High-accuracy determination of internal circadian time from a single blood sample

Nicole Wittenbrink[1,2], Bharath Ananthasubramaniam[1,3], Mirjam Münch[1,4,5], Barbara Koller[1], Bert Maier[1], Charlotte Weschke[1], Frederik Bes[4,5], Jan de Zeeuw[5,6], Claudia Nowozin[4,5], Amely Wahnschaffe[4,5], Sophia Wisniewski[4,5], Mandy Zaleska[6], Osnat Bartok[7], Reut Ashwal-Fluss[7], Hedwig Lammert[8], Hanspeter Herzel[3], Michael Hummel[8], Sebastian Kadener[7,9], Dieter Kunz[4,5,6], Achim Kramer[1]

[1] Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Laboratory of Chronobiology, Berlin, Germany

[2] Humboldt-Universität zu Berlin, Department of Biology, Systems Immunology Lab, Berlin, Germany

[3] Humboldt-Universität zu Berlin, Institute for Theoretical Biology, Berlin, Germany

[4] St-Hedwig-Krankenhaus, Clinic for Sleep & Chronomedicine, Berlin, Germany

[5] Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Physiology, Group Sleep Research and Clinical Chronobiology, Berlin, Germany.

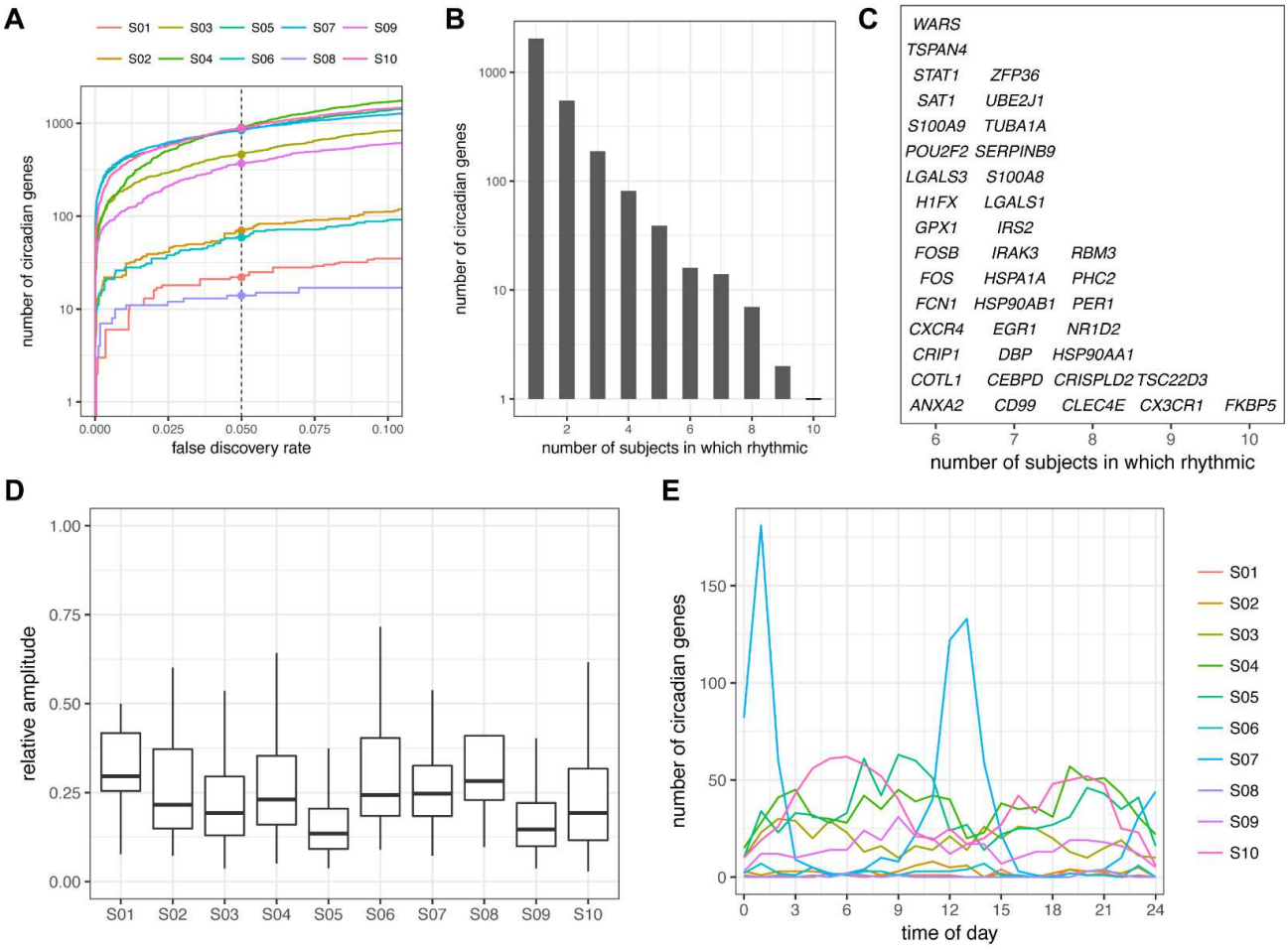
[6] Intellux Berlin GmbH, Berlin, Germany

