



# Joint and Individual Representation of Domains of Physical Activity, Sleep, and Circadian Rhythmicity

Junrui Di, et al. [full author details at the end of the article]

Received: 18 December 2017 / Revised: 7 March 2019 / Accepted: 2 April 2019 / Published online: 15 April 2019  
© International Chinese Statistical Association 2019

## Abstract

Developments in wearable technology have enabled researchers to continuously and objectively monitor various aspects and physiological domains of real life including levels of physical activity, quality of sleep, and strength of circadian rhythm in many epidemiological and clinical studies. Current analytical practice is to summarize each of these three domains individually via a standard inventory of interpretable features, and explore individual associations between the features and clinical variables. However, the features often exhibit significant interaction and correlation both within and between domains. Integration of features across multiple domains remains methodologically challenging. To address this problem, we propose to use joint and individual variation explained, a dimension reduction technique that efficiently deals with multivariate data representing multiple domains. In this paper, we review the most frequently used features to characterize the domains of physical activity, sleep, and circadian rhythmicity and illustrate the approach using wrist-worn actigraphy data from 198 participants of the Baltimore Longitudinal Study of Aging.

**Keywords** Multi-domain · Physical activity · Sleep · Circadian rhythmicity · JIVE · Dimension reduction

## 1 Introduction

Traditional methods of collecting information about human behavior in the free-living environment have relied heavily on self-reported questionnaires, sleep logs, and daily diaries [65]. However, these methods have numerous limitations including recall biases, disease association confounding, and lack of detailed time-of-day information [4,56]. In recent years, there has been an exponential growth of both research- and consumer-grade wearable devices that employ actigraphy to obtain high-resolution

---

✉ Junrui Di  
jdi2@jhu.edu

Extended author information available on the last page of the article

objective assessment of multiple domains including physical activity, sleep, and circadian rest/activity rhythms. This multi-system or multi-domain objective tracking provides tremendous opportunities to identify key influences on the core components of human health and behavior [21]. However, this dramatic growth in the volume and complexity of collected behavioral data also presents difficult analytic challenges. To capture the complexity of information presented in wearable data, multiple features are typically generated to represent each of the domains of physical activity, sleep, and circadian rhythmicity.

Physical activity (PA) refers to motor activity taking place during waking time. In general, daily activity can be categorized into sedentary behavior, light PA (LiPA), and moderate-to-vigorous PA (MVPA) [11,71]. As recently shown, not only total sedentary and active times, but also the patterns of their accumulation demonstrate strong independent associations with adverse health outcomes and mortality [10,14,15]. Finally, the total volume of physical activity has been quantified through alternative aggregate measures of activity counts without applying any cut-points [9,71,72,77].

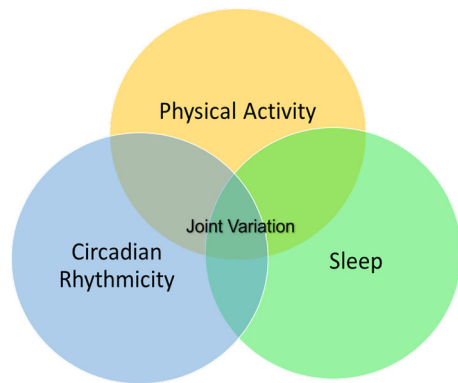
The amount, timing, and quality of sleep (SL) can be accurately estimated with wrist actigraphy. Estimating sleep over multiple days and weeks in the natural environment is a major advantage of using actigraphy over polysomnography, a golden standard for measuring the quality of sleep [3]. Typical actigraphy-derived summaries characterizing SL domain include measures of total sleep time, sleep efficiency, midpoint of time in bed, and sleep fragmentation [58,67].

Circadian rhythms are rhythms that oscillate about every 24 h. Circadian rhythmicity (CR) has been observed in multiple physiological processes including core body temperature, hormone secretion, heart rate, blood pressure, and many others [16,18,44]. Measuring circadian rhythm with wearables is based on a principle that there is increased movement during wake periods and reduced movement during sleep periods, and has been shown to be reliable and valid [26,28,30]. The strength of CR can be assessed using both parametric and non-parametric approaches. The cosinor, the extended cosinor, and the multi-period cosinor are the most popular parametric models [13,41] for CR. Intra-daily variability (IV) and relative amplitude (RA) are the most popular non-parametric summaries that quantify fragmentation and synchronization, and have been widely used in various applications [24,69,70,75]. Functional data analysis provides non-parametric quantification of diurnal patterns via functional principal components [23,57,78].

Guided by specific questions, several lines of research typically focus on one of the three domains without considering the joint dependence of features within and between the domains. There is now a growing interest in understanding both joint and individual effects of all three domains and their relationships with different health outcomes.

When faced with multivariate correlated data, the common practice is to seek a low-dimensional representation of the data using appropriate dimension reduction techniques. Principal component analysis (PCA) is a popular method that transforms the originally correlated variables to orthogonal principal components (PCs). However, PCA fails to properly account for clustering of features within domains. Given the physiological relation between PA, SL, and CR, it is reasonable to expect some shared patterns across the domains. At the same time, since the domains represent

**Fig. 1** Venn diagram shows two parts of variation, domain-specific (individual part of each circle) and joint (the shaded intersection regions)



different physiological systems, a substantial domain-specific variation is expected. Figure 1 schematically illustrates a hypothetical variation allocation across the three domains. Each of the three domains contains domain-specific or individual variation (represented by the specific part of the circle) and the variation that is shared by two of three or all three domains (represented by the shaded intersection).

To deal with this type of data, several integrative dimension reduction techniques have been developed. Canonical Correlation Analysis (CCA) [29] is a popular method to globally examine the relation between two sets of variables. Partial Least Squares (PLS) [76] directions are defined similarly to CCA, but maximize covariance rather than correlation. However, the restriction of PLS to pairwise comparisons limits their utility in finding common structure among more than two data types. Recently Lock et al. [40] introduced joint and individual variation explained (JIVE), a method which decomposes original data matrix into a low-rank approximation capturing joint variation across data types, low-rank approximations for structured variation individual to each data type, and residual noise. It is an efficient extension of PCA that quantifies the amount of joint variation between data types, reduces the dimensionality of the data and reveals principal components that can be used for the visual exploration of joint and individual structures. Several recent generalizations have focused on extending JIVE to (1) incorporate heterogeneous data types (continuous/binary/count) using exponential family distributions Li and Gaynanova [39], (2) separate not only joint and individual, but also joint and partially shared variation [22], and (3) formalize the problem and increase computation feasibility with non-iterative procedure using fast linear algebra [19]. In this study, we propose to use original JIVE as a flexible approach to quantify and separate between- and within-domain variation for objective tracking of human behavior.

In the rest of the paper, we first provide a brief review of typically used actigraphy-derived features for each of the three domains. Then, we describe JIVE method and demonstrate it in a case study that analyses actigraphy data of 198 participants from the Baltimore Longitudinal Study of Aging (BLSA).

## 2 The Domains of PA, SL, and CR

Accelerometer, a key component of most wearable devices and smartphones, is often used to objectively measure 24-h motor activity. It collects high-resolution acceleration signal (e.g., 30–100 HZ). However, to study questions regarding PA, SL, and CR domains, researchers often rely on summary measures of the raw acceleration signal aggregated over certain epochs (e.g., 1 min) [7]. One common example is activity count (AC) which is a unitless proxy measure of activity level within an epoch. AC is usually generated by software using proprietary algorithms (different across platforms) or using open-source software that filters the acceleration signal and aggregates it at epoch level, typically, minute level [1,7,32]. Regardless of the above-mentioned properties, AC is almost always a reflection of intensity and volume of activity [8]. Therefore, in this study, we focus on minute-level AC data, so that subject-specific activity daily profile is represented via a 1440-dimensional vector of ACs.

A large number of features/summaries have been proposed in literature to describe different aspects of PA, SL, and CR. In this section, we review commonly used features from each domain and briefly describe their statistical interpretations.

### 2.1 Notation

Let  $y(t)$  denote the activity count (AC) at epoch  $t$  and minute-level epochs,  $t \in T = \{1, 2, 3, \dots, 1440\}$ . We will represent  $T$  as a union of the two parts, wake time (WT) period and sleep (time in bed) (SL(TiB)) period that will be used to derive features representing PA and SL domains, respectively.

### 2.2 The Domain of Physical Activity

#### 2.2.1 Total Volume of Physical Activity

Total volume of physical activity [72] serves as a proxy for the total amount of accumulated physical activity over a day across all levels of intensity, from sedentary to vigorous. The most commonly used measure is total activity counts (TAC) which is defined as

$$\text{TAC} = \sum_{t \in \text{WT}} y(t). \quad (1)$$

Because, TAC exhibits high levels of skewness [53,72], a Box–Cox transformation is often applied to obtain a more symmetric distribution. The most straightforward approach is to take the log-transformation of TAC (LTAC). Another measure of the total volume is total log-transformed activity counts (TLAC), defined as

$$\text{TLAC} = \sum_{t \in \text{WT}} \log(y(t) + 1). \quad (2)$$

After the log-transformation, zero activity count is transformed to zero and non-zero activity counts are re-scaled to a much smaller range with a more symmetric distribution. TAC and TLAC have been systematically compared by Varma et al. [72] in National Health and Nutrition Examination and Survey (NHANES) 2003–2006. Varma et al. [72] concluded that TAC primarily represents moderate-to-vigorous intensity physical activity (MVPA), while TLAC primarily represents light-intensity physical activity (LiPA).

### 2.2.2 Total Volume of Sedentary Behavior

Sedentary behavior, as a significant risk factor for a wide range of chronic diseases and mortality, had gained a lot of attention in various research fields recently [11,12,17,25,27,35,42,48,73]. Conceptually, it is defined as any waking low-energy behavior while in a sitting, reclining, or lying posture and includes activities such as sitting, lying down, and watching television, and other forms of screen-based entertainment [63].

To identify sedentary behavior from epoch level AC data, one needs to apply pre-defined thresholds. For example, in NHANES 2003–2006, where ActiGraph AM-7164 accelerometer (ActiGraph, LLC, Fort Walton Beach, Florida) was used, sedentary activity is typically defined for minutes with less than 100 AC [35,64]. In studies that use devices with no validated thresholds, a common practice is to explore a grid of thresholds and report the results and their interpretation and sensitivity for the entire grid.

The most common summary of sedentary behavior is total sedentary time (TST) spent during wake period. It can be represented as

$$TST = \sum_{t \in WT} I(y(t) < h), \tag{3}$$

where  $I()$  is the indicator function.

To account for subject-specific differences in total WT (TWT) and try to separate effects of the longer total sedentary time and shorter sleep, the percent of sedentary time (pST) is considered as a normalized measure of the total sedentary time

$$pST = \frac{TST}{TWT}. \tag{4}$$

Similar features can be calculated to represent other time-related volumes, such as the total active time (TAT) and the percent of total active time (pAT).

### 2.2.3 Patterns of Accumulation of Sedentary and Active Time

Recently, there has been a growing interest in exploring and quantifying the patterns of accumulation of the total sedentary and total active times and associating these patterns with various health outcomes [14,15,17,38]. In a systematic review, Di et

al. [14] compared previously proposed fragmentation metrics. Conceptually, fragmentation approaches represent actigraphy-estimated wake period of the daily profile as a sequence of alternating bouts of sedentary and active time and quantify this sequence via summaries of the duration of and frequency of switching between sedentary and active bouts. The main fragmentation features include i) average bout duration (denoted by  $\mu$ ), (ii) Gini index (denoted by  $g$ ), defined as an absolute variability normalized to the average bout duration, (iii) active-to-sedentary and sedentary-to-active transition probabilities (denoted by  $\lambda_a$  and  $\lambda_s$ , respectively), which can be re-expressed as the reciprocal of  $\mu$  (see proof in Di et al. [14]); and (iv)  $\alpha$ , the parameter of the power-law distribution. For a detailed discussion of these measures, their statistical properties, and the distribution in the elderly US population, we refer to Di et al. [14].

