Robust identification of temporal biomarkers in longitudinal omics studies

Ahmed A. Metwally, Tom Zhang, Si Wu, Ryan Kellogg, Wenyu Zhou, Kevin Contrepois, Hua Tang and Michael Snyder

1Department of Genetics, Stanford University, Stanford, CA 94305, USA, 2Illumina Artificial Intelligence Laboratory, Illumina Inc., San Diego, CA 92122, USA, 3Systems and Biomedical Engineering Department, Faculty of Engineering, Cairo University, Giza 12613, Egypt, 4Department of Computer Science, Columbia University, New York, NY 10027, USA and 5Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

*To whom correspondence should be addressed.

Abstract

Motivation: Longitudinal studies increasingly collect rich ‘omics’ data sampled frequently over time and across large cohorts to capture dynamic health fluctuations and disease transitions. However, the generation of longitudinal omics data has preceded the development of analysis tools that can efficiently extract insights from such data. In particular, there is a need for statistical frameworks that can identify not only which omics features are differentially regulated between groups but also over what time intervals. Additionally, longitudinal omics data may have inconsistencies, including non-uniform sampling intervals, missing data points, subject dropout and differing numbers of samples per subject.

Results: In this work, we developed OmicsLonDA, a statistical method that provides robust identification of time intervals of temporal omics biomarkers. OmicsLonDA is based on a semi-parametric approach, in which we use smoothing splines to model longitudinal data and infer significant time intervals of omics features based on an empirical distribution constructed through a permutation procedure. We benchmarked OmicsLonDA on five simulated datasets with diverse temporal patterns, and the method showed specificity greater than 0.99 and sensitivity greater than 0.87. Applying OmicsLonDA to the iPOP cohort revealed temporal patterns of genes, proteins, metabolites and microbes that are differentially regulated in male versus female subjects following a respiratory infection. In addition, we applied OmicsLonDA to a longitudinal multi-omics dataset of pregnant women with and without preeclampsia, and OmicsLonDA identified potential lipid markers that are temporally significantly different between the two groups.

Availability and implementation: We provide an open-source R package (https://bioconductor.org/packages/OmicsLonDA), to enable widespread use.

Contact: ametwall@stanford.edu or mpsnyder@stanford.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

Introduction

Human health is highly dynamic, and there is great interest in better understanding how wellness and disease states fluctuate over time in relation to different variables such as lifestyle or treatment perturbations. While genetics provides a blueprint for life, health states are also reflected by many other ‘omics’ such as transcriptomics, proteomics, metabolomics, lipidomics, and microbiomics. With rapid advances and decreasing costs in sequencing and mass spectrometry, many studies are beginning to measure comprehensive omics profiles at frequent timepoints across many individuals. Longitudinal omics studies generate enormous datasets; however, there is currently a major bottleneck in analyzing these data to extract and interpret meaningful findings. In particular, there is a need for robust statistical methods for longitudinal omics.

Longitudinal omics data have their own properties that differentiate them from cross-sectional experiments, including high dimensional feature space, temporal and intrapersonal variation, and samples characterized by heterogeneity of various natures. These heterogeneities include a different number of samples per subject, uncaptured data points, variable time of sample collection ‘sampled non-uniformly’ and omics features often represent a biological...
Robust identification of temporal biomarkers in longitudinal omics studies

2. Materials and methods

The proposed method, OmicsLonDA, aims to find the feature’s significant time intervals (FSI) of differences between each pair of the tested groups (e.g. healthy versus diseased, male versus female, etc.). The method works on unpaired experiment design, where subjects’ longitudinal samples are related to only one of the tested groups. We model the longitudinal data in a time-series model using a spline kernel. Although, in theory, longitudinal data should be correlated, the first-order auto-correlation is not high [e.g. 0.19 in the longitudinal microbiome data from the Human Microbiome Project (Zhou et al., 2019)]. This is mainly due to the fact that longitudinal samples are taken far apart from each other [e.g. weeks to months in Zhou et al. (2019)]. For this reason, we do not consider auto-correlation in our model due to the complexity of assuming a valid dependency structure.

In this article, we introduce a robust method to perform longitudinal differential analysis on omics features in order to identify time intervals of differences between the study phenotypes. Another method, MetaLonDA, has been proposed to find time intervals of significant microbial biomarkers using a permutation test (Mertwally et al., 2018). MetaLonDA is tailored to microbial experiments through the use of a negative binomial distribution. Similarly, the method introduced recently by Santos Nobre and da Motta Singer (2007) in many situations, these assumptions are violated. With the rapidly increasing size and complexity of omics datasets, non-parametric methods (Conover, 1999; Efron and Tibshirani, 1994) are emerging as the primary methods for biomedical analysis. Non-parametric statistics have the advantage of making minimal distributional assumptions and can scale to fit the complexity of the data. A recent non-parametric robust method, bootLong, was developed for extracting microbial biomarkers from longitudinal microbiome data based on a moving block bootstrap approach (Jeganathan et al., 2018). It accounts for within-subject dependency by using overlapping blocks of repeated observations within each subject. It then infers biomarkers based on approximately pivotal statistics. Although bootLong shows promising results in identifying microbial biomarkers in microbial longitudinal studies, it does not provide time intervals of differences between the study phenotypes. Another method, MetaLonDA, has been proposed to find time intervals of significant microbial biomarkers using a permutation test (Mertwally et al., 2018). MetaLonDA is tailored to microbial experiments through the use of a negative binomial distribution. Similarly, the method introduced recently by Santos Nobre and da Motta Singer (2007) is designed to test whether two groups of omics experiments, such as genes expression from RNAseq experiments (Reuter et al., 2015), proteins levels from proteomics experiments (Van Eyk and Snyder, 2018), metabolites intensities from metabolomics experiments (Kellogg et al., 2018), microbial abundance from metagenomics experiments (Mertwally et al., 2016). The data processing output of each of these omics assays can be summarized in a matrix **C** with a dimension of \( m \times n \) where \( m \) denotes the number of omic features and \( n \) denotes the number of samples. \( C_{i,j} \) represents the quantity from sample \( j \) that is annotated to feature \( i \). The proposed method is based on four main steps as shown in Figure 1: (a) adjust measurements based on each subject’s profile, (b) fitting the Gaussian smoothing spline regression model, (c) permutation test to generate an empirical distribution of the test statistic of each time interval and (d) inference of significant time intervals of omics features. The details of the method are described in the following sections.