[7] The Hebrew University, Biological Chemistry Department, Jerusalem, Israel

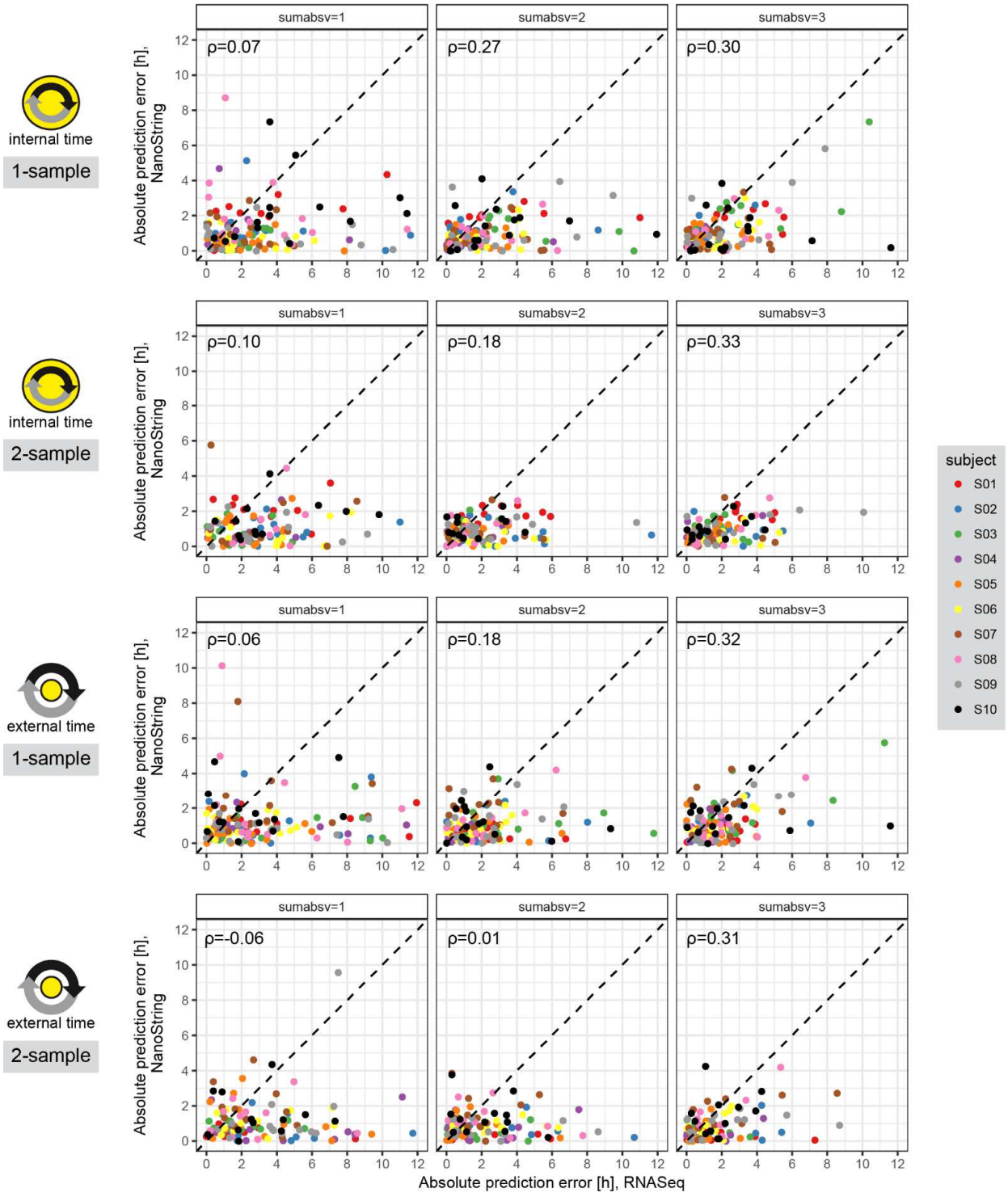
[8] Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Pathology, Berlin, Germany