### 2.3 The Domain of Sleep

Actigraphy devices used in sleep research, such as Actiwatch-2 (Philips Respironics, Bend, OR), often have built-in event buttons or come along with sleep logs that allow to estimate WT and SL (TiB) using participant reports. For devices with no event buttons and studies not using sleep logs, automatic sleep detection algorithms [2,31,46,61,68] can be applied to estimate WT and SL (TiB). Per current practice [3], standard actigraphy-derived sleep parameters, which are derived using algorithms that have been validated against overnight polysomnography, include (i) total sleep time (TSLT; total time slept while in bed), (ii) percentage of total sleep time (pSLT; percentage of total time slept while in bed), (iii) the number of sleep bouts (NSB), (iv) wake after sleep onset (WASO, total time awake after initial sleep onset), (v) percentage of total wake time (pWT), (vi) the number of wake bouts (NWB), (vii) average wake bout duration (AWB; WASO divided by NWB), (viii) sleep efficiency (SEFF; percentage of time in bed spent asleep), (ix) sleep onset latency (SOL; interval from time into bed to sleep onset), and (x) midpoint of the time in bed [67].

Similarly, to the fragmentation metrics for PA domain, we define the wake-to-sleep transition probability (WSTP) as the reciprocal of AWB.

### 2.4 The Domain of Circadian Rhythmicity

#### 2.4.1 Extended Cosinor Model

Marler et al. [41] introduced a family of non-linear parametric transformations of the traditional cosine curve used in the modeling of biological rhythms and often referred as the extended cosinor model (extCosinor). The non-linear transformation is the sigmoidal family, represented by three family members: (i) the Hill function, (ii) the anti-logistic function, and (iii) the arctangent function. These transformations add two additional parameters that must be estimated, in addition to the acrophase ( $\phi$ ), MESOR (*mes*), and amplitude (*amp*). The main advantage of extCosinor is that the estimated curves have shapes that would require more than two additional harmonics to achieve the same (non-linear) fit when modeled with harmonic analysis [41].

The classic cosinor model is defined as

$$y(t) = mes + amp \cdot c(t), \tag{5}$$

where  $c(t) = \cos([t - \phi]2\pi/24)$ .

The Hill-transformed cosine curve is defined as  $h(c(t)) = [c(t) + 1]^\gamma / (m^\gamma + [c(t) + 1]^\gamma)$ ; the anti-logistic-transformed cosine curve as  $l(c(t)) = \exp(\beta[c(t) - \alpha]) / \{1 + \exp(\beta[c(t) - \alpha])\}$ ; and the arctangent-transformed cosine curve as  $\psi(c(t)) = \tan^{-1}[\beta c(t) - \alpha] / \pi + 0.5$ . Thus, the sigmoidally transformed cosine models is defined as  $y(t) = min + amp \cdot F(c(t))$  with  $F()$  being either  $h()$ ,  $l()$ , or  $\psi()$ .

Interpretation wise, *mes* is a half of the deflection; *min* is the minimum value of the function; *amp* is the difference between the minimum and maximum of the function;  $\phi$  is the time at which  $y(t)$  has its mathematically well-defined “peak”;  $\alpha$  and  $m$  controls the “width” of the function; and  $\beta$  and  $\gamma$  controls the “steepness” of the function.

Although cosinor and extCosinor models are popular in the analysis of traditional circadian markers such as cortisol, melatonin, and core body temperature, its use for actigraphy data has a few important limitations. The major limitation is the parametric form of the model that it is very restrictive and may not be very appropriate for real diurnal patterns of motor activity. In addition, assuming the 24-h period, the extCosinor model estimates only five parameters and as can be seen from profiles shown in Fig. 2, actigraphy profiles often have more than five landmarks, thus the profiles often cannot be adequately modeled via extCosinor framework. Non-parametric approaches are often more flexible and more sensitive.

### 2.4.2 Functional Principal Components Scores

As opposed to extCosinor model, functional principal component analysis (FPCA) is a data-driven technique that makes no parametric assumptions about the functional form of diurnal patterns. FPCA represents any sample of diurnal patterns via  $L_2$ -orthogonal functional “principal components” and subject-specific functional PC scores that can be used as scalar covariates. For subject  $i$ , FPCA is typically formulated as

$$y_i(t) = \mu(t) + z_i(t) = \mu(t) + \sum_{k=1}^K \phi_k(t) \xi_{ik}, \tag{6}$$

where  $\mu(t)$  is the overall mean function, and  $z_i(t)$  is the subject-specific deviation from the overall mean. The deviation  $z_i(t)$  can be further decomposed as  $\sum_{k=1}^K \phi_k(t) \xi_{ik}$ , where  $\phi_k(t)$  is the  $k$ -th functional principal component and  $\xi_{ik}$  is the  $k$ -th principal component score. More details can be found in Ramsay [51].

### 2.4.3 Stability of Rest-Activity Rhythms

Van Someren et al had proposed non-parametric methods to summarize rest-activity rhythms and study them in Alzheimer patients [69,70,75].

Intra-daily variability (IV) and relative amplitude (RA) have been used to estimate circadian rhythmicity in various clinical populations. IV measures fragmentation in the rest/activity rhythms and is capable of detecting periods of daytime sleep and nocturnal arousal, and is calculated as the ratio of the mean squares of the difference between successive hours and the mean squares around the grand mean, i.e.,

$$IV = \frac{\sum_{t=2}^T (y(t) - y(t-1))^2}{T-1} \cdot \frac{T}{\sum_{t=1}^T (y(t) - \bar{y})^2}. \quad (7)$$

We assume that time epochs,  $t$ , can be minute, hour, etc and  $\bar{y}$  is the overall mean. RA is defined as  $(M10-L5)/(M10+L5)$ , where M10 is the most active 10-h period and L5 is the least active 5-h period.

### 3 Integrative Analysis of Features from Multiple Domains via JIVE

The list of actigraphy-derived features representing each of the three domains can be easily expanded with more summaries quantifying uncovered aspects of activity profiles, by applying different thresholds, and considering different time (epoch) resolutions [24]. Thus, eliminating possible redundancy while accounting for multi-feature multi-domain representation of wearable data becomes a critical component of efficient data analytical pipelines. Low-dimensional representations of the original high-dimensional summaries extracted from wearable data can often capture most of the relevant information. These representations can provide informative visual insight into the data and reveal hidden clusters of features and subpopulations of subjects.

Intuitively, a significant amount of interdependence is expected across domains of the PA, SL, and CR due to the fact that jointly modeled domains represent the same 24-h day/night cycles from the same group of subjects. At the same time, since the domains represent different physiological systems, a substantial individual domain variation is expected as well. Traditional dimension reduction techniques, such as PCA, can be applied to each domain separately. However, analysis of individual domains will not capture potential dependencies between domains. This motivated recent research on developing methods that explicitly take into account joint and individual domain variation [40].

Joint and individual variation explained (JIVE) has been proposed to deal with scenarios where different sources or views of the data are simultaneously available for the same set of samples [40]. In the original application, JIVE jointly models Mi-RNA and gene expression data collected on the same set of subjects. JIVE decomposes the original multi-block data into a sum of three components: a low-rank approximation capturing joint variation of the domains, low-rank approximations capturing variation individual to each domain, and residual noise. The imposed rank and orthogonality constraints can be considered as extensions of PCA. Let  $\mathbf{Y}^d \in \mathbb{R}^{p_d \times n}$  denote the data structure from domain  $d$ ,  $p_d$  denote the number of features. If  $\mathbf{Y}$  has been row centered and scaled, JIVE can be formulated as follows:



$$\begin{aligned}
 \mathbf{Y}^1 &= J^1 + A^1 + \epsilon^1 = \Phi_J^1 \Gamma_J + \Phi_A^1 \Gamma_A^1 + \epsilon^1 \\
 &\vdots \\
 \mathbf{Y}^D &= J^D + A^D + \epsilon^D = \Phi_J^D \Gamma_J + \Phi_A^D \Gamma_A^D + \epsilon^D,
 \end{aligned}
 \tag{8}$$

or in a general form as

$$\mathbf{Y} = \mathbf{J} + \mathbf{A} + \boldsymbol{\epsilon},
 \tag{9}$$

where  $\mathbf{J} = \begin{bmatrix} J^1 \\ \vdots \\ J^D \end{bmatrix}$  denotes the joint structure matrix with  $rank(\mathbf{J}) = r$ , and  $A^d$  denotes the individual structure with respect to  $\mathbf{Y}^d$  with  $rank(A^d) = r_d$ . Ranks  $r$  and  $r_d$ 's can be determined by either BIC or a permutation test [40]. Matrices  $\Gamma_J$  and  $\Phi_J$  are score and loading matrices to the joint structure, while  $\Gamma_A^d$  and  $\Phi_A^d$  are score and loading matrices for the individual structures. JIVE is fitted by iteratively estimating joint and individual structures. The full details on JIVE and its estimation can be found in Lock et al. [40].

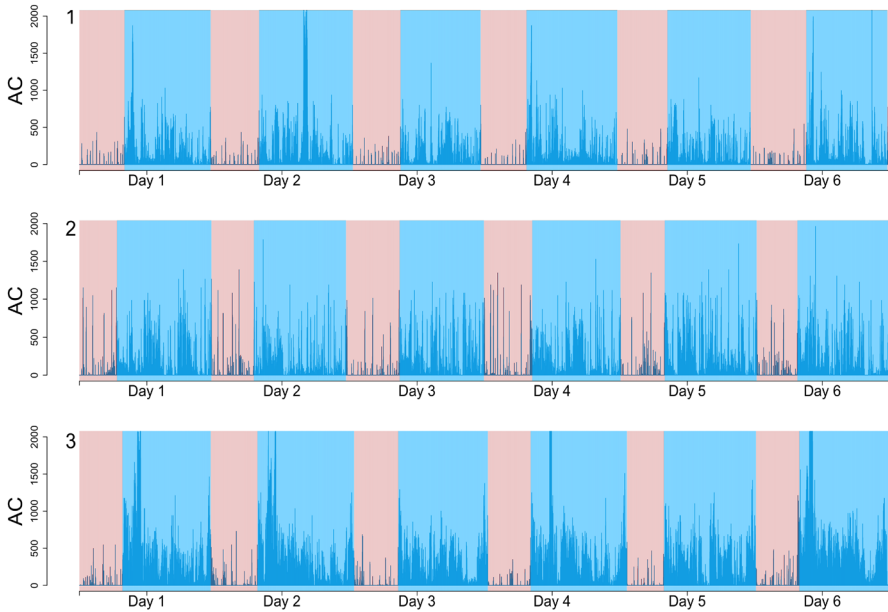
It is important to note that JIVE provides a non-parametric low-rank- based procedure to estimate and exclude joint across-domain variation from original features by creating adjusted domain-specific features such as  $\mathbf{Y}_{adj} = \mathbf{Y} - \mathbf{J}$ . For example, it is possible to filter out the effect of physical activity from the quality of sleep, which regular PCA is not capable of.

## 4 Case Study: BLSA Study

### 4.1 Data Description

The Baltimore Longitudinal Study of Aging (BLSA) is a study of normative human aging, established in 1958 and conducted by the National Institute on Aging Intramural Research Program. Detailed descriptions of the sample and enrollment procedures/criteria have been previously reported [20,59].

In a sub-study, BLSA participants were asked to do wrist actigraphy with Actiwatch-2 (Philips Respironics, Bend, OR), an actigraph worn on the non-dominant wrist for seven consecutive 24-h periods. To assist with sleep scoring of actigraphy data, participants were asked to press the “event” marker button on the device to indicate “lights off” time in the evening and “wake-up” time in the morning when no longer intending to sleep, and to complete a sleep diary confirming these times. Following the convention for processing BLSA actigraphy data [47,53,55,74], days with more than 5% of data missing (more than 72 min per day) were treated as invalid and excluded from the analysis. For the days with less than 5% of missing data, missing values were imputed as the average activity counts per minute over all valid days at the same period for each participant. Participants with at least 3 valid (not necessarily consecutive) days of actigraphy data were included in the analysis. Thus, the analysis



**Fig. 2** Activity profiles of 6 days/nights for three randomly chosen BLSA participants. Red and blue backgrounds indicate sleep and wake periods, respectively (Color figure online)

included 198 BLSA participants (with an average of 5.72 valid days) with age ranging from 31 to 96 years (with median age of 74 years) and 58% of females.

Figure 2 shows 6 days from three randomly selected BLSA participants. Red background shows sleep periods (defined by the event marker/diary), the blue background shows wake periods. As can be easily observed from the example, there is a significant difference across the three subjects both in physical activity and sleep patterns. For example, participants at the top and bottom panels demonstrate more consolidated sleep periods, characterized by lower activity and lower fragmentation, compared with the middle profile. This could potentially be the reason for the subjects to have distinctively different activity patterns during wake time.