In this article, we introduce a robust method to perform longitudinal differential analysis on omics features in order to identify time intervals of differences between study groups. The method is based on a semi-parametric approach, where we use smoothing splines to model longitudinal data and infer significant time intervals of changes in omics features based on an empirical distribution constructed through a permutation procedure. The proposed method can handle all types of inconsistencies in sample collections and adjust for subjects’ specific baseline. Identifying biomarkers and their significant time interval differences can inform intervention strategies (drugs, probiotics, antibiotics and supplements), and most importantly, may indicate the best time for interventions to be administered to patients. The method achieved a correctly calibrated type-I error rate and is robust to data collection inconsistencies that commonly occur in longitudinal human studies. Application of the proposed method to iPOP cohort revealed a multitude of sex differences in dynamic respiratory infections. To our knowledge, this is the first study to investigate segmental sexual dimorphism in infection response with frequent temporal sampling and delineation of the dynamic infection response for each sex. We also applied OmicsLonDA on a longitudinal lipidomics study on preeclampsia for the identification of time intervals that lipids are significantly different between pregnancy with and without preeclampsia. We provide an open-source Bioconductor R package, OmicsLonDA (OMICS Longitudinal Differential Analysis), for widespread availability.
the differing baseline measurements between subjects. However, min–max normalization is not robust in handling outlier’s measurement within any time series. Hence, outliers need to be removed as a pre-processing step prior to performing min–max normalization.

\[
y_{ij}^{adj} = \log \frac{y_{ij}}{y_{i0}}. \tag{1}
\]

\[
y_{ij}^{adj} = \frac{y_{ij} - \min(y_{ij})}{\max(y_{ij}) - \min(y_{ij})}. \tag{2}
\]

### 2.2 Fitting Gaussian smoothing spline regression model

For each omic feature \( f = 1, \ldots, F \) from the candidate list, the data under consideration are the random variables \( Y_{t,f}^{adj} \) or their observations \( y_{t,f,i}^{adj} \) of level or mapped reads of the \( i \)th subject of group \( k \) to the feature \( f \) at timepoint \( t \), where \( t = 1, \ldots, T \), \( k = 1, 2 \), and subject \( i = 1, \ldots, n^k \). The random variable \( Y_{t,f}^{adj} \) is assumed to follow \( Y_{t,f}^{adj} \sim N(\eta(y_{t,f,i}^{adj}), \sigma^2) \), where \( \eta(y_{t,f,i}^{adj}) \) is the cubic spline function to be estimated from the data. We seek the estimation of model parameters by solving the penalized likelihood function in (Eq. 3) using a piecewise cubic polynomial minimizer. In the objective function, \( \mathcal{L} = \log L(\eta|Y) \) encourages the goodness of fit, \( \mathcal{J}(\eta) \) is a roughness penalty that is added to the minus log-likelihood to quantify the smoothness of \( \eta \), which is essentially the inner product in a reproducing kernel Hilbert space (Gu, 2013). The \( \lambda \) in (Eq. 3) controls the trade-off between the goodness of fit and the smoothness of the spline and can be determined using cross-validation (Gu, 2013). For each feature, we solve (Eq. 3) for each one of the two tested groups, which leads to two smoothing splines, one for each group.

\[
\min_{\eta} -\mathcal{L} + \lambda \mathcal{J}(\eta). \tag{3}
\]

Once we have the two smoothing splines, one that fits each group’s longitudinal samples, we then calculate the test statistic for each of the \( T - 1 \) time intervals, where \( T \) is the number of time intervals that span the study period. We developed studentized test statistics that quantify differences between the two splines for each time interval. The formula represents the area between the two splines for each time interval \((t, t+1)\) as shown in (Eq. 4), where \( A^k_{t,t+1} \) and \( A^{k+1}_{t,t+1} \) denote the area under the spline curve from time \( t \) to \( t+1 \) for Group 1 and Group 2, respectively, \( t = 1, \ldots, T - 1 \), and SE represents the standard error. Usually, the predicted time intervals are equidistant, as shown in Figure 1. Therefore, (Eq. 4) can be rewritten in terms of the spline function \( \eta \) as shown in (Eq. 5). Under the null hypothesis of no difference between the groups at the specific window, we expect the test statistics to take values near 0, with variance estimated using a permutation procedure described next.