[9] Brandeis University, Department of Biology, Waltham, MA, USA

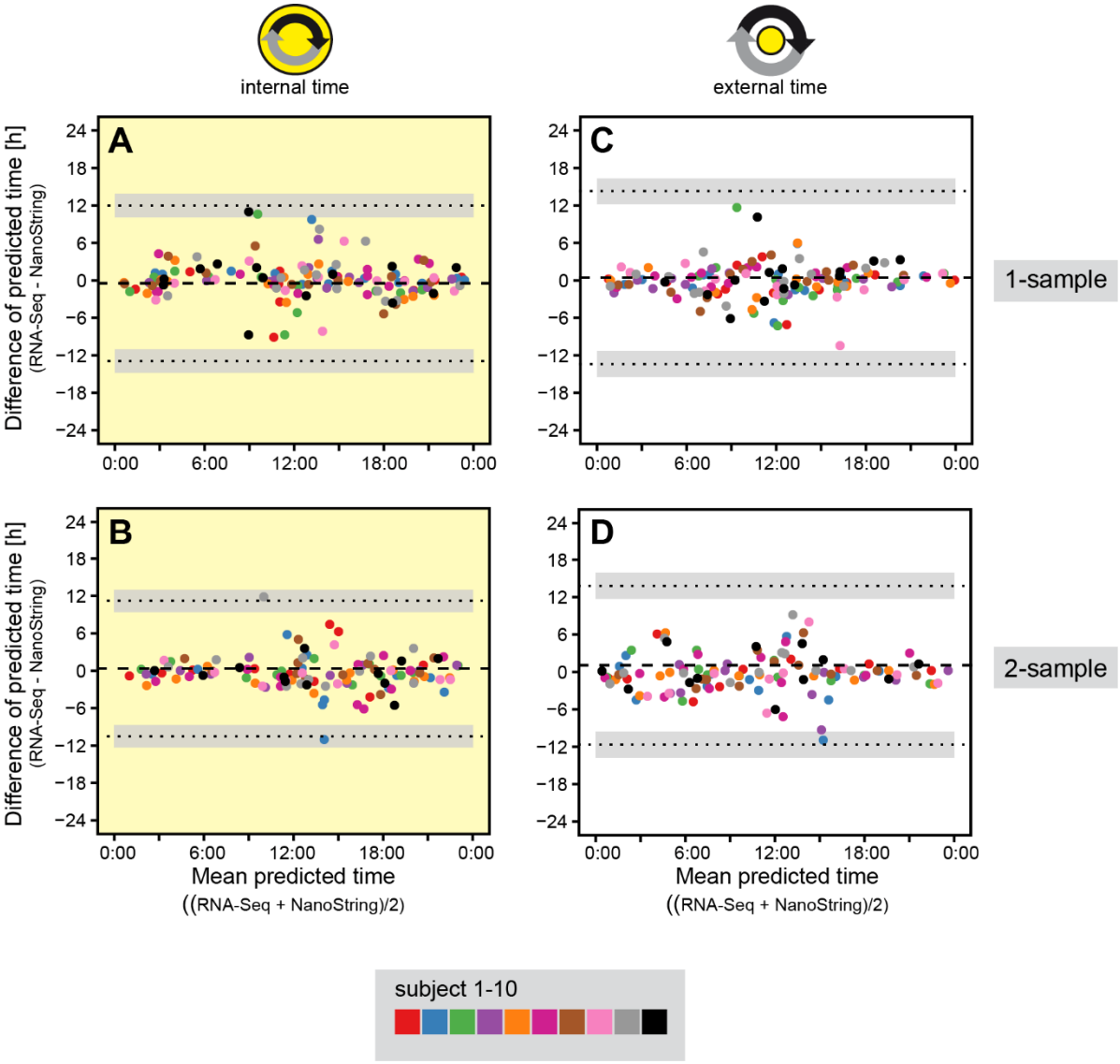
Supplemental Figures and Tables



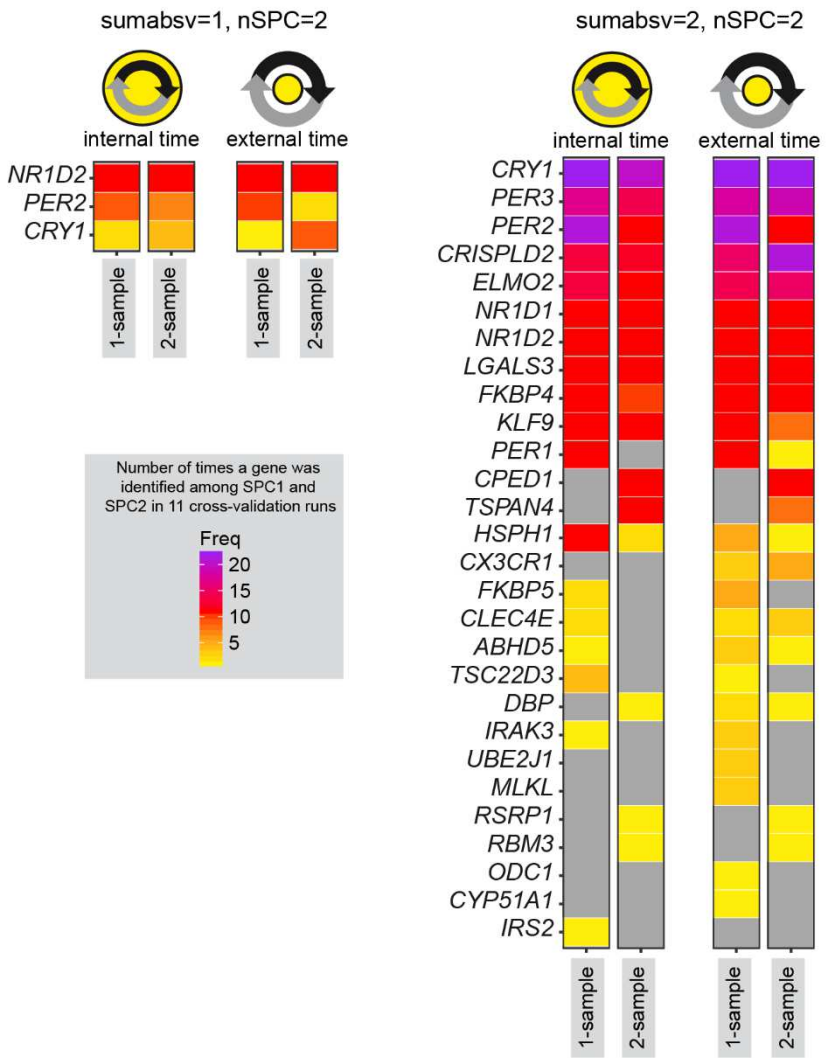
Supplemental Figure 1. Rhythmicity analysis of monocyte transcripts during constant routine. (**A**) The number of circadian transcripts identified at different false discovery rate thresholds for each subject. The circle marks the number of circadian transcripts at the chosen threshold of 0.05. (**B**) A different representation of the circadian transcripts identified at the 0.05 FDR threshold showing the ones rhythmic across different subjects. (**C**) The list of circadian transcripts shared between 6, 7, 8, 9 and 10 subjects (at the 0.05 FDR threshold). (**D, E**) Distributions of relative amplitudes and phases of the circadian transcripts. The boxplots (D) show the median, lower quartile and upper quartile of relative amplitude distribution of circadian genes for each subject. The phases of the circadian transcripts (E) in each subject are counted in 1 h bins over the time of day (external time).