## 4.2 Features

The features described in Sect. 2 have been calculated for each subject. Table 1 summarizes all domain-specific features. In the domain of PA, TAC, TLAC, LTAC, TST, and pST were calculated for each day and averaged across valid days. Fragmentation metrics were calculated by aggregating bouts across all valid days in one summary. We considered the thresholds of 50 AC and 100 AC to define sedentary periods. Sleep domain features were calculated using Actiware 6.0 software (Philips Respironics, Bend, OR), which uses a validated algorithm to derive sleep parameters [37]. SL features were calculated based on two thresholds: 20 AC and 40 AC. All sleep features were derived at nightly level and then averaged across all valid days/nights. Features

**Table 1** Domain-specific Features

Domains	Thresholds	Features
PA	50, 100	TAC, TLAC, LTAC, TST, pST, $\mu$ , $\lambda$ , $g$ , $\alpha$
SL	20, 40	SOL, SEFF, WASO, pWT, NWB, AWB, TSLT, pSLT, NSB, WSTP
CR	–	Min, mesor, amplitude, $\alpha$ , $\beta$ , $\phi$ , RA, IV (at both minute and hourly level), $fPC_1$ , $fPC_2, \dots, fPC_{10}$

of CR included five parameters of the extended cosinor models with anti-logistic-transformation fitted using all valid days. RA and IV were calculated at daily level, and then averaged across valid days. Subject-specific 24 h activity profiles were averaged across valid days and functional PCA was applied to them to obtain first ten functional PCs which explained more than 90% of total variation.

As a result, 23 features represented PA domain, 20 features represented SL domain, and 19 features represented CR domain. Please see Table 1 for the list of all features grouped within each domain.

All features have been pre-normalized. Specifically, Lock et al. [40] suggested to pre-normalize each data domain (i.e., center each individual feature and scale by Frobenius norm of the block) to circumvent cases where “the largest domain wins.” However, in our application, even within each domain, different features can have highly distinct scales. For example, consider TAC and  $\lambda_a$  in the PA domain. TAC is usually in the magnitude of at least  $10^5$ , while  $\lambda_a$  is a probability that takes values between 0 and 1. To address this, we also centered and scaled each feature. As a result, we ended up using the correlation matrices of features from multiple domains. This is an important step to perform in situations when features are measured on very heterogeneous scales.

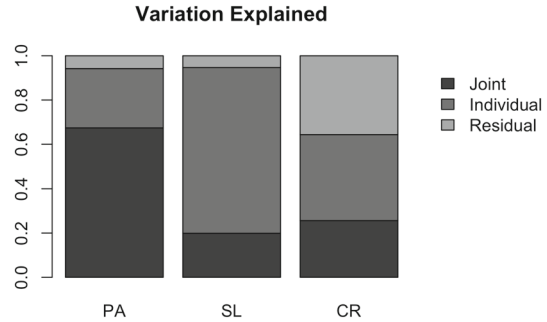
### 4.3 Results

Both standard PCA and JIVE were applied to the domain-specific features. In JIVE, ranks were selected via permutation test proposed in Lock et al. [40]. Row-orthogonality was enforced both between joint and individual components. The optimal choices of the lower rank representation were estimated to be:

- rank 4 joint structure,
- rank 2 structure individual to PA domain,
- rank 3 structure individual to SL domain,
- rank 4 structure individual to CR domain.

Figure 3 displays a bar-chart of the amount of variation explained by joint and individual components in each of the three domains. Interestingly, almost 70% of variation in PA and only around 20% of variation in SL and CR can be explained by the joint components. Partially, this could be explained by PA having the largest number of included features. Individual components explain around 30%, 75%, and 40% of total variation in PA, SL, and CR domains, respectively. Thus, there is almost 40% variation that remain unexplained in CR domain. It is important to put these

**Fig. 3** Joint and individual variations explained by each domain from JIVE. Joint structure explains 67.4%, 19.9%, and 25.6% of total variation in PA, SL, and CR domains, respectively. Individual structures explain 26.8%, 74.7%, and 38.7% of total variation in PA, SL, and CR domains, respectively. And there are still 5.8%, 5.3%, and 35.6% of variation that cannot be explained in PA, SL, and CR, respectively



**Table 2** Joint variation explained by each domain, JIVE

	Joint components				Individual components
	Joint PC1	Joint PC2	Joint PC3	Joint PC4	
PA %	89.4	5.4	25.8	16.4	9.9
SL %	0.3	57.2	55.4	5.9	24.1
CR %	10.3	37.4	18.8	77.7	11.9
Total variation %	39.3				45.9

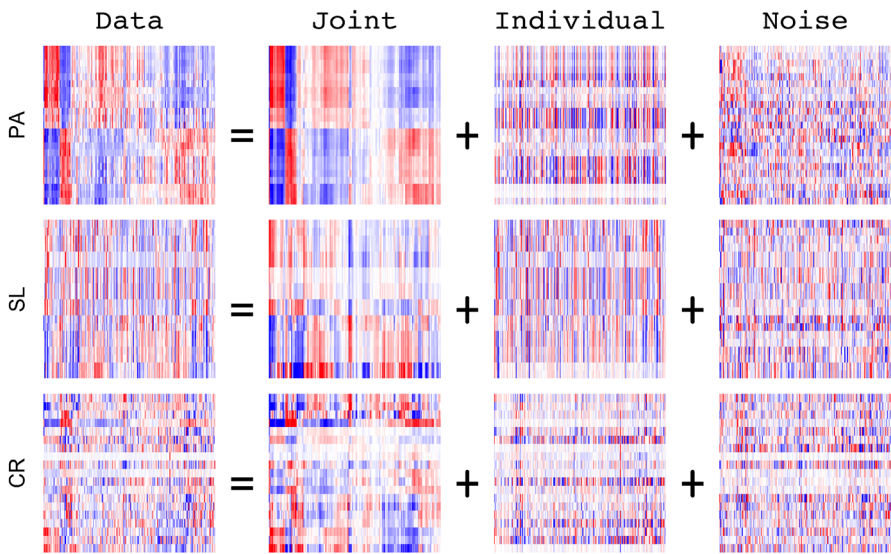
**Table 3** Variation explained by each domain, PCA

	PC1	PC2	PC3	PC4
PA %	90.9	1.7	15.4	64.8
SL %	0.5	93.8	72.7	14.5
CR %	8.6	4.5	11.9	20.7
Cumulative variation %	29.1	48.2	55.9	62.2

results into a perspective by comparing the number of components and the number of features representing each domain.

Table 2 shows the contribution of each domain to the variation explained by joint and individual components. Overall, the joint components explained 39.3% of total variation and the individual components explained 45.9% of total variation. It is interesting to note that the first joint PC is almost exclusively loaded on PA (89.4%) and CR (10.3%) domains. A similar loading allocation is observed for the fourth joint PC which is loaded on PA (16.4%), SL (5.9%), and CR (77.7%) domains. The second joint PC is primarily loaded on SL (57.2%) and CR (37.4%) domains.

To compare and demonstrate the difference between JIVE and PCA, we applied PCA and retained first 4 principal components, which explained 62.2% of total variation. Variation explained by the PCs and their loading on the three domains are displayed in Table 3. The first PC (PC1) is highly loaded on PA (90.9%) and CR (8.6%), but not SL, which is similar to Joint PC1.



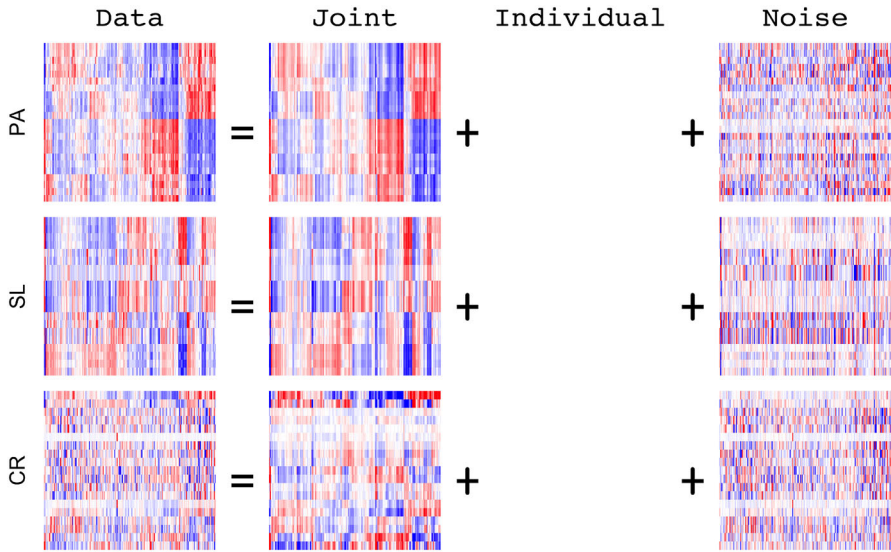
**Fig. 4** JIVE decomposition. Columns represent subjects and rows represent features. The columns are vertically aligned for all heatmaps, with red corresponding to higher values and blue lower values. Rows and columns are ordered by complete-linkage clustering of the joint structure

JIVE estimates in the form of low-rank matrix approximations are shown as heatmaps in Fig. 4 with columns representing subjects and rows representing features. Both columns and rows were clustered by complete-linkage hierarchical clustering of Euclidean distances based on the joint structure to better display underlying latent structures. The joint structure shows clear patterns that can also be seen in the three blocks of data representing the domains. Note that the joint structure appears less prominent in the CR domain, mostly because it explains a smaller amount of the circadian variation.

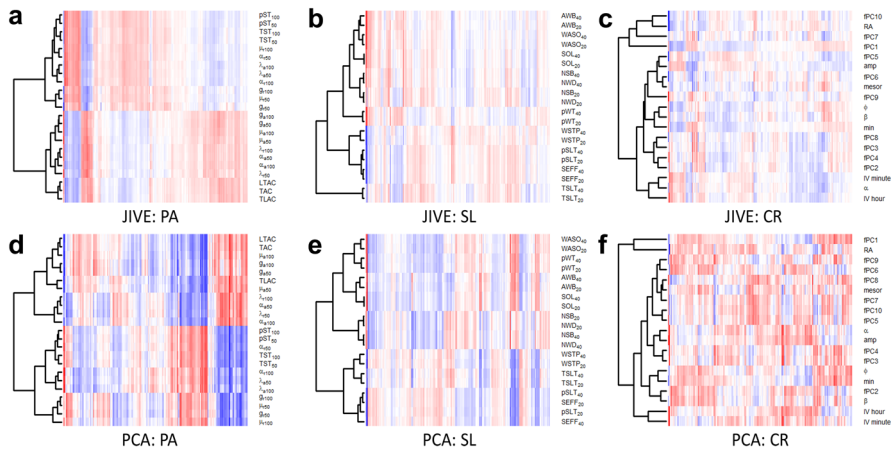
PCA rank-4 approximation was visualized in a similar fashion and is shown in Fig. 5. PCA obviously captured a low-rank structure that is different from the one defined by joint components of JIVE.