\[
testStatistic_{t,t+1} = \frac{A^k_{t,t+1} - A^{k+1}_{t,t+1}}{\sqrt{SE^k_{t,t+1}^2 + SE^{k+1}_{t,t+1}^2}} \tag{4}
\]

\[
testStatistic_{t,t+1} = \frac{\hat{\phi}_1^k + \phi_1^{k+1}}{\sqrt{SE^k_{t,t+1}^2 + SE^{k+1}_{t,t+1}^2}}. \tag{5}
\]

Under the null hypothesis of no difference between the groups at the specific window, we expect the test statistics to take values near 0, with variance estimated using a permutation procedure described next.
2.3 Inference of significant time intervals via permutation procedure

We perform a permutation procedure by permuting the sample group labels $k$. The permutation is done $B$ times, and after each permutation, we calculate the test statistic $U_{j(t)}^{(b)}$ for the null hypothesis for each time interval. Since all longitudinal samples from the same participant have the same group label after each permutation, the autocorrelation correlation is preserved across subjects, and hence, type-1 error remains the same throughout the permutation procedure. Subsequently, the $p$-value of each interval of the tested feature $f$ is calculated using (Eq. 6) when test statistic $U_{j(t)}^{(b)}$ is positive and (Eq. 7) when it is negative, where $T - 1$ denotes the number of time intervals, and $I(.)$ is an indicator function. The $p$-value is adjusted for multiple testing using Benjamini–Hochberg (B/H) to control for the false discovery rate. For each feature $f$, significant time intervals are those with $p$-value $p_{j,t+1} < \alpha$, where $\alpha$ is the significance level.

$$ pvalue_{j,t+1} = \frac{\sum_{b=1}^{B} \sum_{j=0}^{N} I(testStatistic_{j,t+1}^{(b)} > testStatistic_{j,t+1})}{B \times (T - 1)}$$

where $b = 1, \ldots, B$

(6)

$$ pvalue_{j,t+1} = \frac{\sum_{b=1}^{B} \sum_{j=0}^{N} I(testStatistic_{j,t+1}^{(b)} < testStatistic_{j,t+1})}{B \times (T - 1)}$$

where $b = 1, \ldots, B$

(7)

3. Results

3.1 Performance evaluation using simulated data

To measure the performance of our proposed method in identifying the significant time intervals of omics features, we simulated datasets that mimic all variations in human sample collection, such as non-uniform time gap between samples, subjects have a different number of samples over the study period, a different baseline for each subject, subjects drop out late in the study and missing data. We used SimStudy (Golddle 2018) to simulate five datasets with different patterns across the study course, as shown in Figure 2. We simulated 1000 features from each pattern. The Time 0 indicates the start of the study or the start of the perturbation. In our simulations, we assumed that the covariance structure between consecutive timepoints follows first-order autoregression AR(1) with a correlation coefficient $\rho = 0.4$. Datasets were simulated for 10 individuals from each group. Then, to mimic variability in sample collections, we sampled data points with a variable number of subjects and a variable number of samples per subject.

The generated longitudinal data have a varying number of timepoints as well as varying time intervals between each measurement period. We assumed that the number of timepoints per subject follows a truncated Poisson distribution with $\lambda = 20$.

Simulated omic features of the first pattern (Fig. 2A) were simulated with a mean $\mu(t)$, which follows (Eq. 8), where $N$ denotes normal distribution, $t = 0, \ldots, 500$, and $\sigma \in \{1, 5\}$. The first pattern indicates that the change between the two groups happens 50 days from the start of the perturbation and lasts till 150 days. Pattern 2 (Fig. 2B) shows differences between 100 and 200, Pattern 3 (Fig. 2C) shows differences between 150 and 250, Pattern 4 (Fig. 2D) shows differences between 200 and 300 and Pattern 5 (Fig. 2E) does not have change at all between the two groups and act as a negative control. The purpose of simulating these various patterns is to benchmark the proposed method performance while there are fewer samples at the period that have differences between groups since subjects dropping out of most of the longitudinal studies is directly proportional to the time.

$$ \mu(t) = \begin{cases} N(10, \sigma) + t - 50, & 50 \leq t \leq 100 \\ N(10, \sigma) + 500 - t, & 100 < t \leq 150 \\ N(10, \sigma), & \text{otherwise.} \end{cases}$$

(8)

We evaluated the performance of the OmicsLonDA in identifying significant time intervals from each one of the five patterns described above. In our analysis, we used $B = 1000$ permutations to construct the empirical distribution, a significance level of $\alpha = 0.05$, and adjusted for multiple testing using the BH method. We tested 500 intervals for each feature ($T = 1, \ldots, 501$). Specificity = $\frac{TN}{TN + FN}$ and Sensitivity = $\frac{TP}{TP + FN}$ for each pattern, were measured for each feature independently, where $TP$ is the number of significant time intervals that were correctly identified by the method, $FN$ is the number of significant time intervals that were missed by the method, $TN$ is the number of insignificant time intervals that were identified as insignificant, and $FN$ is the number of significant time intervals that were identified as significant. Then, average specificity and sensitivity were measured among the 1000 features for each pattern. We benchmarked two variants of OmicsLonDA based on the model they use to fit the longitudinal data for each group: (i) OmicsLonDA with smoothing spline ANOVA (OmicsLonDA_SSANOVA) and (ii) OmicsLonDA with Gaussian additive mixed models (OmicsLonDA_GAMM). GAMM allows fitting smoothing terms to model time-series data, and it uses subject ID as a random effect. No covariates were added in this simulation study. We also benchmarked OmicsLonDA against MetaLonDA and splinectomeR. We chose to benchmark against MetaLonDA and splinectomeR since, to our knowledge, they are the two methods that were developed mainly to identify time interval of significance between the two tested groups.