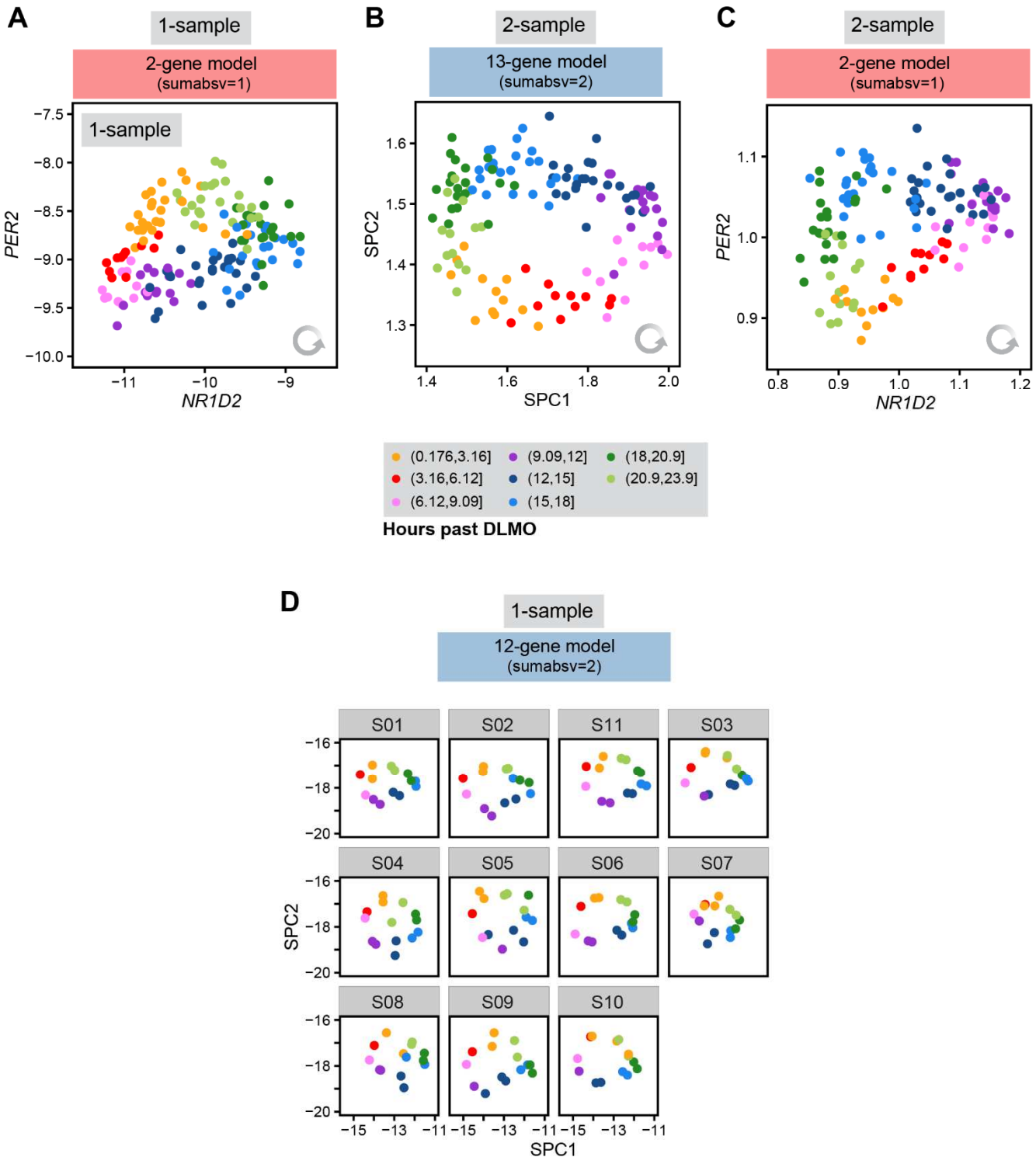
Supplemental Figure 2. Platform comparison correlation analysis. The correlations of the prediction errors estimated based on NanoString and RNA-Seq internal cross-validation models are shown. Spearman ρ correlation coefficients are indicated.



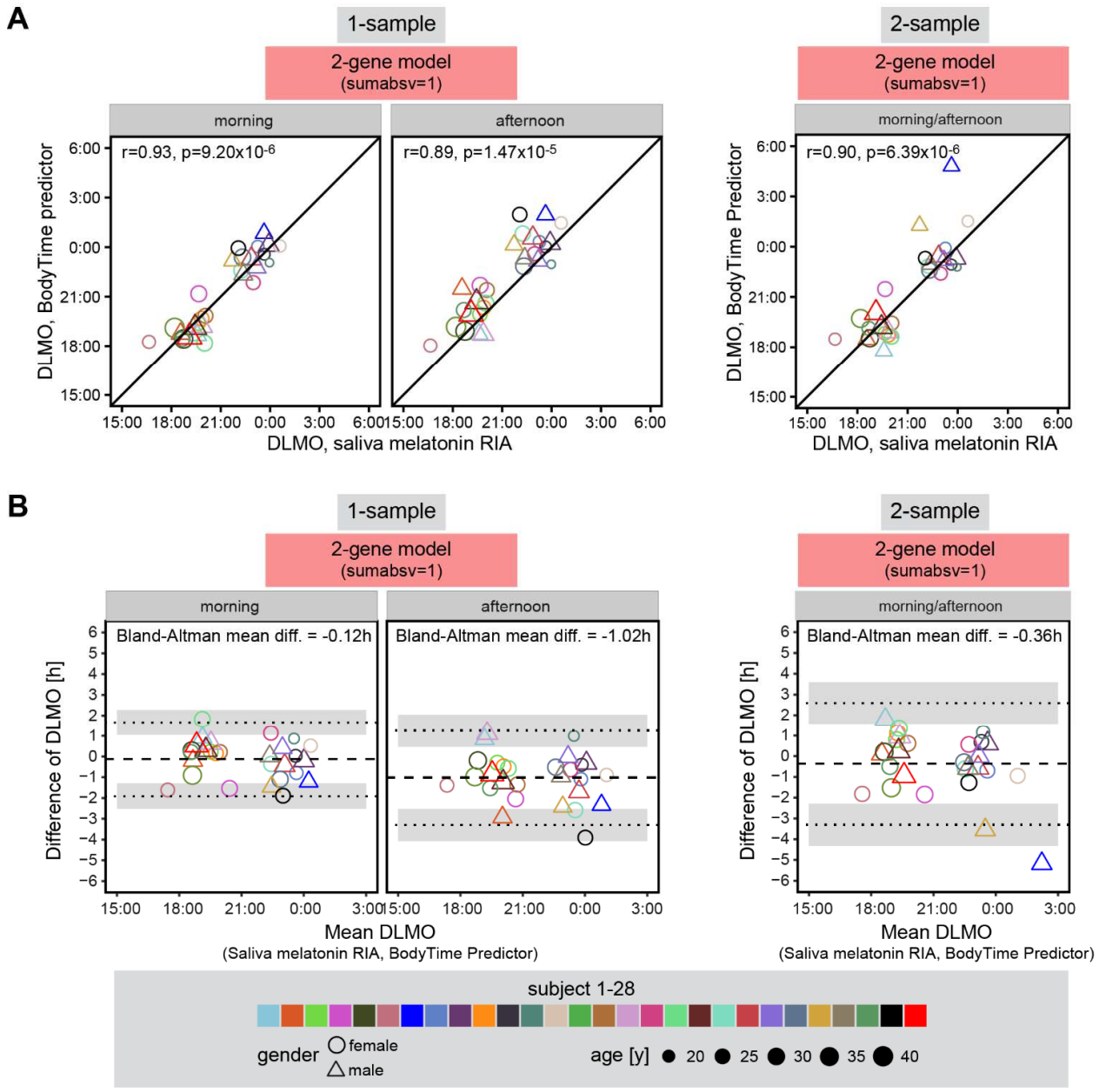
Supplemental Figure 3. Platform comparison Bland-Altman analysis. Bland-Altman analysis of the bias between the prediction errors estimated based on NanoString and RNA-Seq internal cross-validation models. The dashed horizontal line indicates the mean of the differences (bias), dotted lines represent the upper and lower limits (mean of the differences ± 2 standard deviations) with their 95% confidence intervals being shaded light gray.