Figure 6 shows the hierarchical clustering of features corresponding to JIVE joint components (a–c in PA, SL, and CR domains, respectively) and PCA components (d–f in PA, SL, and CR domains, respectively). In PA domain, the clusters are similar both for JIVE and PCA: the features characterizing sedentary and active behavior compose different clusters. However, in SL domain, the clusters created by JIVE are different from those created by PCA. Interestingly, total sleep time (TSLT) forms a distinctive cluster by JIVE, but not by PCA. On the other hand, sleep efficiency (SEFF) and percentage of sleep time (pSLT) are placed in the same clusters both by JIVE and PCA. Finally, in CR domain, although both JIVE and PCA create a distinctive cluster including fPC1 and RA, however, JIVE provides a more nuanced clustering of the other non-parametric and parametric features of circadian rhythmicity

Figure 7 shows the cumulative variation explained by the first four PCA and the four joint JIVE PCs with the relative loading of each domain. Since PCA did not account

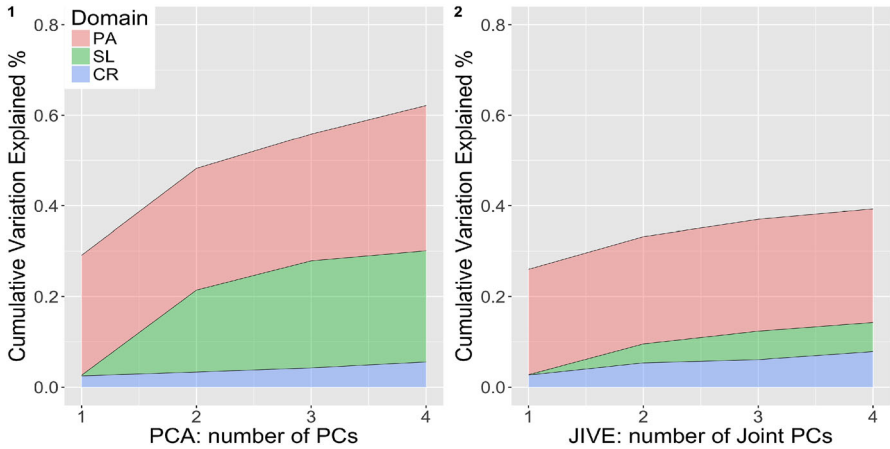


**Fig. 5** Rank 4 approximation based on PCA (no individual domain structure). Columns represent subjects and rows represent features. The columns are vertically aligned for all heatmaps, with red corresponding to higher values and blue lower values. Rows and columns are ordered by complete-linkage clustering of the joint structure (Color figure online)



**Fig. 6** Details of feature clustering for JIVE joint structure and PCA shown in Figs. 4 and 5. **a–c** show clusters for JIVE estimation of PA, SL, and CR, respectively. **d–f** show clusters for PCA estimation of PA, SL, and CR, respectively

for the individual structure, hence, it is less constrained, so it captured more variation than joint JIVE. First, four PCA and joint JIVE PCs explain roughly 60% and 40% of total variation, respectively. Surprisingly, almost 20% difference is primarily driven by the reduced contribution of the SL domain.



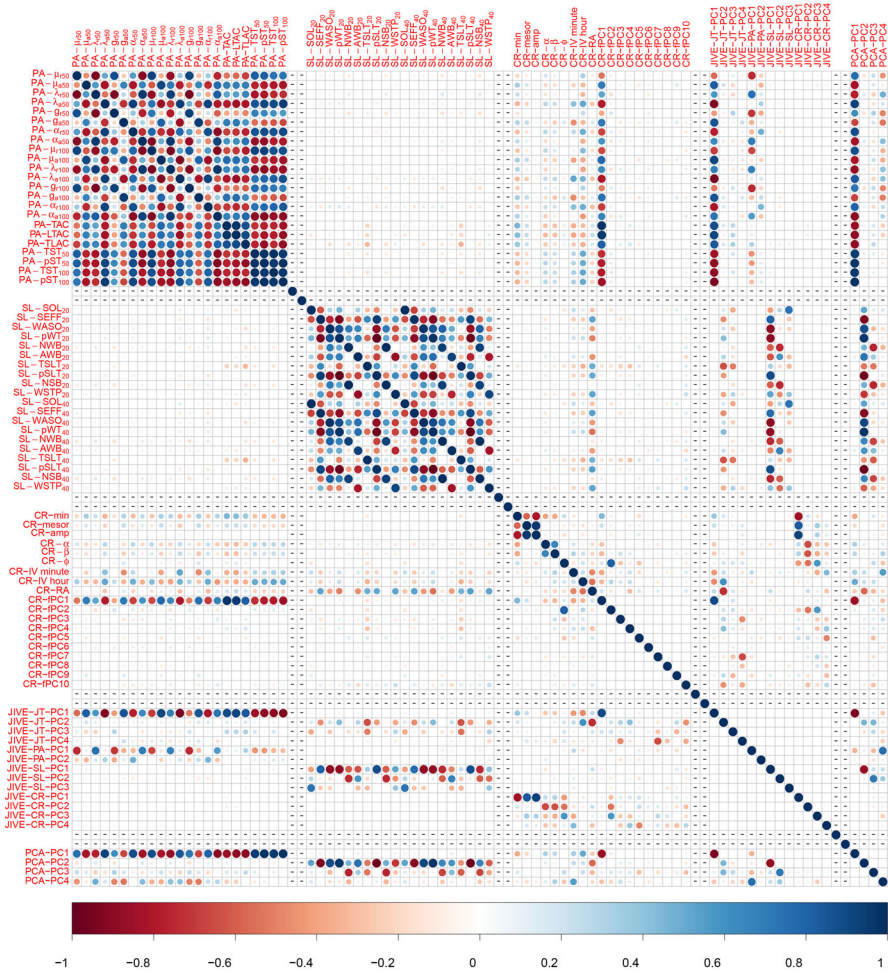
**Fig. 7** Cumulative variation explained by first four PCA-PCs and first four JIVE-JT-PCs

Both PCA and JIVE provide principal component scores, projections of the original data onto the corresponding principal components. Below, we denote a joint JIVE score by “JT-PC” and an individual score corresponding to one of the domains by “PA/SL/CR-PC.” Figure 8 shows the cross-correlation between all features, the 13 joint and individual JIVE scores, and top 4 PC scores from PCA. It is obvious that there is extremely high correlation between features within both the PA and the SL domains. However, the level of correlation in the CR domain is not as high. This is expected because the first 10 functional PC scores have been included to represent this domain. Surprisingly, there is a very low cross-correlation between features of PA and SL domains. Moreover, all features in the PA domains display relatively high correlation with majority of features in CR. This correlation is particularly strong between PA and the first functional PC scores (fPC1), which typically shows the average diurnal profile. As shown in 11 in Appendix A, fPC1 corresponds to overall activity level during waking hours (approximately 6AM to 12AM), the time period where the features of PA domain were derived for majority of participants. We observed that JT-PC1 is highly correlated with majority of features from PA, and some features from CR (especially fPC1), but not correlated with SL features. This is consistent with our finding such that the first Joint PC primarily represents PA and CR, but not SL. On the other hand, JT-PC2 and JT-PC3 represent significant variation of SL domain. Figure 9 shows the age-related change of joint and individual JIVE scores. Clear age trends can be seen for most of the derived JIVE scores with specific landmarks, such as lows and highs and changes in the observed trends.

Unlike the original features that demonstrate significant correlation, the derived JIVE scores are orthogonal by construction. Similar to how PCA scores are used in the principal component regression to substitute original covariates with PC scores, JIVE scores can be used in the same way.

To illustrate this, we studied the association between usual gait speed (a commonly used indicator of physical function/performance) and scores from PCA (PCA regression) and JIVE (JIVE-regression). Both PCA and JIVE scores are uncorrelated by

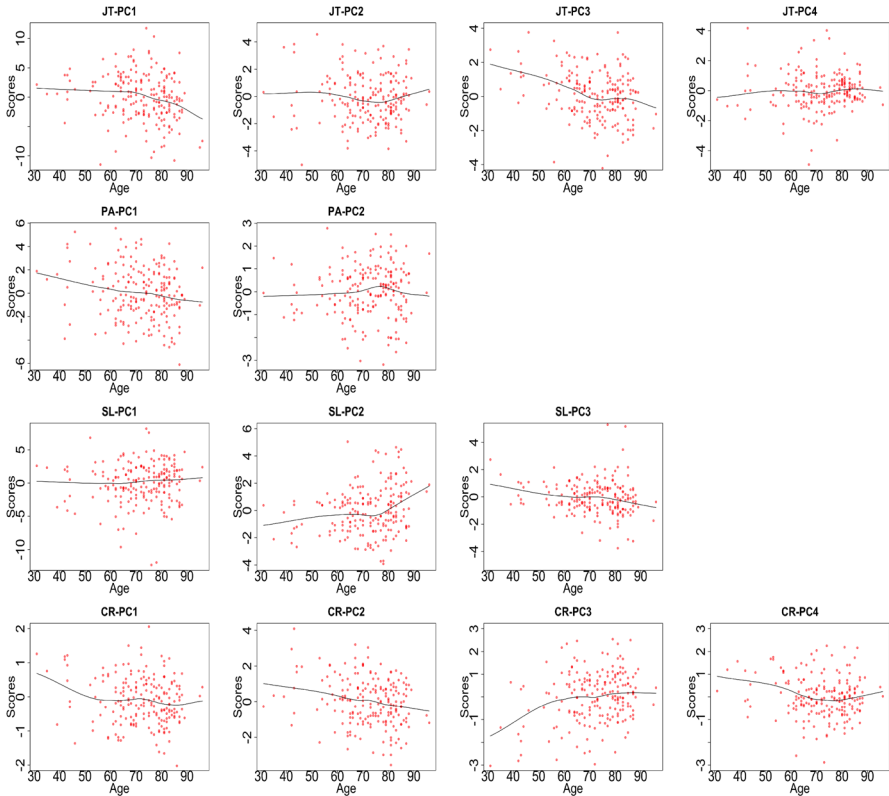




**Fig. 8** Cross-correlations between twenty three PA features, twenty SL features, nineteen CR features, four JIVE joint scores, two JIVE-PA scores, three JIVE-SL scores, four JIVE-CR scores, and four PC scores

construction, therefore, we can include all of them as covariates without worrying about collinearity. Baseline model (Model 0) contains age, gender, and body mass index (BMI) as covariates. Models 1 and 2 expand the baseline model with PCA and JIVE scores, respectively. The results for all three models are shown in Table 4. In general, the JIVE-regression resulted in a higher adjusted- $R^2$  of 0.33 compared to adjusted- $R^2$  of 0.29 for PCA regression. However, this could be partially due to the smaller number of PCA components. In the model with PCA scores, PCA-PC1 and PCA-PC4 are negatively associated with gait speed. PCA-PC1 is highly loaded on all features in PA domain, while PCA-PC4 is slightly loaded on almost all features in all three domains. Thus, the interpretation of PCA regression is quite complicated because it involves almost all features. JIVE-regression provides 0.09 increase in the





**Fig. 9** Relationship between age and joint/individual scores. Black solid lines are smoothing curve fitted using loess

adjusted- $R^2$  compared to the baseline model. Because all scores are uncorrelated and normalized, the magnitude of regression coefficients can be approximately interpreted as a contribution to the increase of the adjusted- $R^2$ . First, JIVE-JT-1 is highly negatively correlated with PCA-PC1, thus, this finding is similar to PCA regression. However, significant scores JIVE-PA-2 in PA domain and JIVE-SL-2 and JIVE-SL-3 in SL domain together contribute more than a half to the increase of the adjusted- $R^2$ . JIVE-PA-2 is highly loaded on all fragmentation metrics of physical activity and JIVE-SL-2 and JIVE-SL-3 are highly loaded on the fragmentation of sleep and do not include any metrics characterizing sleep efficiency. Thus, the results highlight the features quantifying patterns of fragmentation both during wake and during sleep as independently associated with gait speed. Interestingly, this is consistent with a recent finding of a negative association between the gait speed and the fragmentation of physical activity(measured by  $\lambda_a$ , i.e., activity-to-sedentary transition probability), such that increasing of  $\lambda_a$  is associated with a slower gait speed [54]. To the best of our knowledge, the association between gait speed and the sleep characteristics has not been studied. Thus, JIVE-regression allows to separate domain-specific sources

**Table 4** Regression modeling of gait speed and PCA and JIVE scores

Model	Covariates	$\beta$	$p$ value	Adjusted $R^2$
Model 0	BMI	-0.079	0.21	0.24
	Age	<b>-0.50</b>	<b>&lt; 0.0001</b>	
	Female	<b>-0.31</b>	<b>0.016</b>	
Model 1	<b>PC1</b>	<b>-0.17</b>	<b>0.0069</b>	0.29
	PC2	-0.031	0.61	
	PC3	-0.11	0.075	
	<b>PC4</b>	<b>-0.14</b>	<b>0.022</b>	
Model 2	<b>JIVE-JT-1</b>	<b>0.21</b>	<b>0.00096</b>	0.33
	JIVE-JT-2	-0.061	0.33	
	JIVE-JT3	0.036	0.56	
	JIVE-JT4	-0.0055	0.93	
	JIVE-PA-1	-0.016	0.79	
	<b>JIVE-PA-2</b>	<b>0.15</b>	<b>0.012</b>	
	JIVE-SL-1	-0.019	0.75	
	<b>JIVE-SL-2</b>	<b>-0.14</b>	<b>0.024</b>	
	<b>JIVE-SL-3</b>	<b>0.13</b>	<b>0.033</b>	
	JIVE-CR-1	-0.016	0.79	
	<b>JIVE-CR-2</b>	<b>0.14</b>	<b>0.023</b>	
	JIVE-CR3	0.049	0.34	
JIVE-CR4	-0.094	0.11		

Model 0 is the baseline model with age, gender, and BMI. Models 1 and 2 include PCA-PC and JIVE scores, respectively. All continuous variables (age, BMI, and all scores) have been centered and normalized. Statistically significant values are given in bold ( $p < 0.05$ )

of variability and model their independent association with the outcome of interest while addressing possible multicollinearity.