There are three key differences between OmicsLonDA and MetaLonDA: (i) MetaLonDA does not correct for personal baseline, (ii) MetaLonDA uses negative binomial smoothing spline when used with microbiome data and LOESS regression otherwise and (iii) MetaLonDA uses a different formula for test statistic that only include the area between the curves of the two groups without adjusting of the standard error in their estimation. In our benchmarking experiments, we ran MetaLonDA with LOESS regression and 1000 permutations, and all other parameters were left as default. For splinectomeR, we ran the benchmarking evaluation with cut_low = 0 inorder not to filter out any subjects based on small number of time points, and the rest of parameters were left as default.

Table 1 demonstrates the high level of specificity of OmicsLonDA (>0.99) among all five tested patterns, with OmicsLonDA_GAMM having slightly more specificity over the first four patterns, and OmicsLonDA_SSANOVA has slightly more specificity in Pattern 5. On the other hand, MetaLonDA specificity is ~0.80 among all first four patterns, and 0.97 in Pattern 5. splinectomeR followed similar trends to MetaLonDA, high sensitivity and low specificity. Table 1 shows the sensitivity of all benchmarked methods for all patterns, except Pattern 5. This is because all features in Pattern 5 were simulated to not have any significant differences in the time intervals between the two compared groups. splinectomeR has the highest sensitivity (~1) among the compared methods. This high sensitivity can also be seen as a trade-off with the low specificity of splinectomeR. In general, there is a decrease in sensitivity for all methods from Pattern 1 to Pattern 4. This decrease in sensitivity is expected due to the fact that as the time intervals that are significantly different between the two groups shift to the right (later in the study course, which was implemented in our simulations), there are more participants dropping out of the study, and hence there is lower power of each method to detect the significantly differential time intervals. OmicsLonDA_SSANOVA maintains reasonably high sensitivity across all patterns (Pattern 1: 0.98, Pattern 2: 0.92, Pattern 3: 0.90 and Pattern 4: 0.87). OmicsLonDA_GAMM has a similar sensitivity pattern to OmicsLonDA_SSANOVA, but surprisingly, the sensitivity drops significantly at Pattern 4 (0.72). These results demonstrate that OmicsLonDA_SSANOVA is a better choice than OmicsLonDA_GAMM when few samples cover the tested time interval. Additionally, Table 1 shows that low variance in simulated data ($\sigma = 1$ versus $\sigma = 5$ in Eq. 8) increases the sensitivity and specificity of all methods.
Additionally, we evaluated the two implemented personal baseline adjusting methods (log-ratio and min–max normalization) on OmicsLonDA performance. Table 2 shows the sensitivity and specificity of OmicsLonDA when running on the five simulated patterns after each of the baseline adjusting methods. While log-ratio and min–max baseline adjusting methods have yielded a similar effect on OmicsLonDA specificity, min–max has yielded higher sensitivity.

### 3.2 Time and memory evaluation
The running time of OmicsLonDA depends primarily on the number of permutations used to construct the empirical distribution for each feature. Additionally, OmicsLonDA can be run in parallel on multi-core platforms. However, we used one thread in our time evaluation. In our analysis of the simulated data with 1000 permutations, for each feature, OmicsLonDA analysis took, on average, 43 min and 47 s. The evaluation was conducted on a MAC machine with a 2.5 GHz Intel Core i7 processor and 16 GB 1600 MHz RAM.

### 3.3 Application of OmicsLonDA on real-world datasets
#### 3.3.1 Ipop infection multi-omics cohort
As an application for our proposed method, we used the integrative Personal Omics Profiling (iPOP) cohort, a longitudinal cohort that aims to characterize the complex host–microbial interactions in type 2 diabetes mellitus (T2DM) (Zhou et al., 2019). The iPOP cohort was established to better understand T2DM at its earliest stages, where healthy or prediabetic individuals are sampled over ~4 years in a deep multi-omics profiling of transcriptomes, metabolomes, proteomes and cytokines, as well as gut and nasal microbiome. In a total of 1091 visits, 105 participants (25–75 years old, body mass index of 19–41 kg/m², 55 females and 50 males) were profiled during healthy periods and extensively during periods of respiratory viral infection (RVI), immunization and other situations that perturb human host–microbial physiology.

