Supporting Information for

Platform-independent estimation of human physiological time from single blood samples

Yitong Huang and Rosemary Braun

Rosemary Braun
E-mail: rbraun@northwestern.edu

This PDF file includes:

Figs. S1 to S3
Tables S1 to S3
SI References

Genes	ZeitZeiger	PLSR	TimeSignature	Genes	ZeitZeiger	PLSR	TimeSignature
DDIT4	X	X	X	HMGCS1		X	
GHRL			X	FOSL2	X	X	
PER1	X	X	X	ADAT2		X	
EPHX2			X	ABCB1		X	
GNG2			X	HSBP1L1		X	
IL1B		X	X	TBX21		X	
DHRS13		X	X	ZNF101		X	
NR1D1			X	SDF2		X	
ZNF438		X	X	PHF21A	X	X	
NR1D2	X	X	X	SMAP2	X	X	
CD38			X	ZNF397		X	
TIAM2			X	MGST1		X	
CD1C		X	X	AMD1		X	
LLGL2		X	X	NUCKS1		X	
GZMB	X	X	X	FGFBP2		X	
CLEC10A			X	BACH1		X	
PDK1		X	X	UBE2B		X	
GPCPD1		X	X	MAPK8		X	
MUM1			X	SESN2		X	
STIP1		X	X	ITGA6		X	
CHSY1			X	NDFIP1		X	
AK5			X	CD82		X	
CYB561			X	RNF144B		X	
SLPI			X	IRS2	X	X	
PARP2			X	SUGT1		X	
PGPEP1			X	CDC42EP2		X	
C12orf75			X	GNL1		X	
FKBP4		X	X	ICAM1		X	
CAMKK1	X	X	X	AVIL		X	
DTYMK			X	STAU2		X	
NPEPL1			X	TXNL4B		X	
MS4A3			X	SH2D2A		X	
IL13RA1	X		X	CHN2		X	
ID3			X	GOLPH3L		X	
MEGF6		X	X	FAM127A		X	
TCN1			X	HSPA8		X	
NSUN3			X	PARD6A		X	
POLH			X	CAPN7		X	
SYT11		X	X	ABI2		X	
SH2D1B		X	X	TRIM8		X	
REM2			X	STMN3		X	
RBM3		X		PRR14		X	
FKBP5	X	X		TIGIT		X	
S1PR5		X		PPP2R2B		X	
CXCR4		X		TMEM79		X	
ZNF276		X		FUS		X	
ERLIN1		X		CASP4		X	
MKNK2		X		FAM198B		X	
SAMD3		X		HSPD1		X	
DUSP1	X	X		TIMELESS		X	
SERTAD1		X		TRA2A		X	
PDIA5		X		CIAPIN1		X	
HSPH1		X		SFXN2		X	
IL6ST		X		STEAP4		X	
TP53BP1		X		ERP27		X	
SYTL2		X		DYNLL1		X	
ODF2		X		NELL2	X		
CD36		X		CLEC4E	X		
PRRG4	X	X		CBX5	X		
CCDC117		X		FBP1	X		
C1orf21		X		HUWE1	X		
NUDT5		X		TTLL4	X		
TLR2		X		KIAA0922	X		
MED23		X		CDK11B	X		
MPPE1		X		OSBP	X		
IMPAD1		X		KPNA6	X		
SLC39A11		X		RELA	X		
CXorf65		X					

Table S1. Predictor genes identified by ZeitZeiger (1), the PLSR-based method (2), and TimeSignature (3). x marks the predictor gene identified by each algorithm when applied to the same training set from TrTe, resulting in a total union of 135 genes.

SERTAD1	ZNF101	PHF21A	STMN3	GNG2	IL1B	IL13RA1
NR1D1	IRS2	FKBP5	ID3	UBE2B	ZNF438	CLEC4E
ITGA6	DDIT4	DUSP1	STEAP4	GHRL	NR1D2	NELL2
AK5	FOSL2	DHRS13	MPPE1	CYB561	CAMKK1	PER1
AVIL	SMAP2	CDC42EP2	NUDT5	EPHX2	MKNK2	SFXN2
SESN2	PRRG4					

Table S2. A set of 37 predictor genes used in TimeMachine, which is selected by applying *JTK_Cycle* (4, 5) to the 135 genes in Supplementary Table S1.