We also explored a simple alternative to JIVE by using a two-step procedure: (1) performing PCA to each individual domain and (2) applying JIVE to the joint truncated data representing all three domains. To keep high enough percent of variation and to be able to potentially recover structures similar to original JIVE, we considered two scenarios for step one (PCA for individual domains): (1) keeping 5 PCs in each domain and (2) keeping 8 PCs in each domain. The full results are provided in Appendix B.

We believe this approach may have two limitations. First, it requires rank selection both at step 1 (PCA for individual domains) and step 2 (JIVE to truncated data). This seems to introduce extra uncertainty. Second, if we do not keep a sufficiently large number of PCs for each domain, this can result in not capturing some parts of joint variation. One intuitive analogy is Principal Component Regression (PCR) that does PCA on covariates at step 1 and regress PC scores on an outcome at step 2. PCR has no guarantee that variation kept at Step 1 is relevant to variation of the outcome. In our settings, PCA done in each individual domain is not informed by variation from other domains and, thus, in the same way as PCR, may discard a part of joint variation.

#### 4.4 Sensitivity Analyses to Missing Data

In the main analysis, we follow the convention for BLSA actigraphy data processing pipeline to handle missingness in raw minute-level AC data. Specifically, days with more than 5% of missingness (more than 72 min per day) were treated as invalid and excluded from the analysis. For the days with less than 5% of missing data, missing values were imputed as the average activity counts per minute over all valid days at the same period for each participant.

In this section, we conduct sensitivity analyses using alternative ways of handling missing data. In sensitivity analysis 1 (denoted by S1), missing data were imputed using median of the specific time for each participant. We still have 198 subjects with the total of 1134 valid days (the mean is 5.72 days and standard deviation is 0.62). In sensitivity analysis 2 (denoted by S2), instead of using 5% of the threshold, we excluded days with more than 7% of missing data (101 min per day). We have 198 subjects with the total of 1163 days (the mean is 5.76 days and standard deviation is 0.59). In sensitivity analysis 3 (denoted by S3), we considered the most aggressive approach that removed all days with any missing data. This approach resulted in 189 subjects with the total 986 days (the mean is 5.22 days and standard deviation is 0.98).

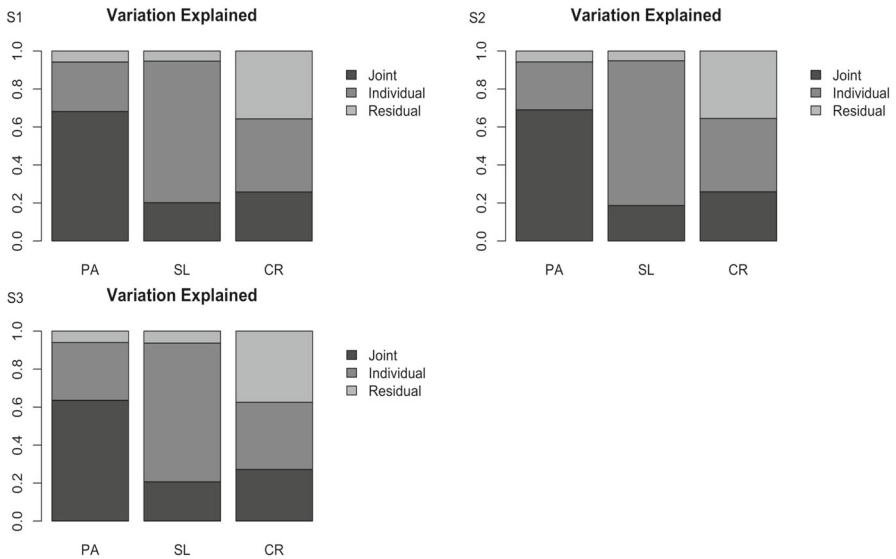
All three approaches provided the results very similar to the results in the main analysis. In all three sensitivity analyses, JIVE chose joint rank to be 4, and individual ranks for PA/SL/CR domains to be 2, 3, and 4, respectively. Fig 10 displays bar-charts of the amount of variation explained by joint and individual components in each of the three domains. PA is dominated by the joint variation, SL is dominated by individual variations, and a significant amount of variation remained unexplained in CR domain.

Figures 15, 16, and 17 in the appendix display the cross-correlation plots between features and JIVE/PCA scores. The patterns are quite similar to what we observed from main analysis as well.

## 5 Discussion

We proposed to use JIVE to model individual and joint variation of domains of physical activity, sleep, and circadian rhythmicity represented by multiple domain-specific actigraphy-derived features. Using the BLSA study, we estimated and separated joint and individual components and scores and examined the correlation between those scores and the original features. We also explored age-specific changes of the scores and their association with usual gait speed. To the best of our knowledge, this is the first attempt to simultaneously model the domains of PA, SL, and CR using JIVE.

Our results demonstrate that the first JIVE Joint PC (JT-PC1) primarily represents shared PA and CR variation. Recent studies have shown that fPC1, which is also highly correlated with JIVE-JT-PC1, can be used as a biomarker for “biological age” [50]. Even though JIVE is developed to identify joint variation across all three domains, this is an interesting finding. This can be further investigated by applying recently developed “structural learning and integrative decomposition” [22] that estimates partially shared variation by assuming certain block-sparsity across domains.



**Fig. 10** Joint and individual variations explained by each domain from JIVE for S1–S3. Results of all three sensitivity analyses are similar to the variation pattern in the main analysis

Visual exploration of age-related changes in JIVE scores revealed interesting trends across joint and individual JIVE scores. These trends may provide additional insights in studies of PA/SL/CR that look at age effects, and try to understand joint and individual effects to each domain. This deserves further analyses and proper regression modeling in the future to identify a more rigorous relationship between age and the joint and individual scores.

From regression modeling perspective, JIVE provides joint and individual scores that can be included simultaneously in regression models as covariates, which can be seen as an extension of the standard principal component regression [34]. A big advantage of JIVE is its ability to separate joint and individual variation and to explore associations between domain-specific variation and outcomes. As an example, we studied the association between usual gait speed with PCA and JIVE scores. The results showed that features quantifying patterns of fragmentation both during wake and sleep time were independently associated with gait speed, which supports results from a recent study [54]. Thus, our results demonstrate the potential of JIVE to reveal important biological mechanisms in the joint context of PA, SL, and CR.

It is important to note that, JIVE, similar to PCA, is an unsupervised, data-driven approach for dimension reduction, which depends on pre-selected features and, potentially, the study sample [52]. Therefore, when JIVE results are interpreted, it is crucial to understand that the inferences are based on pre-selected features. Even though we tried to incorporate a comprehensive list of commonly used features, the research fields of physical activity, sleep, and circadian rhythmicity are under a rapid development with a constant in-flow of novel summaries. Therefore, the number of features

within each domain is expected to increase drastically. A big advantage of JIVE is that it efficiently handles redundancy of highly correlated features. Thus, for those features that require thresholding of AC or depend on a parameter which needs to be tuned or there is an interest in considering multiple time scale, it is possible to include several features corresponding to a wide grid of thresholds/parameters/time scales and eliminate created redundancy through JIVE. This is important because (i) this eliminates the need for choosing thresholds/parameters and (ii) different thresholds/parameters may correspond to different aspects of the phenomena of interest. It is also important to note that BLSA is a study of healthy aging and on average tend to focus on healthier sub-population, so the results in another cohort may be different.

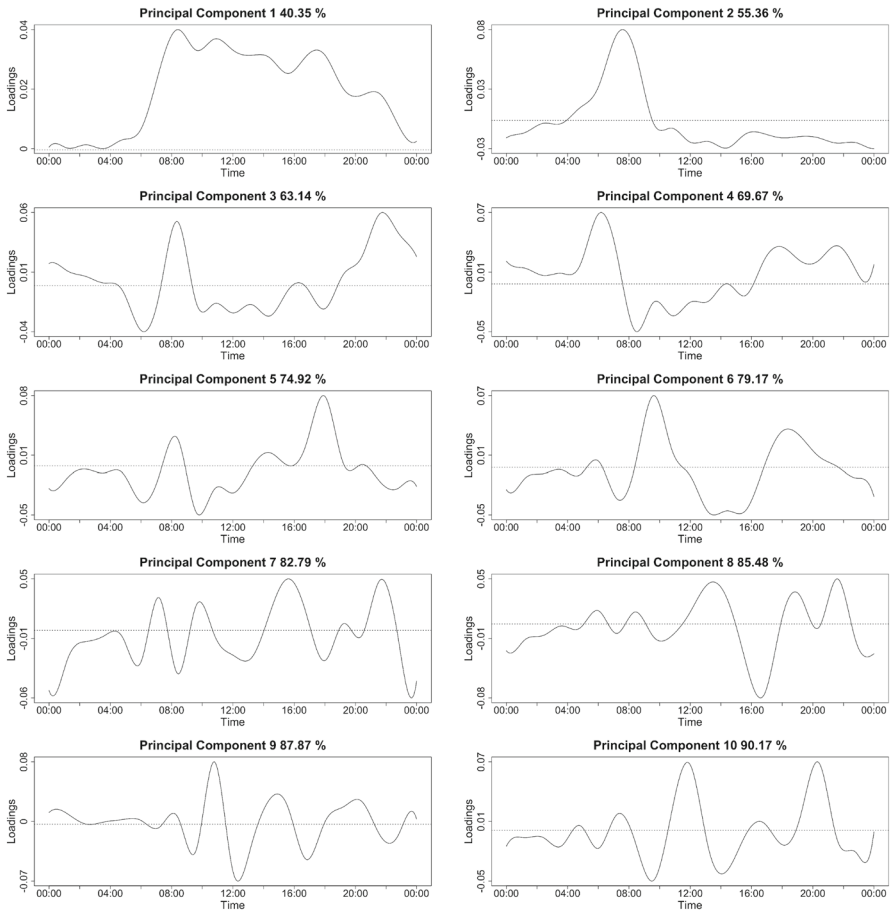
Surprisingly, in our study the cross-correlation between domains of PA and SL was small. There are a few possible explanations for this. First, it may be due to non-linear relationships between features of PA and SL. For example, it has been shown that the total sleep time usually exhibits a U-shaped association with various health outcomes such as obesity, diabetes, heart disease, and mortality [5,6,36,49,60]. Second, domain-specific features were generated by averaging across valid days, which likely resulted in damping dynamic associations between adjacent sleep–wake periods. Thus, potentially strong within-week temporal patterns clustered within subjects were averaged out. Estimating those weekly temporal patterns and incorporating them into JIVE will need to be done as a next step.

JIVE can be readily extended to include an arbitrary number of domains. This is especially important considering the active embedding of multiple sensors in wearables for multi-domain tracking of physical activity, ambient light, skin temperature, sweat, blood pressure, glucose, heart rate, and many others.

There are a few limitations. First, JIVE only accounts for second-order correlation structure and ignores higher-order mutual dependencies that may be quite informative when dealing with continuous multivariate non-Gaussian data. Second, JIVE is not capable of working with features that follow discrete distributions such as binary, Poisson, ordinal, or categorical distributions. Third, JIVE only models joint and individual patterns without accounting for partially shared structure, when, for example, two domains can share a joint component that is not present in the third domain. Thus, future work will need to focus on developing and applying methods that decompose joint dependence structure using measures of higher-order dependency, such as multivariate skewness and kurtosis [33,43,45], and are capable of handling mixed data types [39,62,66] and accounting for partially shared structure [22].

In summary, our proposal of using JIVE to model the domains of PA, SL, and CR is an important step forward to provide analytical methods for multi-feature multi-domain wearable data that will facilitate our understanding of complex interactions of multiple physiological systems.