We leveraged the power of the longitudinal multi-omics nature of the iPOP study to reveal sexual dimorphism at the molecular level following RVI episodes. Sex is considered to be an important epidemiological factor that can determine the risk for some diseases. However, the sex-dependent responses to RVIs are not well explored, especially in a multi-omics and microbiological fashion. Most of the previous studies were based on epidemiological strategy and reported the prevalence of RVI in different sex (Channappanavar et al., 2017; Chiarella et al., 2017; Granados et al., 2017; Wang et al., 2016). In this work, we utilized OmicsLonDA to identify longitudinal transcriptomic, metabolomic, cytokines and microbial changes between females and males following RVI. In the context of this work, we included 25 (12 male and 13 female) participants who were followed before and after RVI (44 episodes of RVI in a total of 180 RVI visits; Fig. 3). We selected episodes that have at least three samples during the first 39 days after RVI. We first adjusted each feature using min–max normalization (Eq. 2). Each feature (gene, protein, metabolite, cytokine or

![Fig. 2. Examples of simulated features from the 5 patterns we have in this study. The first pattern indicates that the change between the two groups happened 50 days from the start of the perturbation and lasts till 150 days. Pattern 2 shows differences between 100 and 200. Pattern 3 shows differences between 150 and 250. Pattern 4 shows differences between 200 and 300. Pattern 5 has no change at all between the two groups and acts as a negative control](image-url)
microbe) was tested independently. Time interval inference is based on an empirical distribution that is built for all intervals of the same feature as described previously. We tested 38 intervals (between Day 1 and Day 39). For proteins, metabolites, cytokines and microbes, we tested all quantified features from the iPOP cohort. In our analysis, we used $B = 1000$ permutations to construct the empirical distribution, significance level $\alpha = 0.05$, and adjusted for multiple testing using BH method.

In total, 104 features (36 genes, 29 proteins, 35 metabolites, 3 cytokines and 1 microbe) exhibit temporal differences between males and females following RVI. Figure 4 shows a timeline summary of omic features that show the difference between males and females after RVI episodes (Supplementary Table S1). The results reveal that females were more responsive to RVI with 58 omic features being overexpressed, while 44 features were over-expressed in males, and 2 genes (MFSD7 and SCN5A) flipped their over-expression trajectory during the course of the infection episode. Females have stronger antibody responses IGLV3-19 (LV319) and IGHV3-53 (HV353). Females have a stronger adaptive response early than males, while males have more innate responses than females with increased complement proteins and increased red blood cells (HBA1, HBB, HBH). Interestingly, males have over-expression of vitamin D3 (dihydroxyvitamin) during the early infection period (Day 1 to Day 21). Females have higher leptin during the whole course of infection (1–39 days).

### Table 1. Performance evaluation of identifying significant time intervals from simulated features

| Pattern, Time of change | OmicsLonDA (SSANOVA) $\sigma = 5$ | OmicsLonDA (GAMM) $\sigma = 5$ | OmicsLonDA (SSANOVA) $\sigma = 1$ | OmicsLonDA (GAMM) $\sigma = 1$ |
|-------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
|                         | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity |
| Pattern_1 (early change) | 0.976      | 0.998      | 0.982      | 0.993      | 0.945      | 0.991      | 0.965      | 0.990      |
| Pattern_2 (middle change)| 0.924      | 0.999      | 0.952      | 0.999      | 0.948      | 0.993      | 0.963      | 0.997      |
| Pattern_3 (late change)  | 0.999      | 0.698      | 0.991      | 0.818      | 0.999      | 0.682      | 0.991      | 0.818      |
| Pattern_4 (very late change) | 0.906     | 0.999      | 0.921      | 0.999      | 0.934      | 0.995      | 0.937      | 0.998      |
| Pattern_5 (no change)    | 0.965      | 0.804      | 0.991      | 0.876      | 0.965      | 0.804      | 0.991      | 0.876      |

### Table 2. Evaluation of adjusting subject’s profile

| Pattern, Time of change | OmicsLonDA (log-ratio) $\sigma = 5$ | OmicsLonDA (min–max) $\sigma = 5$ | OmicsLonDA (log-ratio) $\sigma = 1$ | OmicsLonDA (min–max) $\sigma = 1$ |
|-------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
|                         | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity |
| Pattern_1 (early change) | 0.970      | 0.965      | 0.970      | 0.999      | 0.970      | 0.965      | 0.970      | 0.999      |
| Pattern_2 (middle change)| 0.930      | 0.965      | 0.930      | 0.999      | 0.930      | 0.965      | 0.930      | 0.999      |
| Pattern_3 (late change)  | 0.880      | 0.989      | 0.880      | 0.999      | 0.880      | 0.989      | 0.880      | 0.999      |
| Pattern_4 (very late change) | 0.810     | 0.972      | 0.810      | 0.999      | 0.810      | 0.972      | 0.810      | 0.999      |
| Pattern_5 (no change)    | 0.970      | 0.965      | 0.970      | 0.999      | 0.970      | 0.965      | 0.970      | 0.999      |

Fig. 3. Study design of the iPOP infection cohort. Time points distributions of 44 infection episodes whose corresponding subject has at least three timepoints within 40 days following and infection incidence. Total of 180 samples from 25 subjects (12 male and 13 female). Timeline annotation of RVI episodes, where Day 0 is the first day of infection.
3.3.2 Preeclampsia lipidomics cohort

We applied OmicsLonDA on a longitudinal lipidomics study on preeclampsia (Maric et al., 2021) as a case study to demonstrate the value of OmicsLonDA for the identification of time intervals that lipids are significantly different between pregnancy with and without preeclampsia. Preeclampsia is a serious pregnancy complication affecting 5–10% of pregnant women, accounting for approximately 40% of fetal deaths worldwide. It not only harms maternal health but also inhibits fetal growth and causes babies to be born with immature development. Therefore, detection of preeclampsia biomarkers at early gestational age and identification of the time intervals with dramatic lipid changes in preeclampsia is crucial for preeclampsia early diagnosis and treatment.