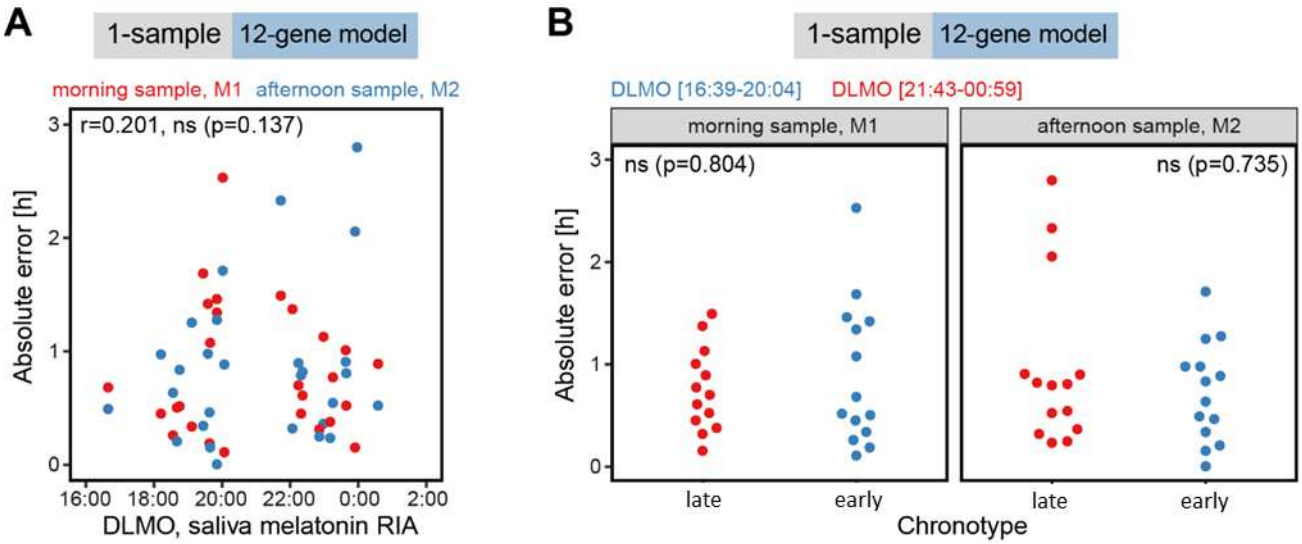
Supplemental Figure 4. Global gene sets of best-performing internal cross-validation predictors trained on the NanoString data of the BOTI study. Each column depicts a type of predictor defined by the predicted variable (internal or external time), the format of the data input (1-sample or 2-sample) and the optimal ZeitZeiger parameters (sumabsv, nSPC). Each predictor includes eleven leave-one-subject-out cross-validation runs, i.e. eleven gene sets. The order (from top to bottom) and the colors indicate how often a gene was extracted by ZeitZeiger and assigned to SPC1, SPC2 or both among those eleven gene sets.



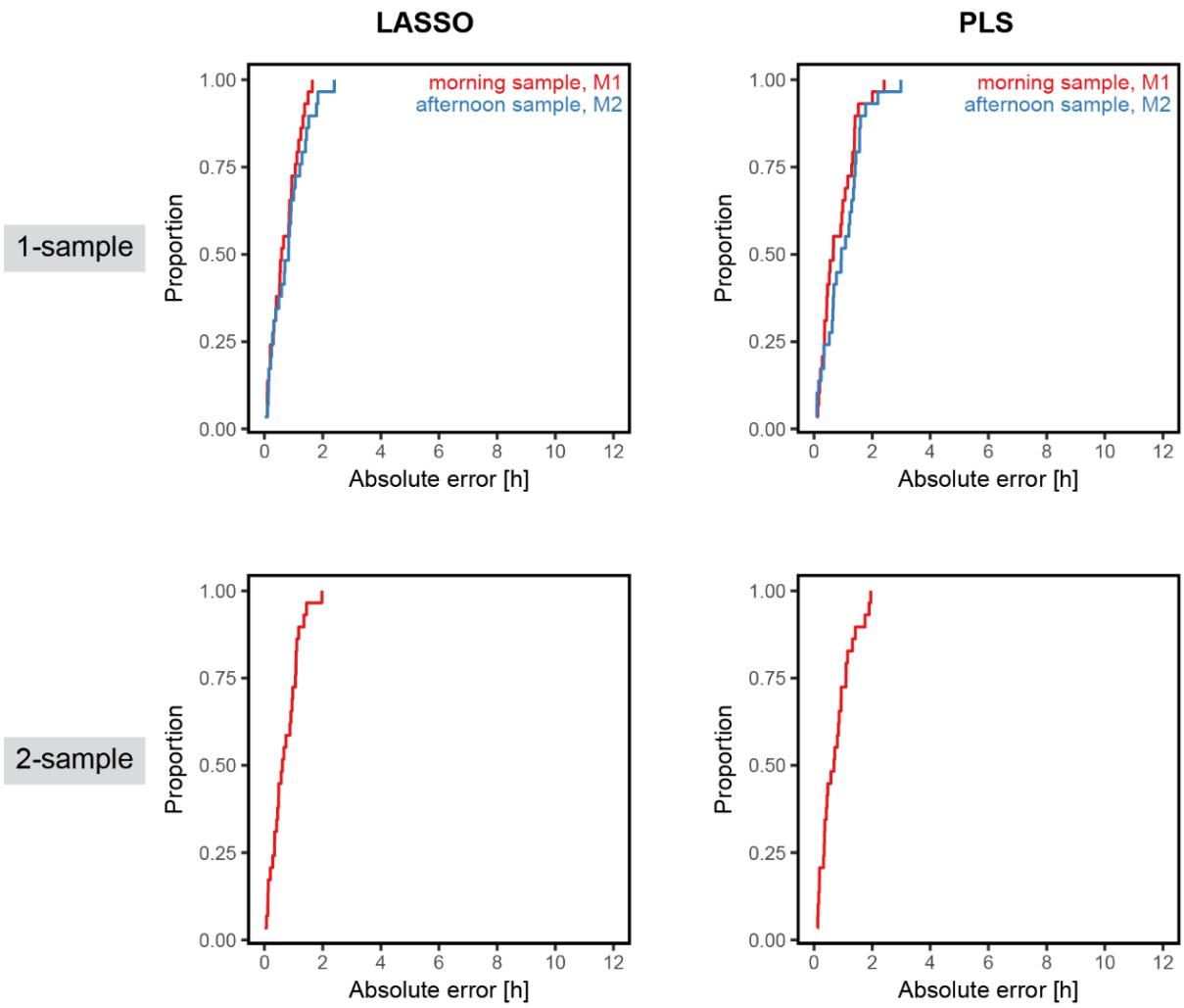
Supplemental Figure 5. Properties of the final BodyTime predictors. (**A-C**). NanoString expression profiles of the BOTI study's samples (n=154) in the SPC space of the 1-sample 2-gene, 2-sample 13-gene or 2-sample 2-gene predictor. (**D**) NanoString expression profiles of the BOTI study's samples (n=154) in the SPC space of the 1-sample 12-gene predictor faceted by subject. Colors indicate bins of the internal time.