**Funding** This study was supported in part by the Intramural Research Program (IRP), National Institute on Aging (NIA), National Institutes of Health (NIH), and by Research and Development Contract HHSN-260-2004-00012C. Dr. Adam Spira was supported in part by R01AG050507 from the National Institute on Aging. Dr Adam Spira received an honorarium from Springer Nature Switzerland AG for Guest Editing



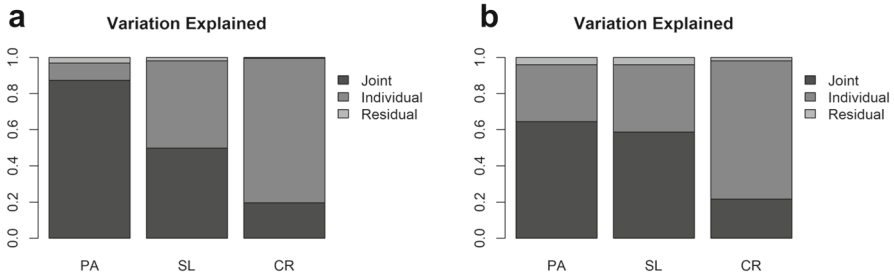
**Fig. 11** First 10 functional principal components reported with cumulative variance explained

a Special Issue of Current Sleep Medicine Reports. Dr. Mark Wu was supported by R01AG054771. Dr. Jennifer Schrack was supported by U01AG057545 and R21AG053198.

## Appendix

### A Functional PCA

Functional PCs for circadian rhythmicity domain, for each subject, 1440-min activity profile were averaged across valid days which resulted in an average daily profile that is representative. Then functional PCA using sandwich smoother for covariance matrix smoothing was applied [78]. The first 10 principal components explained more than 90% of total variation. Figure 11 shows the first 10 functional PCs.



**Fig. 12** Joint and individual variations explained by each domain from the alternative approach to conduct JIVE. **a** 5 PCs were kept from the first step. Joint structure explains 87.3%, 49.8%, and 19.6% of total variation in PA, SL, and CR domains, respectively. Individual structures explain 9.6%, 48.2%, and 79.8% of total variation in PA, SL, and CR domains, respectively. And there are only 3.2%, 2.0%, and 0.5% of variation that cannot be explained in PA, SL, and CR, respectively. **b** 8 PCs were kept from the first step. Joint structure explains 64.4%, 58.6%, and 21.6% of total variation in PA, SL, and CR domains, respectively. Individual structures explain 31.4%, 37.3%, and 76.4% of total variation in PA, SL, and CR domains, respectively. And there are only 4.2%, 4.1%, and 2.0% of variation that cannot be explained in PA, SL, and CR, respectively

## B Results from the Two-Step Alternative Approach

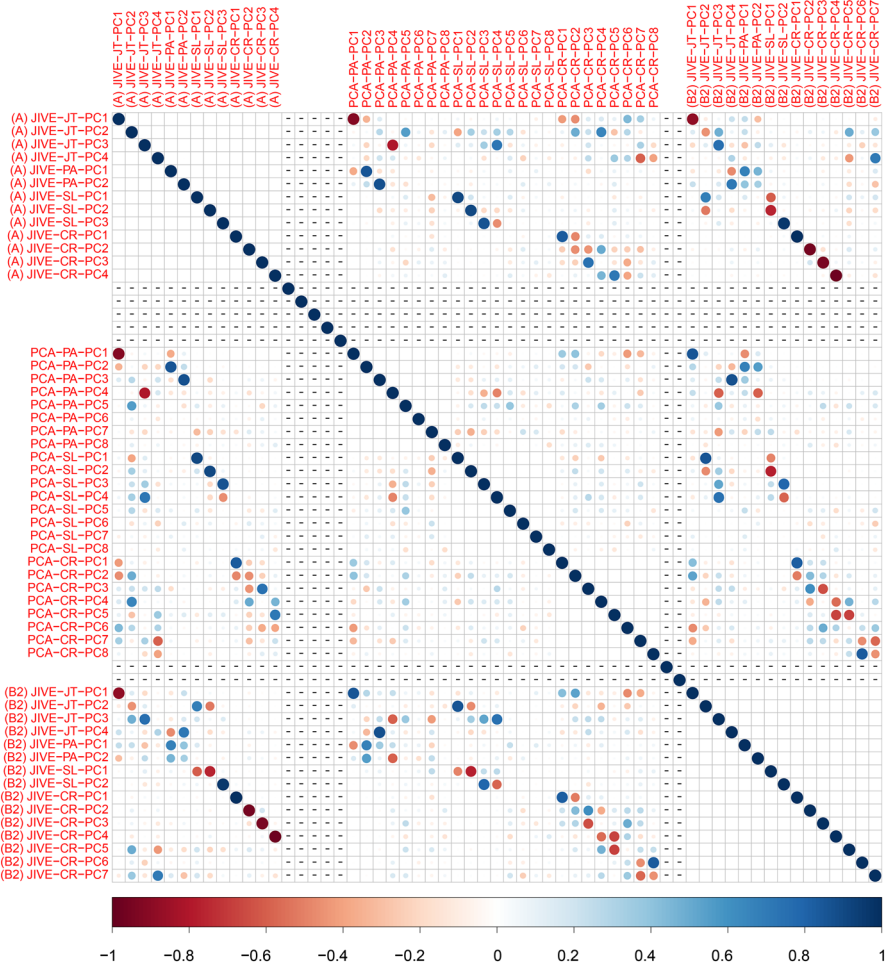
We considered an alternative two-step approach by first performing PCA to each domain individually, and then applying JIVE to the joint truncated data from all three domains.

In order to keep as much joint variation as possible, we considered in the first step (i.e., individual PCA) to keep 5 and 8 PCs from each domain. When 5 PCs were kept in each domain, JIVE selected joint rank to be 4, and 1, 2, and 5 as the individual ranks for PA, SL, and CR, respectively. Meanwhile, when 8 PCs were kept in each domain, JIVE selected joint rank to be 4 as well, but individual ranks to be 2, 2, and 7 for domains for PA, SL, and CR, respectively. Figure 12 shows the proportion of variation explained. Even though it seems like there are more joint variation recovered, but we have to keep in mind that potentially useful information may have been removed, and the proportion here is with respect to the remained variation in the truncated data.

Figures 13 and 14 show the relationships between JIVE results from the main analysis (denoted by (A)) and the alternative analysis with 5 PCs selected (denoted by (B1)), and with 8 PCs selected (denoted by (B2)). The first joint PCs between (A) and (B1)/(B2) are highly correlated, which shows high consistency between the two approaches. There is also certain level of correlation between JT-PC2 from the two approaches. The individual PCs from the two approaches are highly correlated.



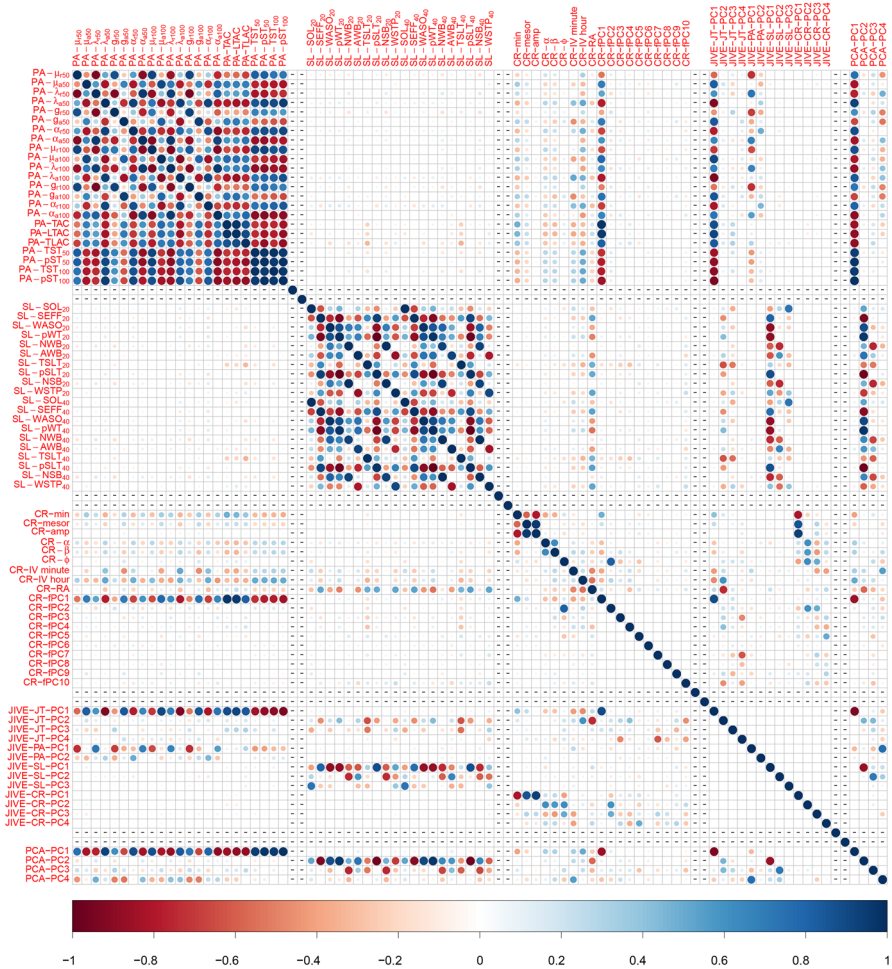




**Fig. 14** Cross-correlations between (A) original JIVE and (B2) two-step JIVE with 8 PCs selected in the first step

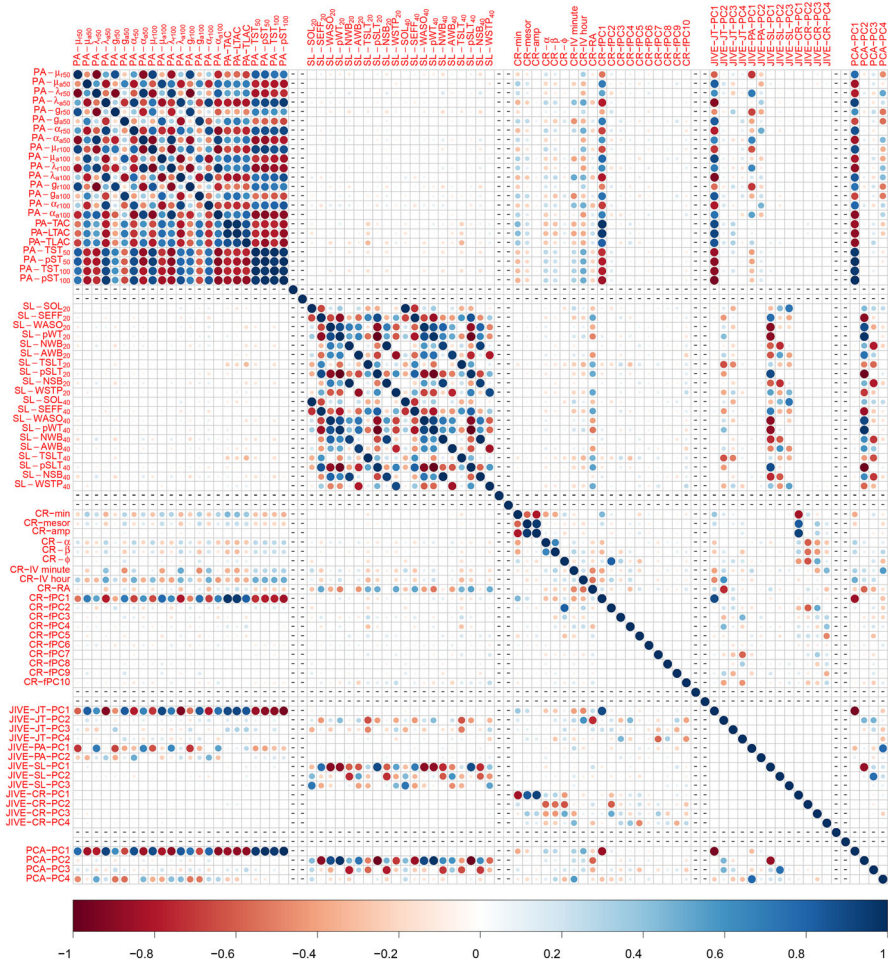
0.59). In sensitivity analysis 3 (S3), we considered the most aggressive approach which is to remove all subject days with missing values. This approach left us 189 subjects contributed by 986 days (with mean 5.22 days and standard deviation of 0.98).