In this longitudinal prospective study, the cohort was previously described (Maric et al., 2021) with 27 and 20 women with and without preeclampsia, respectively. The plasma samples were collected from each subject at two or three-time points during pregnancy. The gestational age distribution of the preeclampsia and control groups in each trimester is shown in Figure 5A. For each plasma sample, we conducted target lipidomics analysis by applying the Lipidyzer platform for 750 lipid species composed of 13 diverse lipid classes (Figure 5B).

Following OmicsLonDA analysis workflow, we first adjusted the levels of each lipid at later time points using min–max normalization method (Eq.2) and normalized them to baseline. The OmicsLonDA test was conducted for each lipid independently. We set 1 week as one time interval unit and tested on 30 time intervals (Week 8–38). In our analysis, we used 1000 times permutations to construct the empirical distribution for each lipid. All the results were adjusted for multiple testing using BH method with a significance level $\alpha = 0.05$.

We identified 19 lipid species, accounting for eight lipid classes, with significant temporal differences between preeclampsia and control groups during pregnancy. Figure 5C demonstrates the time intervals of these 19 lipids that have significantly different profiles between preeclampsia and control groups (Supplementary Table S2). Interestingly, we found that most of the significant lipids belonging to the same lipid classes exhibit the same changing trends, indicating homogeneity of chemical properties and potential biological roles of lipid species from the same classes. There are two exceptions: one is CER (14:0) and CER (24:0), which may be due to the different lengths of fatty acid chains in these two ceramides. Another interesting exception is TAG 46:3 (FA 18:2), TAG 50:5 (FA16:1) and TAG 56:6 (FA20:4), showing the higher levels of the first two triglycerides in the preeclampsia group at later gestational age compared to the control group. In contrast, TAG 56:6 (FA20:4) displays the increased abundance at early pregnancy in control subjects. Further experiments are needed to investigate the explicit reasons.

4. Discussion

In this work, we have developed a statistical method that provides robust identification of time intervals where omics features are significantly different between groups in longitudinal multi-omics. The
method is able to simultaneously identify time intervals and differential signatures by analyzing each feature separately, but across all patients. The proposed method is based on a semi-parametric approach, where using smoothing splines to model longitudinal data and infer significant time intervals of omics features based on an empirical distribution constructed through the permutation procedure. A critical need in longitudinal omics is for robust frameworks that incorporate the time dimension in statistical significance analysis. Our method was evaluated through extensive simulations (five patterns). The performance evaluation demonstrated that OmicsLonDA has achieved a correctly calibrated type I error rate and is robust to data collection inconsistencies that commonly occur in longitudinal human studies. Moreover, the sensitivity is high in Pattern 1 and then declines slightly through Pattern 4. This decrease in sensitivity can be explained due to the decreasing number of samples collected towards the end of the study (i.e., patient dropout).

We further applied OmicsLonDA to two real-world datasets: (i) the iPOP longitudinal omics study for investigating sexual dimorphism on molecular response following RVI and (ii) preeclampsia cohort to identify time intervals that lipids are significantly different between pregnancy with and without preeclampsia. Recently, OmicsLonDA has been utilized to identify the seasonal time intervals of differentially abundant/expressed omics features between insulin-resistant and insulin-sensitive individuals (Saliani et al., 2020).

Sex differences in response to infection are known (Casimir et al., 2013). For both viral and bacterial infections, males are more susceptible than females, while females produce a more vigorous inflammatory response (Casimir et al., 2013). Sexual dimorphism in infection response likely arises from differences in hormone status, with both testosterone and estrogen shown to modulate infection and inflammatory processes (Chrousos, 2010). Our study further adds evidence that sexual dimorphism may contribute to stages of inflammatory responses (Klein and Flanagan, 2016), with females having a stronger adaptive response early but less innate responses than male. Our analysis is the first to our knowledge that revealed a multitude of sex differences in RSV infection response with frequent temporal sampling and delineation of the dynamic infection response for each sex.