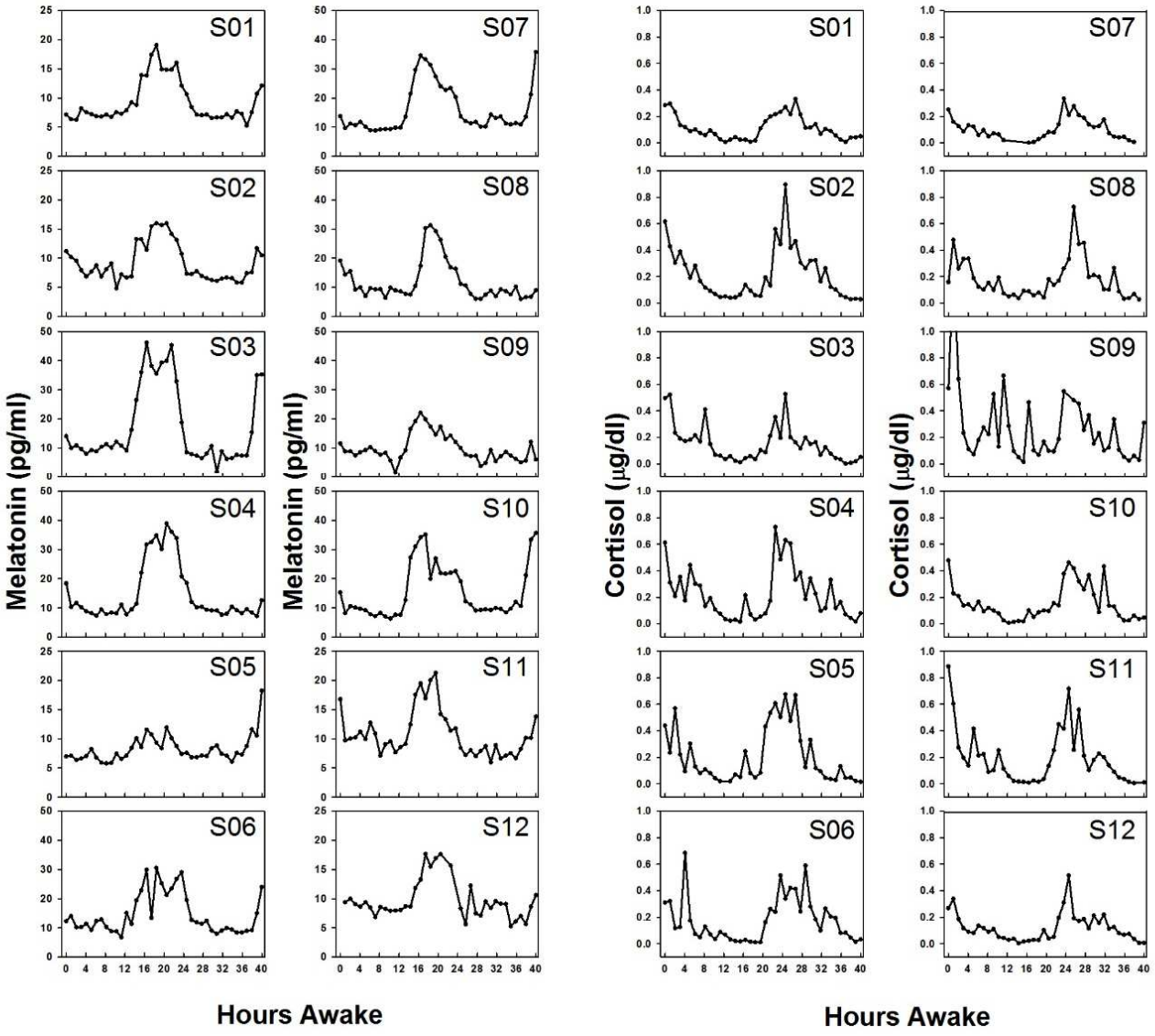
Supplemental Figure 6. Comparison of DLMO estimated by the 2-gene BodyTime predictors to DLMO derived from saliva melatonin concentrations (gold standard) assayed by RIA. (**A**) Circular correlation analysis. Circular Pearson correlation coefficients (r) and p-values are indicated. (**B**) Bland-Altman analysis. The dashed horizontal line indicates the mean of the differences (bias), dotted lines represent the upper and lower limits (mean of the differences ± 2 standard deviations) with their 95% confidence intervals being shaded light gray.



