-Delta Delta C(T))</sup> Method. *Methods*. 2001;25(4):402-8.
- [54] Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssens ACJW, et al. Eye color and the prediction of complex phenotypes from genotypes. *Curr Biol*. 2009;19:R192-R3.

- [55] Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A, et al. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci Int Genet.* 2013;7:98-115.
- [56] Webb AR. Linear discriminant analysis. *Statistical Pattern Recognition, Second Edition.* 2003;123-168.
- [57] Burges CJ. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery.* 1998;2(2):121-167.
- [58] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science.* 1999;286:531-7.
- [59] Arendt J. Melatonin and Human Rhythms. *Chrono Int.* 2006;23:1-2, 21-37
- [60] Weitzman ED, Fukushima D, Nogeire C, Roffwarg H, Gallagher TF, Hellman L. Twenty-four hour pattern of the episodic secretion of cortisol in normal subjects. *J Clin Endocrinol Metab.* 1971;33(1):14-22.
- [61] Moller-Levet CS, Archer SN, Bucca G, Laing EE, Slak A, Kabiljo R, et al. Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc Natl Acad Sci USA.* 2013;110:E1132-E41.
- [62] Kohane IS, Valtchinov VI. Quantifying the white blood cell transcriptome as an accessible window to the multiorgan transcriptome. *Bioinformatics.* 2012;28:538-45.



**Figure legends:**

Figure 1. Individual profiles of 11 genes/mRNA markers with significantly rhythmic expression patterns in 2-hourly collected blood samples from 12 male individuals over a period of 36h hours under controlled real-life conditions. Data are presented as *ACTB*-normalized z-scores, with individual-based colour coding. The black line represents the superimposed mean cosine curve, as calculated by the nlcf method.



**Table 1.** Results of the single cosinor, non-linear curve fitting (nlcf) or non-linear mixed model methods for all 21 genes/mRNA markers tested.

Gene/mRNA marker	SINGLE COSINOR	NON-LINEAR CURVE FITTING/NON-LINEAR MIXED MODEL
	% of significant subjects	
<u>BMAL1</u>	50	significant
<u>CAPRN1</u>	25	significant
<u>CLOCK</u>	42	significant
<u>CRY1</u>	25	not significant
<u>CRY2</u>	42	not significant
<u>DBP</u>	25	not significant
<u>DEC1</u>	0	not significant
<u>HSPA1B</u>	42	significant
<u>MKNK2</u>	42	significant
<u>PER1</u>	25	significant
<u>PER2</u>	25	not significant
<u>PER3</u>	67	significant
<u>REV-ERBa</u>	58	not significant
<u>ROCK2</u>	25	significant
<u>RORa</u>	8	not significant
<u>SIRT1</u>	42	significant
<u>SREBF1</u>	17	not significant
<u>STAT3</u>	25	significant
<u>THRA1</u>	75	significant
<u>TRIB1</u>	58	significant
<u>USF1</u>	17	not significant

Presented are the percentages of subjects with significant rhythms in expression (as calculated with the single cosinor) and outcomes of the nlcf or nlm method). Underlined are the significantly rhythmic genes selected for prediction modelling.

**Table 2.** Results of time prediction modelling using multinomial logistic regression and LOOCV.

<i>melatonin &amp; cortisol</i>					
<i>Predicted time category</i>	<i>AUC</i>	<i>sens</i>	<i>spec</i>	<i>ppv</i>	<i>npv</i>
<i>morning/noon</i>	0.85	0.66	0.85	0.68	0.84
<i>afternoon/evening</i>	0.83	0.81	0.69	0.68	0.82
<i>night/early morning</i>	0.85	0.46	0.95	0.71	0.86
<i>HSPA1B, PER1, PER3, TRIB1, THRA1, MKNK2</i>					
<i>Predicted time category</i>	<i>AUC</i>	<i>sens</i>	<i>spec</i>	<i>ppv</i>	<i>npv</i>
<i>morning/noon</i>	0.75	0.51	0.90	0.71	0.79
<i>afternoon/evening</i>	0.80	0.83	0.69	0.68	0.84
<i>night/early morning</i>	0.93	0.78	0.94	0.79	0.93
<i>MKNK2, HSPA1B, PER3, melatonin &amp; cortisol</i>					
<i>Predicted time category</i>	<i>AUC</i>	<i>sens</i>	<i>spec</i>	<i>ppv</i>	<i>npv</i>
<i>morning/noon</i>	0.88	0.73	0.91	0.80	0.87
<i>afternoon/evening</i>	0.88	0.79	0.79	0.75	0.83
<i>night/early morning</i>	0.95	0.75	0.93	0.75	0.93

*AUC* – area under the receiver operating characteristic (ROC) curves; *PPV* – positive predictive value; *NPV* – negative predictive value, *spec* – specificity; *sens* – sensitivity.