Preeclampsia is a potentially life-threatening complication during pregnancy identified by increased blood pressure and proteinuria. It is one of the leading causes of maternal and perinatal mortality and morbidity (Ghulmiyyah and Sibai, 2012). Substantial efforts have been made to detect molecular changes of preeclampsia during pregnancy at gene, protein and metabolite levels (He et al., 2020; Lapaire et al., 2012; Nobakht M. Gh, 2018). Nowadays, lipids are growingly recognized as key players involved in pathophysiology of preeclampsia (Anand et al., 2016; Wojcik-Baszko et al., 2018). For instance, arachidonic acid and its downstream products were reported to be significantly changed in preeclampsia (Balazy, 2012). Oxidized lipid species were also selected as biomarkers of preeclampsia which are related to increased reactive oxygen species (Butterfield and Lauderback, 2002). However, most of these studies mainly focused on single timepoint instead of monitoring dynamic molecular changes with multiple timepoints during pregnancy. Herein, in this study, we applied the developed OmicsLonDA on a longitudinal lipidomics dataset to compare lipid levels in women with and without preeclampsia. We successfully identified 19 distinct lipid species that are significantly different in preeclampsia pregnancy compared to normal pregnancy with different time intervals. Interestingly, most of the significant lipids that belong to the same lipid classes exhibit the same changing trends, indicating potentially similar biological functions these lipids may exert in preeclampsia progression. Importantly, by OmicsLonDA, we detected several lipids that harbored significantly different time intervals at early pregnancy phase (i.e. the first trimester) such as CE (20:4), PC (17:0/20:4), PC (18:0/20:4) and PC (18:1/20:4), which may serve as clinically meaningful biomarkers for preeclampsia early diagnosis. Intriguingly, all of these four lipid biomarkers share the same fatty acid chain arachidonic acid (fatty acid 20:4). Arachidonic acid is a polyunsaturated fatty acid-containing 20 carbons and four double bonds with a final double bond in the $\omega-6$ position. It is well-documented that arachidonic acid and its products eicosanoids play important roles in inflammatory processes (Higgins and Lees, 1984). They have been reported as biomarkers of preeclampsia previously (Balazy, 2004). Our results not only support the previous findings but also revealed more potentially involved lipid biomarkers by leveraging the advantages of the longitudinal data as well as the merit of OmicsLonDA.
There are two factors that need to be considered before applying OmicsLonDA to any longitudinal dataset: (i) baseline adjustment method and (ii) number of permutations. Firstly, if subjects have samples right before perturbation, baseline adjustment based on log-ratio (Eq. 1) would be preferred since the adjusted time series would reflect the perturbation effect for each person. However, obtaining a sample right before perturbation may not be accessible in many longitudinal multi-omics studies. Hence, adjusting based on min-max (Eq. 2) would be a more feasible solution. The downside of using the min-max method is that the adjusted time series may not be following normal distribution, violating the assumption of Gaussian smoothing spline. This may not be a concern for OmicsLonDA since the identification of significant time intervals is based on a non-parametric permutation test. Secondly, the running time of OmicsLonDA depends primarily on the number of permutations. In our simulations and real-world datasets, we used 1000 permutations to ensure stable null distribution construction. However, we have seen that a similar stable null distribution can be achieved with 100 permutations, which can be used as a fast approach to identify time intervals. However, we do not recommend running OmicsLonDA with less than 100 permutations since the results may not be consistent between different runs.

OmicsLonDA elucidates not only differentially regulated molecules but indicates the temporal window over which the differential regulation occurs to provide a nuanced and detailed understanding of biological dynamics. In the future, we plan to utilize the identified multi-omics features and their significant time intervals through non-parametric Bayesian dynamic networks to infer the causality of phenotypes based on a phased correlation between features’ time intervals and phenotype onset. Another avenue for improving the proposed method is to incorporate auto-correlation between longitudinal samples into the model fitting. Also, in the proposed method, time intervals to be tested are a user-defined parameter. In the future, we plan to develop a learning method that selects non-trivial intervals that span several timepoints. OmicsLonDA is publicly available on the Bioconductor repository (https://bioconductor.org/packages/OmicsLonDA/).

Acknowledgements
We thank Robert Tibshirani from Stanford University for his valuable suggestions and critical comments throughout the development of OmicsLonDA. We would like to thank iPOP participants and numerous colleagues from Stanford University who contributed to the sample collection, preparation and processing of the iPOP cohort.

Funding
This work was supported by NIH Common Fund Human Microbiome Project (HMP) [U154DE02378901], NIH [S10OD020141] to the SGPIM Genome Sequencing Service Center, Stanford Clinical and Translational Science Award [UL1TR001085] and Diabetes Genomics and Analysis Core of the Stanford Diabetes Research Center [P30DK116074].

Conflict of Interest: A.A.M. is currently an employee of Illumina. M.P.S. is a co-founder and a member of the scientific advisory board of Personalis, QBio, January, SentOmics, Protos and Mirense. He is on the scientific advisory board of Danaher, GenapSys and Jupiter. No other authors have competing interests.

Data and Code Availability
OmicsLonDA is publicly available on the Bioconductor repository (https://bioconductor.org/packages/OmicsLonDA/). The development source code is located on Github (https://github.com/aaemetwally/OmicsLonDA), iPOP data are publicly available at the human microbiome project portal (https://portal.hmpdacc.org/).

References
Anand, S. et al. (2016) Detection and confirmation of serum lipid biomarkers for preeclampsia using direct infusion mass spectrometry. J. Lipid Res., 57, 687–696.

Balazy, M. (2004) Eicosanomics: targeted lipidomics of eicosanoids in biological systems. Prostaglandins Other Lipid Mediat., 73, 173–180.

Bar-Joseph, Z. (2004) Analyzing time series gene expression data. Bioinformatics, 20, 2493–2503.

Bordbar, A. et al. (2011) A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. BMC Syst. Biol., 5, 180.

Butterfield, D.A., and Lauderback, C.M. (2002) Lipid peroxidation and protein oxidation in Alzheimer’s disease brain: potential causes and consequences involving amyloid beta-peptide-associated free radical oxidative stress. Free Radic. Biol. Med., 32, 1050–1060.

Casimir, G.J. et al. (2013) Sex and inflammation in respiratory diseases: a clinical viewpoint. Biol. Sex Differ., 4, 16.

