Research Articles

Open camera or QR reader and
scan code to access this article
and other resources online.

Multi-Omics Integration in a Twin Cohort and Predictive Modeling of Blood Pressure Values

Gabin Drouard,[1] Miina Ollikainen,[1] Juha Mykkänen,[2,3] Olli Raitakari,[2–4] Terho Lehtimäki,[5] Mika Kähönen,[6] Pashupati P. Mishra,[5] Xiaoling Wang,[7] and Jaakko Kaprio[1]

Abstract

Abnormal blood pressure is strongly associated with risk of high-prevalence diseases, making the study of blood pressure a major public health challenge. Although biological mechanisms underlying hypertension at the single omic level have been discovered, multi-omics integrative analyses using continuous variations in blood pressure values remain limited. We used a multi-omics regression-based method, called sparse multi-block partial least square, for integrative, explanatory, and predictive interests in study of systolic and diastolic blood pressure values. Various datasets were obtained from the Finnish Twin Cohort for up to 444 twins. Blocks of omics—including transcriptomic, methylation, metabolomic—data as well as polygenic risk scores and clinical data were integrated into the modeling and supported by cross-validation. The predictive contribution of each omics block when predicting blood pressure values was investigated using external participants from the Young Finns Study. In addition to revealing interesting inter-omics associations, we found that each block of omics heterogeneously improved the predictions of blood pressure values once the multi-omics data were integrated. The modeling revealed a plurality of clinical, transcriptomic, and metabolomic factors consistent with the literature and that play a leading role in explaining unit variations in blood pressure. These findings demonstrate (1) the robustness of our integrative method to harness results obtained by single omics discriminant analyses, and (2) the added value of predictive and exploratory gains of a multi-omics approach in studies of complex phenotypes such as blood pressure.

Keywords: hypertension, blood pressure, twins, multi-omics, phenomics, predictive modeling, sparse multi-block partial least square

Introduction

HYPERTENSION IS A PATHOLOGICAL ELEVATION of blood pressure associated with greater risk of high-prevalence diseases. In particular, hypertension is known to increase the risk of cardiovascular disease (Jordan et al., 2018) as well as cerebrovascular and renal diseases (Kelly and Rothwell, 2020; Ku et al., 2019), making its study of major public health importance. In addition to its broad effects, hypertension has multiple origins, including environmental causes

[1]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland.
[2]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland.
[3]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland.
[4]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland.
[5]Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.
[6]Department of Clinical Physiology, Tampere University Hospital, and Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.
[7]Georgia Prevention Institute (GPI), Medical College of Georgia, Augusta University, Augusta, Georgia, USA.

such as nutrition and excessive alcohol consumption (Puddey et al., 2019; Schwingshackl et al., 2017). It also has a substantial genetic component, as demonstrated by twin and molecular genetic studies (Arnett and Claas, 2018). The existence of genetic and environmental influences on blood pressure further motivates the use of omics data.

The advent of high-throughput technologies has made it possible to obtain sufficiently large volumes of data to highlight significant findings and to gain insight into the biological mechanisms underlying hypertension. Many studies have thus examined the structural and functional genomics of blood pressure using genetic variants and transcriptomics, respectively (Huang et al., 2020; Surendran et al., 2020). Environmental influences have also been investigated, for example, through methylation studies and high-throughput clinical phenotypes in the field of phenomics (Irvin et al., 2021).

Although biological mechanisms underlying hypertension at the single omic level have been discovered, multi-omics integrative analyses using continuous variations in blood pressure values remain limited. Evaluation of the integrated predictive value of various molecular substrates of hypertension is also actively being pursued (Baek et al., 2020; Kwong et al., 2018; Wang et al., 2018). A better understanding of the mechanisms reflecting unitary changes in blood pressure could allow for fine mapping of interindividual differences than those captured by discriminant or categorical analyses. Binary discretization of individuals into normotensive and hypertensive status fails to capture risk factors likely to increase or decrease blood pressure within the normotensive or hypertensive patient groups.

Integration across multiple omics knowledge domains to dissect the phenotypes associated with blood pressure regulation and hypertension is much needed in the present moment. It is in response to these challenges and prospects that this study was undertaken.