The patterns are quite similar to what we observed from the main analysis. In S1, joint structure explains 68.1%, 20.1%, and 25.8% of total variation in PA, SL, and CR domains, respectively. Individual structures explain 26.1%, 74.5%, and 38.5% of total variation in PA, SL, and CR domains, respectively. And there are only 5.8%, 5.3%, and 35.7% of variation that cannot be explained in PA, SL, and CR respectively. In S2, joint structure explains 69.0%, 18.7%, and 25.9% of total variation in PA, SL, and CR domains, respectively. Individual structures explain 25.2%, 76.2%, and 38.6% of

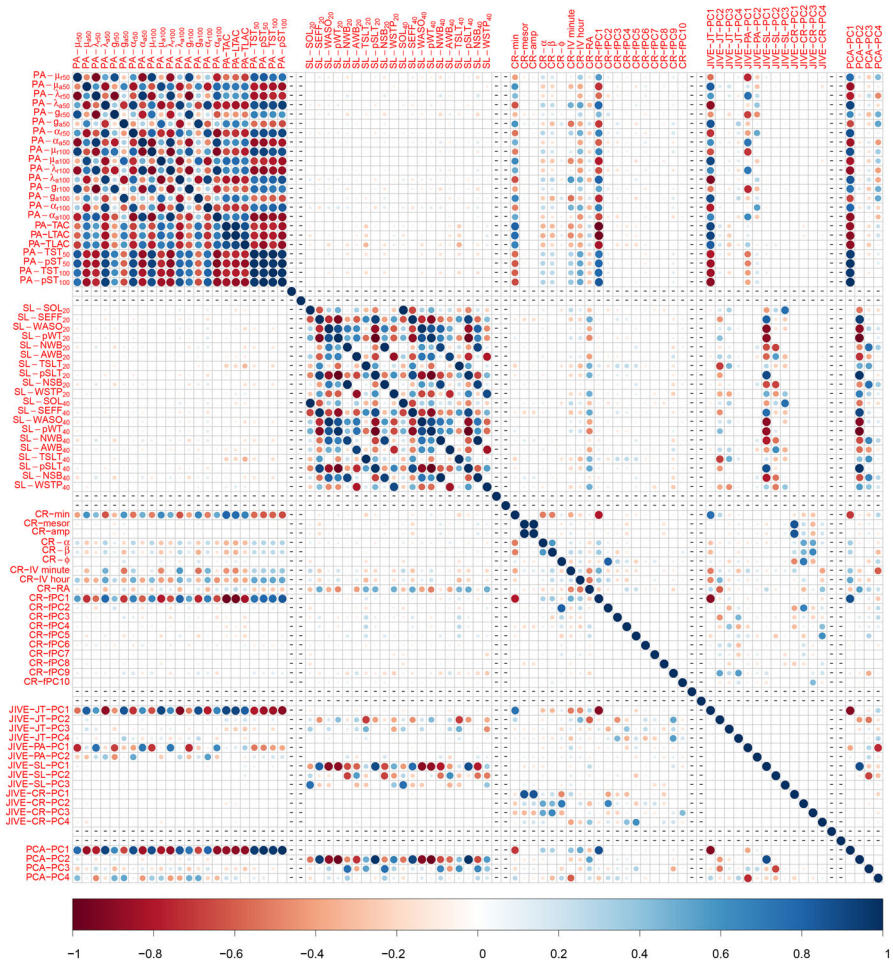


**Fig. 15** Cross-correlations between features and JIVE/PC scores, for S1, where median is used to impute missing data

total variation in PA, SL, and CR domains, respectively. And there are only 5.7%, 5.2%, and 35.5% of variation that cannot be explained in PA, SL, and CR respectively. Finally, in S3, joint structure explains 63.6%, 20.7%, and 27.2% of total variation in PA, SL, and CR domains, respectively. Individual structures explain 30.5%, 73.0%, and 35.3% of total variation in PA, SL, and CR domains, respectively. And there are only 6.0%, 6.3%, and 37.5% of variation that cannot be explained in PA, SL, and CR, respectively. The cross-correlation plots between features and JIVE/PC scores are shown in Figs. 15, 16, and 17.



**Fig. 16** Cross-correlations between features and JIVE/PC scores, for S2, where days with more than 7% were removed



**Fig. 17** Cross-correlations between features and JIVE/PC scores, for S3, where all subject days with missing values were removed

## References

1. Actigraph (2018a) What are counts. <https://actigraph.desk.com/customer/en/portal/articles/2515580-what-are-counts->
2. Actigraph (2018b) What does the “detect sleep periods” button do and how does it work? <https://actigraph.desk.com/customer/en/portal/articles/2515836-what-does-the-%22detect-sleep-periods%22-button-do-and-how-does-it-work->
3. Ancoli-Israel S, Martin JL, Blackwell T, Buenaver L, Liu L, Meltzer LJ, Sadeh A, Spira AP, Taylor DJ (2015) The sbm guide to actigraphy monitoring: clinical and research applications. *Behav Sleep Med* 13(sup1):S4–S38
4. Argiropoulou EC, Michalopoulou M, Aggeloussis N, Avgerinos A (2004) Validity and reliability of physical activity measures in greek high school age children. *J Sports Sci Med* 3(3):147
5. Ayas NT, White DP, Al-Delaimy WK, Manson JE, Stampfer MJ, Speizer FE, Patel S, Hu FB (2003a) A prospective study of self-reported sleep duration and incident diabetes in women. *Diabetes Care* 26(2):380–384. <https://doi.org/10.2337/diacare.26.2.380>




6. Ayas NT, White DP, Manson JE, Stampfer MJ, Speizer FE, Malhotra A, Hu FB (2003b) A prospective study of sleep duration and coronary heart disease in women. *Arch Intern Med* 163(2):205. <https://doi.org/10.1001/archinte.163.2.205>
7. Bai J, Di C, Xiao L, Evenson KR, LaCroix AZ, Crainiceanu CM, Buchner DM (2016) An activity index for raw accelerometry data and its comparison with other activity metrics. *PLoS ONE* 11(8):e0160. <https://doi.org/10.1371/journal.pone.0160644>
8. Bassett DR (2012) Device-based monitoring in physical activity and public health research. *Physiol Meas* 33(11):1769–1783. <https://doi.org/10.1088/0967-3334/33/11/1769>
9. Bassett DR, Troiano RP, McClain JJ, Wolff DL (2015) Accelerometer-based physical activity: total volume per day and standardized measures. *Med Sci Sports Exerc* 47(4):833–8
10. Belletiere J, Winkler EA, Chastin SF, Kerr J, Owen N, Dunstan DW, Healy GN (2017) Associations of sitting accumulation patterns with cardio-metabolic risk biomarkers in Australian adults. *PLoS ONE* 12(6):e0180,119
11. Buman MP, Winkler EAH, Kurka JM, Hekler EB, Baldwin CM, Owen N, Ainsworth BE, Healy GN, Gardiner PA (2014) Reallocating time to sleep, sedentary behaviors, or active behaviors: associations with cardiovascular disease risk biomarkers, NHANES 2005–2006. *Am J Epidemiol* 179(3):323–334. <https://doi.org/10.1093/aje/kwt292>
12. Chau JY, Grunseit AC, Chey T, Stamatakis E, Brown WJ, Matthews CE, Bauman AE, van der Ploeg HP (2013) Daily sitting time and all-cause mortality: a meta-analysis. *PLoS ONE* 8(11):e80,000. <https://doi.org/10.1371/journal.pone.0080000>
13. Cornelissen G (2014) Cosinor-based rhythmometry. *Theor Biol Med Model* 11(1):16
14. Di J, Leroux A, Urbanek J, Varadhan R, Spira A, Schrack J, Zipunnikov V (2017) Patterns of sedentary and active time accumulation are associated with mortality in US adults: the NHANES study. <http://biorxiv.org/content/early/2017/08/31/182337.abstract>
15. Diaz KM, Howard VJ, Hutto B, Colabianchi N, Vena JE, Safford MM, Blair SN, Hooker SP (2017) Patterns of sedentary behavior and mortality in U.S. middle-aged and older adults. *Ann Intern Med* 167:465–475. <https://doi.org/10.7326/M17-0212>
16. Dunlap JC, Loros JJ (2004) *Chronobiology: biological timekeeping*. Sinauer Associates, Sunderland
17. Dunstan DW, Kingwell BA, Larsen R, Healy GN, Cerin E, Hamilton MT, Shaw JE, Bertovic DA, Zimmet PZ, Salmon J, Owen N (2012) Breaking up prolonged sitting reduces postprandial glucose and insulin responses. *Diabetes Care* 35(5):976–983. <https://doi.org/10.2337/dc11-1931>
18. Edgar RS, Green EW, Zhao Y, van Ooijen G, Olmedo M, Qin X, Xu Y, Pan M, Valekunja UK, Feeney KA, Maywood ES, Hastings MH, Baliga NS, Merrow M, Millar AJ, Johnson CH, Kyriacou CP, O’Neill JS, Reddy AB (2012) Peroxiredoxins are conserved markers of circadian rhythms. *Nature* 485(7399):459. <https://doi.org/10.1038/nature11088>
19. Feng Q, Hannig J, Marron J (2015) Non-iterative joint and individual variation explained. *arXiv preprint arXiv:1512.04060*
20. Ferrucci L (2008) The Baltimore longitudinal study of aging (BLSA): a 50-year-long journey and plans for the future. *J Gerontol Ser A* 63(12):1416–1419. <https://doi.org/10.1093/gerona/63.12.1416>
21. Ferrucci L, Alley D (2007) Obesity, disability, and mortality: a puzzling link. *Arch Intern Med* 167(8):750–751
22. Gaynanova I, Li G (2017) Structural learning and integrative decomposition of multi-view data. *arXiv preprint arXiv:1707.06573*
23. Goldsmith J, Zipunnikov V, Schrack J (2015) Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics* 71(2):344–353
24. Gonçalves BS, Cavalcanti PR, Tavares GR, Campos TF, Araujo JF (2014) Nonparametric methods in actigraphy: an update. *Sleep Sci* 7(3):158–164
25. Hamilton MT, Healy GN, Dunstan DW, Zderic TW, Owen N (2008) Too little exercise and too much sitting: inactivity physiology and the need for new recommendations on sedentary behavior. *Curr Cardiovasc Risk Rep* 2(4):292–298. <https://doi.org/10.1007/s12170-008-0054-8>
26. Hauri PJ, Wisbey J (1992) Wrist actigraphy in insomnia. *Sleep* 15(4):293–301. <https://doi.org/10.1093/sleep/15.4.293>
27. Healy GN, Clark BK, Winkler EA, Gardiner PA, Brown WJ, Matthews CE (2011) Measurement of adults’ sedentary time in population-based studies. *Am J Prev Med* 41(2):216–227. <https://doi.org/10.1016/j.amepre.2011.05.005>
28. Hofstra WA, de Weerd AW (2008) How to assess circadian rhythm in humans: a review of literature. *Epilepsy Behav* 13(3):438–444. <https://doi.org/10.1016/j.yebeh.2008.06.002>