		Proportion of samples within ± 2 hr error			Proportion of samples within ± 4 hr error		
	Dataset	rTM	zTM	PLSR	rTM	zTM	PLSR
Predicting Circadian Time	TrTe	55.7%	57.3%	47.0%	83.8%	83.8%	83.8%
	V1	39.5%	34.9%	39.5%	65.9%	66.3%	64.7%
	V3	50.3%	52.9%	40.5%	78.4%	79.7%	80.4%
Predicting Local Time	TrTe	51.2%	47.4%	35.2%	75.6%	75.1%	62.9%
	V1	39.2%	37.4%	28.0%	69.2%	66.4%	55.6%
	V2	47.4%	46.2%	37.3%	75.1%	71.1%	59.8%
	V3	41.8%	36%	41.1%	68%	59.5%	62.7%

Table S3. Performance of algorithms on predicting circadian phase and local time in four distinct validation sets.

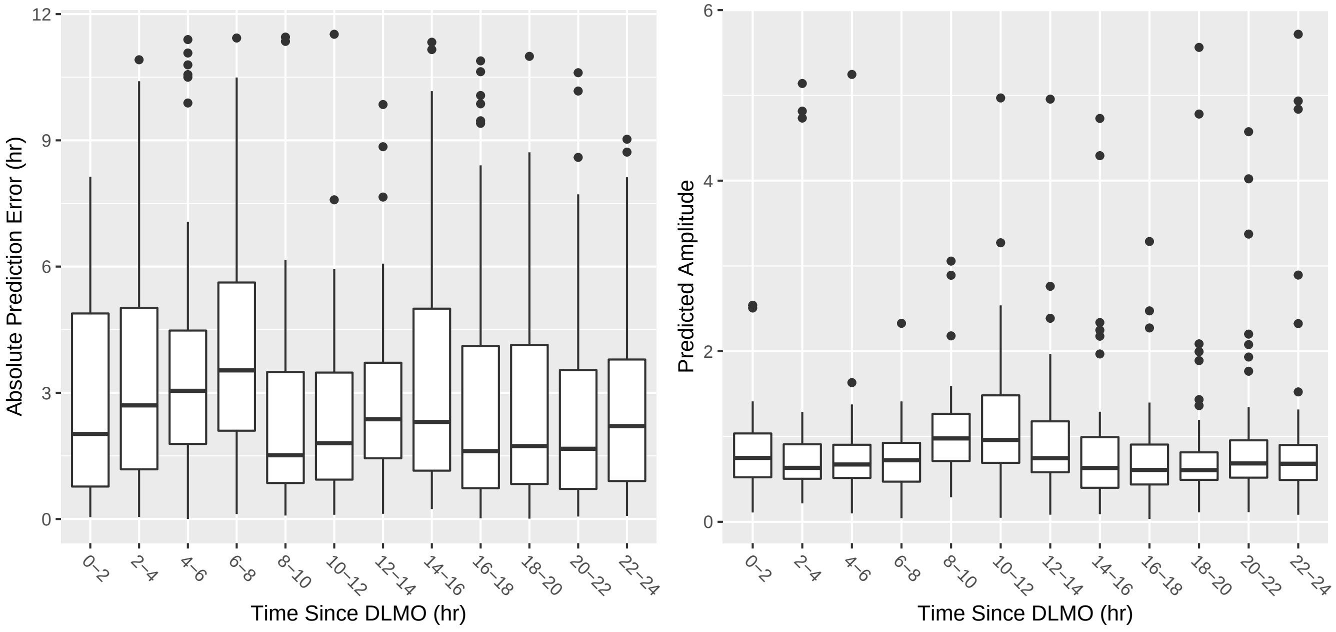

Fig. S1. ratio TimeMachine's performance with respect to time.