Supplemental Figure 7. The prediction error of the BodyTime predictor is independent of the chronotype of the subject. (**A**) Correlation plot of the absolute prediction error and DLMO derived from saliva melatonin concentrations (gold standard) assayed by RIA. The Pearson correlation coefficient and its significance are indicated in the top-left corner. (**B**) Boxplot of the absolute prediction error for early and late chronotypes. The p-values of Mann-Whitney U-tests are indicated.



Supplemental Figure 8. External validation of predictors in the independent VALI study using LASSO or partial least squares (PLS). Cumulative frequency distributions of the absolute prediction errors of the 1-sample and 2-sample NanoString predictors when they were applied to the VALI study data set using LASSO or PLS. In case of the 1-sample assay, the internal time stamps of all morning (M1) or afternoon (M2) samples were predicted; in case of the 2-sample assay the time stamp of the sample difference was predicted (M1 - M2). Proportion refers to the number of predictions with an absolute error that is less or equal to the specified value divided by the total number of predictions (1-sample, M1: n=28, 1-sample, M2: n=28, 2-sample, M1-M2: n=28).



Supplemental Figure 9. Melatonin and cortisol profiles during constant routine. Individual secretion profiles for melatonin (left two columns) and cortisol (right two columns) during the CR (study 1; n=12). Melatonin concentrations are indicated in pg/ml and cortisol concentrations in µg/dl. Please note: y-scales for melatonin are either between 0 and 25 pg/ml or between 0 - 50 pg/ml due to inter-individual secretion variability. For cortisol, y-scales are between 0 - 1 µg/dl. For subject 9, one value (1.5 µg/dl) is not depicted. Data were aligned relative to hours awake. There were a total of three missing melatonin and ten missing cortisol samples due to technical problems or insufficient material.

Supplemental Table 1: BOTI study – Participants' Body Mass Index (BMI), scores of PSQI, HO, MCTQ (MSF-sc)[a], SPAQ as well as habitual bedtimes (BT) and habitual wake times (WT) and DLMO times based on the 2 SD method.[b]

BMI	PSQI	HO	MCTQ (MSF-sc)	SPAQ	BT	WT	DLMO (2-SD)
23.6 (1.6)	3.4 (0.9)	50.3 (7.4)	4.5 (0.6)	7.1 (2.6)	23:50 (0:43)	7:51 (0:43)	21:17 (1:09)
20.5-25.9	2-5	34-60	3.5-5.3	4-13	22:49-1:26	6:49-9:26	19:13-22:36

[a] Local time of mid-sleep on free days corrected for sleep debt accumulated over the workweek.
[b] Given are mean scores, SD (in brackets) and ranges. BT, WT and DLMO are given in clock time (hh:mm).

Supplemental Table 3: Gene ontology (GO) functional enrichment analysis.[a]

GO.ID	Term	Annotated	Significant	Expected	p-value
GO:0002376	immune system process	1508	54	21.85	6.00E-12
GO:0009605	response to external stimulus	923	38	13.37	7.80E-10
GO:0042221	response to chemical	1836	55	26.6	4.30E-09
GO:0043207	response to external biotic stimulus	433	23	6.27	4.20E-08
GO:0051707	response to other organism	433	23	6.27	4.20E-08
GO:0050896	response to stimulus	3805	83	55.13	8.00E-08
GO:0006955	immune response	1092	38	15.82	8.60E-08
GO:0009607	response to biotic stimulus	454	23	6.58	1.00E-07
GO:0051704	multi-organism process	1209	40	17.52	1.40E-07
GO:0006952	defense response	750	30	10.87	1.60E-07
GO:0051239	regulation multicell. organism. process	1135	38	16.45	2.40E-07
GO:0048518	positive regulation of biological process	2685	65	38.9	3.90E-07
GO:0032501	multicellular organismal process	2622	64	37.99	3.90E-07
GO:0048856	anatomical structure development	2200	57	31.88	4.40E-07
GO:0010033	response to organic substance	1480	44	21.44	5.50E-07
GO:0044707	single-multicellular organism process	2460	61	35.64	5.80E-07
GO:0001775	cell activation	807	30	11.69	8.00E-07
GO:0070887	cellular response to chemical stimulus	1396	42	20.23	8.40E-07
GO:0032502	developmental process	2429	60	35.19	9.30E-07
GO:0080144	regulation of response to stress	773	29	11.2	1.10E-06

[a] GO analysis performed on the 119 genes extracted by ZeitZeiger at least once by RNA-Seq data based internal cross-validation (see Figure 2B in the main manuscript). As background set, 9115 genes identified as expressed in monocytes across all subjects were used.

Supplemental Table 5: Performance of internal cross-validation predictors built for platform comparison.[a]

Platform	Model	Model parameters (sumabsv, nSPC)	Absolute prediction error median [IQR]	Absolute prediction error ≤ 1 h [% of samples]	Absolute prediction error ≤ 2 h [% of samples]
NanoString	Internal, 1-sample	1, 2	0.89 [1.29]	56.6	82.4
	Internal, 1-sample	2, 2	0.77 [1.06]	61.0	88.2
	Internal, 1-sample	3, 2	0.83 [1.03]	57.4	83.8
RNA-Seq	Internal, 1-sample	1, 2	2.05 [2.75]*	24.3	50.0
	Internal, 1-sample	2, 2	1.61 [2.87]	39.0	58.8
	Internal, 1-sample	3, 2	1.53 [2.24]	38.0	61.0
NanoString	Internal, 2-sample	1, 2	0.83 [1.20]	47.8	71.3
	Internal, 2-sample	2, 2	0.81 [0.90]	51.5	80.1
	Internal, 2-sample	3, 2	0.84 [0.93]	57.4	80.9
RNA-Seq	Internal, 2-sample	1, 2	2.57 [3.00]	16.2	33.1
	Internal, 2-sample	2, 2	1.42 [2.36]	32.4	51.5
	Internal, 2-sample	3, 2	1.50 [2.11]	32.4	53.7
NanoString	External, 1-sample	1, 2	0.84 [0.96]	53.7	83.8
	External, 1-sample	2, 2	0.82 [0.98]	60.3	83.8
	External, 1-sample	3, 2	0.90 [1.29]	55.9	80.1
RNA-Seq	External, 1-sample	1, 2	1.96 [3.19]*	32.4	51.5
	External, 1-sample	2, 2	1.55 [1.90]*	33.1	59.6
	External, 1-sample	3, 2	1.40 [1.71]*	37.5	66.2
NanoString	External, 2-sample	1, 2	0.81 [1.12]*	50.7	72.8
	External, 2-sample	2, 2	0.76 [0.81]	55.9	78.7
	External, 2-sample	3, 2	0.59 [0.97]	56.6	77.2
RNA-Seq	External, 2-sample	1, 2	2.22 [3.53]	19.1	39.7
	External, 2-sample	2, 2	1.26 [2.80]*	36.0	50.7
	External, 2-sample	3, 2	1.13 [1.77]*	36.8	58.8