29. Hotelling H (1936) Relations between two sets of variates. *Biometrika* 28(3—4):321–377. <https://doi.org/10.2307/2333955>
30. Jean-Louis G, von Gizycki H, Zizi F, Fookson J, Spielman A, Nunes J, Fullilove R, Taub H (1996) Determination of sleep and wakefulness with the actigraph data analysis software (ADAS). *Sleep* 19(9):739–743
31. Jean-Louis G, Kripke DF, Cole RJ, Assmus JD, Langer RD (2001) Sleep detection with an accelerometer actigraph: comparisons with polysomnography. *Physiol Behav* 72(1—2):21–28. [https://doi.org/10.1016/S0031-9384\(00\)00355-3](https://doi.org/10.1016/S0031-9384(00)00355-3)
32. John D, Freedso P (2012) ActiGraph and actual physical activity monitors: a peek under the hood. *Med Sci Sports Exerc* 44(1 Suppl 1):S86–89. <https://doi.org/10.1249/MSS.0b013e3182399f5e>
33. Jondeau E, Jurczenko E, Rockinger M (2010) Moment component analysis: an illustration with international stock markets. *SSRN Electron J* 36(4):576–598. <https://doi.org/10.2139/ssrn.1694643>
34. Kaplan A, Lock EF (2017) Prediction with dimension reduction of multiple molecular data sources for patient survival. arXiv preprint [arXiv:1704.02069](https://arxiv.org/abs/1704.02069)
35. Koster A, Caserotti P, Patel KV, Matthews CE, Berrigan D, Van Domelen DR, Brychta RJ, Chen KY, Harris TB (2012) Association of sedentary time with mortality independent of moderate to vigorous physical activity. *PLoS ONE* 7(6):e37,696. <https://doi.org/10.1371/journal.pone.0037696>
36. Kripke DF, Garfinkel L, Wingard DL, Klauber MR, Marler MR (2002) Mortality associated with sleep duration and insomnia. *Arch Gen Psychiatr* 59(2):131. <https://doi.org/10.1001/archpsyc.59.2.131>
37. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC (2001) Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med* 2(5):389–396. [https://doi.org/10.1016/S1389-9457\(00\)00098-8](https://doi.org/10.1016/S1389-9457(00)00098-8)
38. Leroux A, Di J, Smirnova E, McGuffey EJ, Cao Q, Bayatmokhtari E, Tabacu L, Zipunnikov V, Urbanek JK, Crainiceanu C (2019) Organizing and analyzing the activity data in NHANES. *Stat Biosci*. <https://doi.org/10.1007/s12561-018-09229-9>
39. Li G, Gaynanova I (2017) A general framework for association analysis of heterogeneous data. arXiv preprint [arXiv:1707.06485](https://arxiv.org/abs/1707.06485)
40. Lock EF, Hoadley KA, Marron JS, Nobel AB (2013) Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *Ann Appl Stat* 7(1):523–542. <https://doi.org/10.1214/12-AOAS597>
41. Marler MR, Gehrman P, Martin JL, Ancoli-Israel S (2006) The sigmoidally transformed cosine curve: a mathematical model for circadian rhythms with symmetric non-sinusoidal shapes. *Stat Med* 25(22):3893–3904. <https://doi.org/10.1002/sim.2466>
42. Matthews CE, George SM, Moore SC, Bowles HR, Blair A, Park Y, Troiano RP, Hollenbeck A, Schatzkin A (2012) Amount of time spent in sedentary behaviors and cause-specific mortality in US adults. *Am J Clin Nutr* 95(2):437–445. <https://doi.org/10.3945/ajcn.111.019620>
43. Miettinen J, Taskinen S, Nordhausen K, Oja H (2015) Fourth moments and independent component analysis. *Stat Sci* 30(3):372–390. <https://doi.org/10.1214/15-STSS520>
44. Mills JN (1966) Human circadian rhythms. *Physiol Rev* 46(1):128–171
45. Morton J, Lim LH (2009) Principal cumulant component analysis. preprint
46. Nakazaki K, Kitamura S, Motomura Y, Hida A, Kamei Y, Miura N, Mishima K (2014) Validity of an algorithm for determining sleep/wake states using a new actigraph. *J Physiol Anthropol* 33(1):31. <https://doi.org/10.1186/1880-6805-33-31>
47. Nastasi AJ, Ahuja A, Zipunnikov V, Simonsick EM, Ferrucci L, Schrack JA (2018) Objectively measured physical activity and falls in well-functioning older adults. *Am J Phys Med Rehabil* 97(4):255–260. <https://doi.org/10.1097/PHM.0000000000000830>
48. Owen N, Healy GN, Matthews CE, Dunstan DW (2010) Too much sitting: the population-health science of sedentary behavior. *Exerc Sport Sci Rev* 38(3):105–113. <https://doi.org/10.1097/JES.0b013e3181e373a2>
49. Patel SR, Malhotra A, Gao X, Hu FB, Neuman MI, Fawzi WW (2012) A prospective study of sleep duration and pneumonia risk in women. *Sleep* 35(1):97–101. <https://doi.org/10.5665/sleep.1594>
50. Pyrkov TV, Slipensky K, Barg M, Kondrashin A, Zhurov B, Zenin A, Pyatnitskiy M, Menshikov L, Markov S, Fedichev PO (2018) Extracting biological age from biomedical data via deep learning: too much of a good thing? *Sci Rep* 8(1):5210
51. Ramsay JO (2006) *Functional data analysis*. Wiley, New York
52. Ringnér M (2008) What is principal component analysis? *Nat Biotechnol* 26(3):303–304. <https://doi.org/10.1038/nbt0308-303>

53. Schrack JA, Zipunnikov V, Goldsmith J, Bai J, Simonsick EM, Crainiceanu C, Ferrucci L (2014) Assessing the “Physical Cliff”: detailed quantification of age-related differences in daily patterns of physical activity. *J Gerontol Ser A*. <https://doi.org/10.1093/gerona/glt1199>
54. Schrack JA, Kuo PL, Wanigatunga AA, Di J, Simonsick EM, Spira AP, Ferrucci L, Zipunnikov V (2018a) Active-to-sedentary behavior transitions, fatigability, and physical functioning in older adults. *J Gerontol* 74(4):560–567. <https://doi.org/10.1093/gerona/gly243>
55. Schrack JA, Leroux A, Fleg JL, Zipunnikov V, Simonsick EM, Studenski SA, Crainiceanu C, Ferrucci L (2018b) Using heart rate and accelerometry to define quantity and intensity of physical activity in older adults. *J Gerontol* 73(5):668–675
56. Shephard RJ (2003) Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med* 37(3):197–206
57. Shou H, Zipunnikov V, Crainiceanu CM, Greven S (2015) Structured functional principal component analysis. *Biometrics* 71(1):247–257
58. Spira AP, An Y, Peng Y, Wu MN, Simonsick EM, Ferrucci L, Resnick SM (2017) Apoe genotype and nonrespiratory sleep parameters in cognitively intact older adults. *Sleep* 40(8):zszx076
59. Stone JL, Norris AH (1966) Activities and attitudes of participants in the baltimore longitudinal study. *J Gerontol* 21(4):575–580. <https://doi.org/10.1093/geronj/21.4.575>
60. Taheri S, Lin L, Austin D, Young T, Mignot E (2004) Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index. *PLoS Med* 1(3):e62. <https://doi.org/10.1371/journal.pmed.0010062>
61. Tilmanne J, Urbain J, Kothare MV, Wouwer AV, Kothare SV (2009) Algorithms for sleep-wake identification using actigraphy: a comparative study and new results. *J Sleep Res* 18(1):85–98. <https://doi.org/10.1111/j.1365-2869.2008.00706.x>
62. Tran T, Phung D, Venkatesh S (2011) Mixed-variate restricted boltzmann machines. In: Asian conference on machine learning, pp 213–229
63. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, Chastin SF, Altenburg TM, Chinapaw MJ (2017) Sedentary behavior research network (SBRN)—terminology consensus project process and outcome. *Int J Behav Nutr Phys Activity* 14(1):75. <https://doi.org/10.1186/s12966-017-0525-8>
64. Troiano R, Berrigan D, Dodd K, Masse L, Tilert T, McDowell M (2008) Physical activity in the united states measured by accelerometer. *Med Sci Sports Exerc* 40(1):181
65. Troiano RP, Macera CA, Ballard-Barb R (2001) Be physically active each day. How can we know? *J Nutr* 131(2S–1):451S–460S
66. Udell M, Horn C, Zadeh R, Boyd S et al (2016) Generalized low rank models. *Found Trends Mach Learn* 9(1):1–118
67. Urbanek JK, Spira AP, Di J, Leroux A, Crainiceanu C, Zipunnikov V (2018) Epidemiology of objectively measured bedtime and chronotype in us adolescents and adults: Nhanes 2003–2006. *Chronobiol Int* 35(3):416–434. <https://doi.org/10.1080/07420528.2017.1411359>
68. van Hees VT, Sabia S, Jones SE, Wood AR, Anderson KN, Kivimäki M, Frayling TM, Pack AI, Bucan M, Trenell M et al (2018) Estimating sleep parameters using an accelerometer without sleep diary. *Sci Rep* 8(1):12,975
69. van Someren EJ, Hagebeuk EE, Lijzenga C, Scheltens P, de Rooij SE, Jonker C, Pot AM, Mirmiran M, Swaab DF (1996) Circadian rest—activity rhythm disturbances in Alzheimer’s disease. *Biol Psychiatry* 40(4):259–270. [https://doi.org/10.1016/0006-3223\(95\)00370-3](https://doi.org/10.1016/0006-3223(95)00370-3)
70. Van Someren EJW, Swaab DF, Colenda CC, Cohen W, McCall WV, Rosenquist PB (1999) Bright light therapy: improved sensitivity to its effects on rest-activity rhythms in alzheimer patients by application of nonparametric methods. *Chronobiol Int* 16(4):505–518. <https://doi.org/10.3109/07420529908998724>
71. Varma VR, Dey D, Leroux A, Di J, Urbanek J, Xiao L, Zipunniko V (2017a) Re-evaluating the effect of age on physical activity over the lifespan. *Prev Med* 101:102–108. <https://doi.org/10.1016/j.ypmed.2017.10.028>
72. Varma VR, Dey D, Leroux A, Di J, Urbanek J, Xiao L, Zipunnikov V (2017b) Total volume of physical activity: TAC, TLAC or TAC( $\lambda$ ). *Prev Med*. <https://doi.org/10.1016/j.ypmed.2017.10.028>
73. Varo JJ, Martínez-González MA, de Irala-Estévez J, Kearney J, Gibney M, Martínez JA (2003) Distribution and determinants of sedentary lifestyles in the European Union. *Int J Epidemiol* 32(1):138–146. <https://doi.org/10.1093/ije/dyg116>

74. Wanigatunga AA, Simonsick EM, Zipunnikov V, Spira AP, Studenski S, Ferrucci L (2018) Perceived fatigability and objective physical activity in mid- to late-life. *J Gerontol* 73(5):630–635. <https://doi.org/10.1093/gerona/glx181>
75. Witting W, Kwa I, Eikelenboom P, Mirmiran M, Swaab D (1990) Alterations in the circadian rest-activity rhythm in aging and Alzheimer's disease. *Biol Psychiatr* 27(6):563–572
76. Wold H (2004) Partial least squares. In: *Encyclopedia of statistical sciences*, John, Hoboken. <https://doi.org/10.1002/0471667196.ess1914>
77. Wolff-Hughes DL, Fitzhugh EC, Bassett DR, Churilla JR (2015) Waist-worn actigraphy: population-referenced percentiles for total activity counts in us adults. *J Phys Activity Health* 12(4):447–453
78. Xiao L, Zipunnikov V, Ruppert D, Crainiceanu C (2016) Fast covariance estimation for high-dimensional functional data. *Stat Comput* 26(1–2):409–421

## Affiliations

Junrui Di<sup>1</sup>  · Adam Spira<sup>2,3,4</sup> · Jiawei Bai<sup>5</sup> · Jacek Urbanek<sup>6</sup> · Andrew Leroux<sup>1</sup> · Mark Wu<sup>7</sup> · Susan Resnick<sup>8</sup> · Eleanor Simonsick<sup>8</sup> · Luigi Ferrucci<sup>8</sup> · Jennifer Schrack<sup>2,9</sup> · Vadim Zipunnikov<sup>2,10</sup>

- <sup>1</sup> Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe St. E3039, Baltimore, MD 21205, USA
- <sup>2</sup> Johns Hopkins Center on Aging and Health, 2024 E. Monument Street Suite 2-700, Baltimore, MD 21205, USA
- <sup>3</sup> Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, 624 N. Broadway RM794, Baltimore, MD 21205, USA
- <sup>4</sup> Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, 624 N. Broadway RM794, Baltimore, MD 21205, USA
- <sup>5</sup> Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe St. E3136, Baltimore, MD 21205, USA
- <sup>6</sup> Division of Geriatric Medicine and Gerontology, Department of Medicine, Johns Hopkins University School of Medicine, 2024 E. Monument Street Suite 2-700, Baltimore, MD 21205, USA
- <sup>7</sup> Department of Neurology and Neuroscience, Johns Hopkins University School of Medicine, 855 N. Wolfe Street, Baltimore, MD 21205, USA
- <sup>8</sup> Intramural Research Program, National Institute on Aging, National Institutes of Health, 251 Bayview Boulevard Suite 100, Baltimore, MD 21224, USA
- <sup>9</sup> Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe St. E7144, Baltimore, MD 21205, USA
- <sup>10</sup> Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe St. E3644, Baltimore, MD 21205, USA