Channappanavar, R. et al. (2017) Sex-based differences in susceptibility to severe acute respiratory syndrome coronavirus infection. J. Immunol. (Baltimore, MD 1950), 198, 4046–4053.

Chiarella, F.C. et al. (2017) Characterization and circulation of seasonal influenza viruses in Madrid, 2010-2016. J. Med. Virol., 89, 1726–1733.

Chrousos, G.P. (2010) Stress and sex versus immunity and inflammation. Sci. Signal., 3.

Conover, W.J. (1999) Practical Nonparametric Statistics. Vol. 350. Wiley, New York.

Efron, B. and Tibshirani, R. (1994) An Introduction to the Bootstrap. CRC Press LLC, Boca Raton, Florida.

Fieuws, S. and Verbeke, G. (2004) Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. Stat. Med., 23, 3093–3104.

Ghulmuyah, L. and Sibai, B. (2012) Maternal mortality from preeclampsia/eclampsia. Semin. Perinatol., 36, 56–59.

Goldfeld, K. (2018). Smstud simulator: real data study. https://cran.r-project.org/web/packages/simstudy/index.html.

Gonçalves, E. et al. (2013) Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. Mol. BioSyst., 9, 1576–1583.

Granados, A. et al. (2017) Influenza and rhinovirus viral load and disease severity in upper respiratory tract infections. J. Clin. Virol., 86, 14–19.

Gu, C. (2013) Smooth spline ANOVA Models, 2nd edn. Springer Science & Business Media, New York.

Gueorguieva, R.V. and Sanacora, G. (2006) Joint analysis of repeatedly observed continuous and ordinal measures of disease severity. Stat. Med., 25, 1307–1322.

He, A. et al. (2020) Potential protein biomarkers for preeclampsia. Curr. Eye Res., 12, 16.

Higgins, A.J. and Lees, P. (1984) The acute inflammatory process, arachidonic acid metabolism and the mode of action of anti-inflammatory drugs. Equine Vet J., 16, 163–175.

Jeganathan, P. et al. (2018) The block bootstrap method for longitudinal microbiome data. September.

Kelloff, R.A. et al. (2018) Personal omics for precision health. Circ. Res., 122, 1169–1171.

Klein, S.L. and Flanagan, K.L. (2016) Sex differences in immune responses. Nat. Rev. Immunol., 16, 626–638.

Lapaire, O. et al. (2012) Microarray screening for novel preeclampsia biomarker candidates. Fetal Diagn. Ther., 31, 147–153.

Mardirogla, A. et al. (2013) Integration of clinical data with a genome-scale metabolic model of the human adipocyte. Mol. Syst. Biol., 9, 649.

Maric, I. et al. (2021) Multimomics longitudinal modeling of preeclamptic pregnancies. Research Square, January. https://assets.researchsquare.com/files/rs-138586/v1_covered.pdf?c=1631835001.

Metwally, A.A. et al. (2016) WEVOTE: weighted voting taxonomic identification method of microbial sequences. PLoS One, 11, e0163527.

Metwally, A.A. et al. (2018) MetaLonDA: a flexible R package for identifying time intervals of differentially abundant features in metagenomic longitudinal studies. Microbiome, 6, 32.

Molnbergius, G. and Veerle, G. (2000) Linear Mixed Models for Longitudinal Data. Springer Series in Statistics, New York, NY.

Nobakht M. Gh, B.F. (2018) Application of metabolomics to preeclampsia diagnosis. Syst. Biol. Reprod. Med., 64, 324–339.

Reuter, J.A. et al. (2015) High-throughput sequencing technologies. Mol. Cell, 58, 586–597.

Sailani, M.R. et al. (2020) Deep longitudinal multimomics profiling reveals two biological seasonal patterns in California. Nat. Commun., 11, 1–12.

Santos Nobre, J. and da Motta Singer, J. (2007) Residual analysis for linear mixed models. Biom. J., 49, 863–875.

Shield-Cutler, R.R. et al. (2018) SplinectomeR enables group comparisons in longitudinal microbiome studies. Front. Microbiol., 9, 785.
Storey, J.D. et al. (2005) Significance analysis of time course microarray experiments. Proc. Natl. Acad. Sci. USA, 102, 12837–12842.
Van Eyk, J.E. and Snyder, M.P. (2018) Precision medicine: role of proteomics in changing clinical management and care. J. Proteome Res., 18, 1–6.
Verbeke, G. et al. (2014) The analysis of multivariate longitudinal data: a review. Stat. Methods Med. Res., 23, 42–59.
Wang, H. et al. (2016) Prevalence of respiratory viruses among children hospitalized from respiratory infections in Shenzhen, China. Virol. J., 13, 39.
Wang, Y. et al. (2017) Comparing the performance of approaches for testing the homogeneity of variance assumption in one-factor ANOVA models. Educ. Psychol Meas., 77, 305–329.
Wojcik-Baszko, D. et al. (2018) Role of dyslipidemia in preeclampsia - a review of lipidomic analysis of blood, placenta, syncytiotrophoblast microvesicles and umbilical cord artery from women with preeclampsia. Prostaglandins Other Lipid Mediat., 139, 19–23.
Zhou, W. et al. (2019) Longitudinal multi-omics of host-microbe dynamics in prediabetes. Nature, 569, 663–671.