[a] Asterisks indicate cases where the absolute prediction error showed significant variation across 3 h time bins (Kruskal-Wallis test p-value < 0.05). IQR=interquartile range.

Supplemental Table 6: VALI study – demographics, screening questionnaires and chronotype related parameters.[a]

	ET (n=14)	Range	MT (n=14)	Range	T-Test
Sex	8F/6M	-	9F/5M	-	-
Age	24.2 (3.6)	18-31	29.6 (6.2)	22-41	*
BMI	23.0 (3.1)	18.4-29.1	22.5 (3.8)	18.4-30.4	ns
PSQI	4.2 (1.8)	2-8	2.1 (1.5)	0-5	*
MCTQ (MSF-sc)	6.66 (0.70)	5.75-7.98	2.50 (0.73)	1.26-3.53	**
HO	33.3 (6.0)	23-44	67.0 (5.6)	60-77	**
SPAQ	7.9 (5.2)	1-16	4.8 (3.4)	0-11	ns
BT	1:47 (0:42)	0:07-2:49	21:36 (0:35)	20:39-22:34	**
WT	9:53 (0:34)	9:01-11:24	5:54 (0:50)	4:27-7:22	**
Sleep Duration	8:05 (0:45)	6:36-9:27	8:18 (0:27)	7:40-9:08	ns
Midpoint Sleep	5:50 (0:31)	4:50-6:52	1:45 (0:41)	0:30-2:58	**
DLMO 3 pg/ml	23:03 (0:50)	21:43-0:35	19:08 (0:55)	16:39-20:04	**
DLMO 2-SD	23:02 (0:42)	21:40-0:28	19:07 (0:56)	16:42-20:04	**

[a] Given are mean, SD (in brackets) and ranges for both chronotypes (MT=morning types; n=14; ET=evening types; n=14). BMI= body mass index; PSQI, MCTQ (MSF-sc, i.e. local time of mid-sleep on free days corrected for sleep debt accumulated over the workweek), HO and SPAQ are derived from screening questionnaires; BT= bedtimes; WT=wake times; sleep duration and mid-sleep: all derived from averaged BTs and WTs, sleep duration (hh:mm) and mid-sleep times based on rest-activity recordings and sleep logs during 7 days before the study session. BT, WT, mid-sleep, DLMO 3 pg/ml method and DLMO 2-SD method are all indicated in clock times (hh:mm). The last column depicts significant differences between both chronotypes; *$p < 0.05$; ** $p < 0.0001$; ns=not significant; two-tailed t-test.

Supplemental Table 8: External validation of the BodyTime predictors in the independent VALI study using LASSO or PLS.

Predictor	Type of validation sample	Absolute prediction error median [IQR]	Absolute prediction error ≤ 1 h [% of samples]	Absolute prediction error ≤ 2 h [% of samples]
1-sample, LASSO	morning	0.59 [0.76]	72.4	100.0
	afternoon	0.82 [0.94]	69.0	96.6
1-sample, PLS	morning	0.66 [0.93]	65.5	93.1
	afternoon	0.94 [0.89]	51.7	93.1
2-sample, LASSO	morning/afternoon	0.62 [0.72]	72.4	100.0
2-sample, PLS	morning/afternoon	0.68 [0.75]	72.4	100.0

Supplemental Table 9: Quality controls for melatonin and cortisol assays.[a]

	Melatonin (Study 1)	Cortisol (Study 1)	Melatonin (Study 2)
Intra-Assay CV% (low control)	10.3	6.7	9.9
Intra-Assay CV% (high control)	6.4	3.8	10.1
Inter-Assay CV% (low control)	12.3	6.5	13.5
Inter-Assay CV% (high control)	12.2	5.6	9.7

[a]Intra- and inter-assay correlation coefficients of variability (CV) are shown for low and high dose controls in percentage (%).

Supplemental Table 12: External validation of the BodyTime predictors in the independent VALI study using ZeitZeiger and sample-by-sample normalization.

Predictor	Type of validation sample	Absolute prediction error median [IQR]	Absolute prediction error ≤ 1 h [% of samples]	Absolute prediction error ≤ 2 h [% of samples]
1-sample, 12-gene	morning	0.61 [0.85]	66.7	92.6
	afternoon	0.84 [0.81]	64.3	89.3
1-sample, 2-gene	morning	0.53 [0.86]	71.4	100.0
	afternoon	1.02 [0.98]	50.0	82.1

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	3
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	5-7
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	7
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	23-25; 25-27
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	23; 25
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	23-25; 25-27
	5b	D;V	Describe eligibility criteria for participants.	24-25; 26
	5c	D;V	Give details of treatments received, if relevant.	NA
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed. DLMO	27-28
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	28-31
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	
Sample size	8	D;V	Explain how the study size was arrived at.	26-27
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	31-32
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	31-32
	10c	V	For validation, describe how the predictions were calculated.	31-32
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	31-32
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	NA
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	NA
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	25-27
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	9; 16 Figure 1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	9; 16 Figure 1, SupTable S1 and S6
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	Figure 1
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	9, 13-14
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	NA
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	Figure 3
	15b	D	Explain how to the use the prediction model.	18; 32-33
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	Table 1 + 2, Figure 4
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	NA
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	20-21
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	(18-19)?
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	18-21
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	21-22
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	2

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.