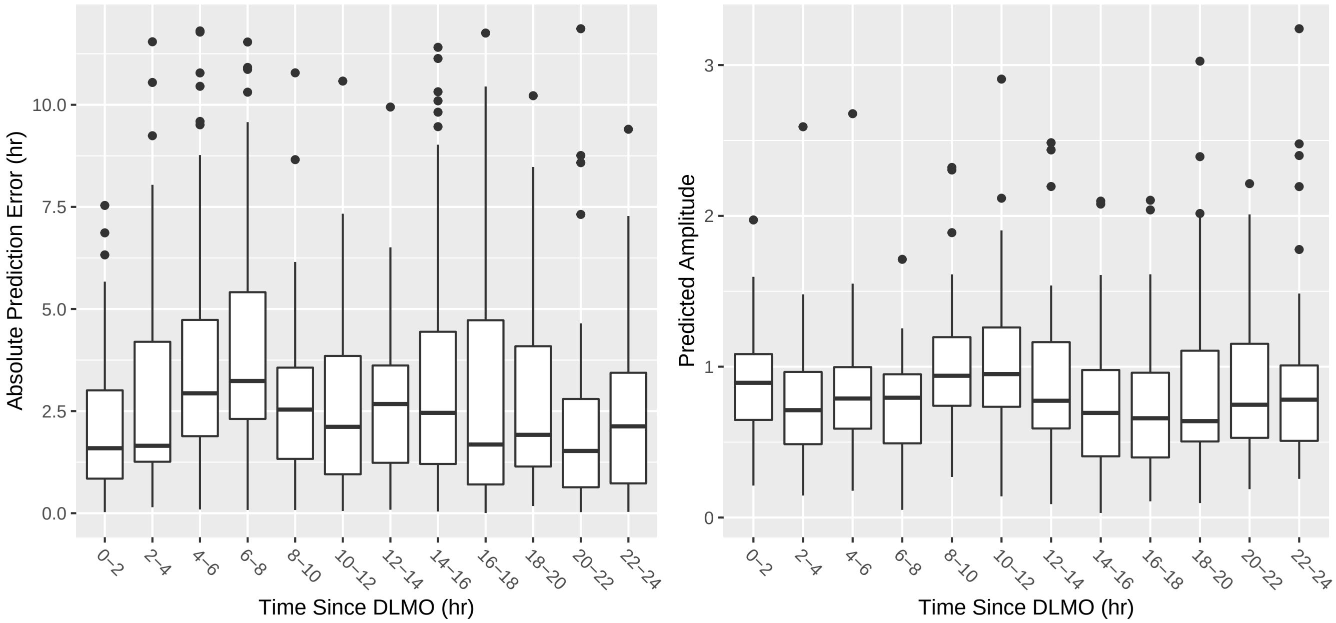

Fig. S2. Z-score TimeMachine's performance with respect to time.

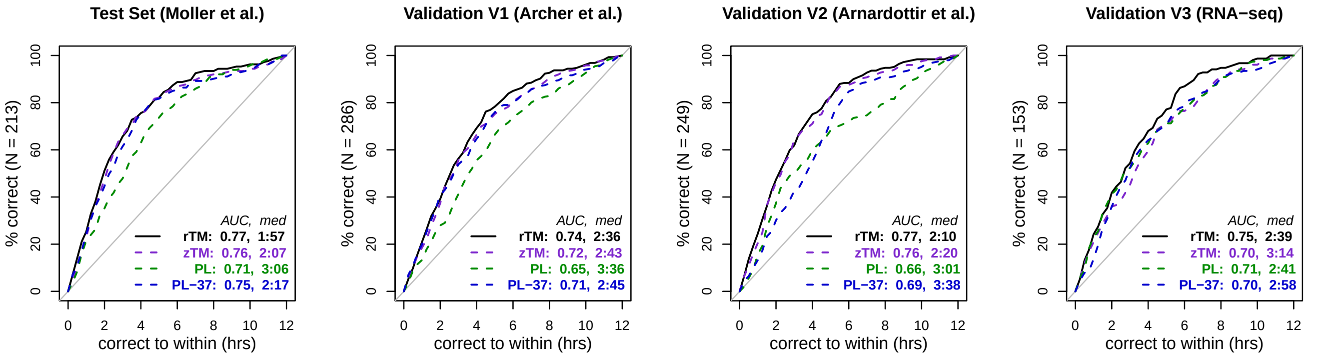

Fig. S3. Predictions of blood-draw time using different one-time methods, ratio TimeMachine (denoted by rTM), Z-score TimeMachine (denoted by zTM), PLSR with quantile normalization of 7615 genes (denoted by PL), and the revised PLSR model with Z-score of 37 predictor genes (denoted by PL-37).

References

1. JJ Hughey, Machine learning identifies a compact gene set for monitoring the circadian clock in human blood. *Genome medicine* **9**, 1–11 (2017).
2. EE Laing, et al., Blood transcriptome based biomarkers for human circadian phase. *Elife* **6**, e20214 (2017).
3. R Braun, et al., Universal method for robust detection of circadian state from gene expression. *Proc. Natl. Acad. Sci.* **115**, E9247–E9256 (2018).
4. ME Hughes, JB Hogenesch, K Kornacker, Jtk_cycle: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J. biological rhythms* **25**, 372–380 (2010).
5. G Wu, RC Anafi, ME Hughes, K Kornacker, JB Hogenesch, Metacycle: an integrated r package to evaluate periodicity in large scale data. *Bioinformatics* **32**, 3351–3353 (2016).