We integrated blood pressure data, specifically transcriptomic, methylation, clinical, metabolomic, and polygenic risk scores (PRS) from participants of the Finnish Twin Cohort (FTC) to gain insight into the intra- and inter-omics biological mechanisms underlying unitary increases in systolic blood pressure (SBP) and diastolic blood pressure (DBP). We also present the predictive performance of each of these omics blocks within a multi-omics model based on a regression-type method called sparse multi-block partial least square (sMBPLS). Predictive performance was assessed by comparing the predictions of SBP and DBP values in a test cohort of substantial size with their measured values.

Materials and Methods

Data blocks and sources

The study protocol was approved by the Institutional Ethics Board of the Hospital District of Helsinki and Uusimaa, Finland (ID 154/13/03/00/11) and the Institutional Review Board of Augusta University. Omics datasets were obtained from within the FTC (Kaprio et al., 2019) for up to 444 twins, and all applicable written and informed consent was obtained in relation to the data generated or used for analysis.

Twins were selected based on responses to items on blood pressure and hypertension in the fourth survey of the FTC in 2011–2012; twin pairs with a difference in blood pressure were targeted, as previously described in detail (Kaprio et al., 2019). The twins came in for 1 day of measurement of blood pressure, completed interviews and questionnaires and provided a fasting blood sample for biochemical measures, and samples for omics. In addition, weight, height, and waist and hip circumference were measured (Tuomela et al., 2019).

In total, clinical, metabolomic, methylation, transcriptomic, and PRS data were collected for a subset of this initial number of participants. Metabolomic data for 434 participants were collected with nuclear magnetic resonance spectroscopy and included in the study. The proportion of individuals with methylation (Illumina 450k) and transcriptomic data (Microarray) was lower (360 participants for methylation, 389 participants for transcriptomic data) (Fig. 1). Four PRSs related to SBP, DBP, body mass index (BMI), and coronary artery disease (CAD) were also included.

The preprocessing steps of each omics block before integration into the model sometimes required, for example, imputation of missing values and selection of variables (Supplementary Document, Section S1). Once these preprocessing steps were completed, four omics blocks of different dimensions were considered for the modeling phase (Fig. 1) (Abayomi et al., 2005; Aryee et al., 2014; Benton et al., 2017; Boks et al., 2009; Cazaly et al., 2016; Domingo-Relloso et al., 2021; Du et al., 2008; Friedman et al., 2010; Hayati Rezvan et al., 2015; Honaker et al., 2011; Kaprio et al., 1987; Keil et al., 1991; Lin et al., 2008; Nikpay et al., 2015; Price et al., 2006; Salvador et al., 2019; Triche et al., 2013; van Buuren and Groothuis-Oudshoorn, 2011; Vilaplana, 2006; Waldmann et al., 2013; Yengo et al., 2018; and Zou and Hastie, 2005).

In addition to the FTC participants, data from the Young Finns Study (YFS) (Raitakari et al., 2008) were used for the predictive phase of our study. This test cohort consists of a total of 1350 participants for whom the same omics blocks as described above for the FTC were available (Supplementary Document, Section S2 for details of the methylation preprocessing methodology) (Ahola-Olli et al., 2019; Elovainio et al., 2015; McCartney et al., 2021; Soininen et al., 2015; Võsa et al., 2021). A large number of variables within each block have been retrieved, although some were missing (Performance Criteria and Data Linkage sections). Clinical differences between the YFS and FTC cohorts were noteworthy, as reflected in the blood pressure values and age distributions (Table 1).

Integrative methods

Latent structures and integration.   Partial least square (PLS) regressions, sometimes referred to as latent structure projections, are a family of methods that proceed by deriving latent variables defined as linear combinations of variables (Abdi and Williams, 2013). One of these PLS-based methods adapted to a multi-omics context, called sMBPLSs, was used to integrate the different omics blocks into a single model. sMBPLS calculates latent components for each block (hereafter referred to as block-related components) and for the outcome matrix Y before averaging the block-related components to obtain upscaled latent components (Li et al., 2012). These computations were carried out by iteratively maximizing the covariance between the latent components,
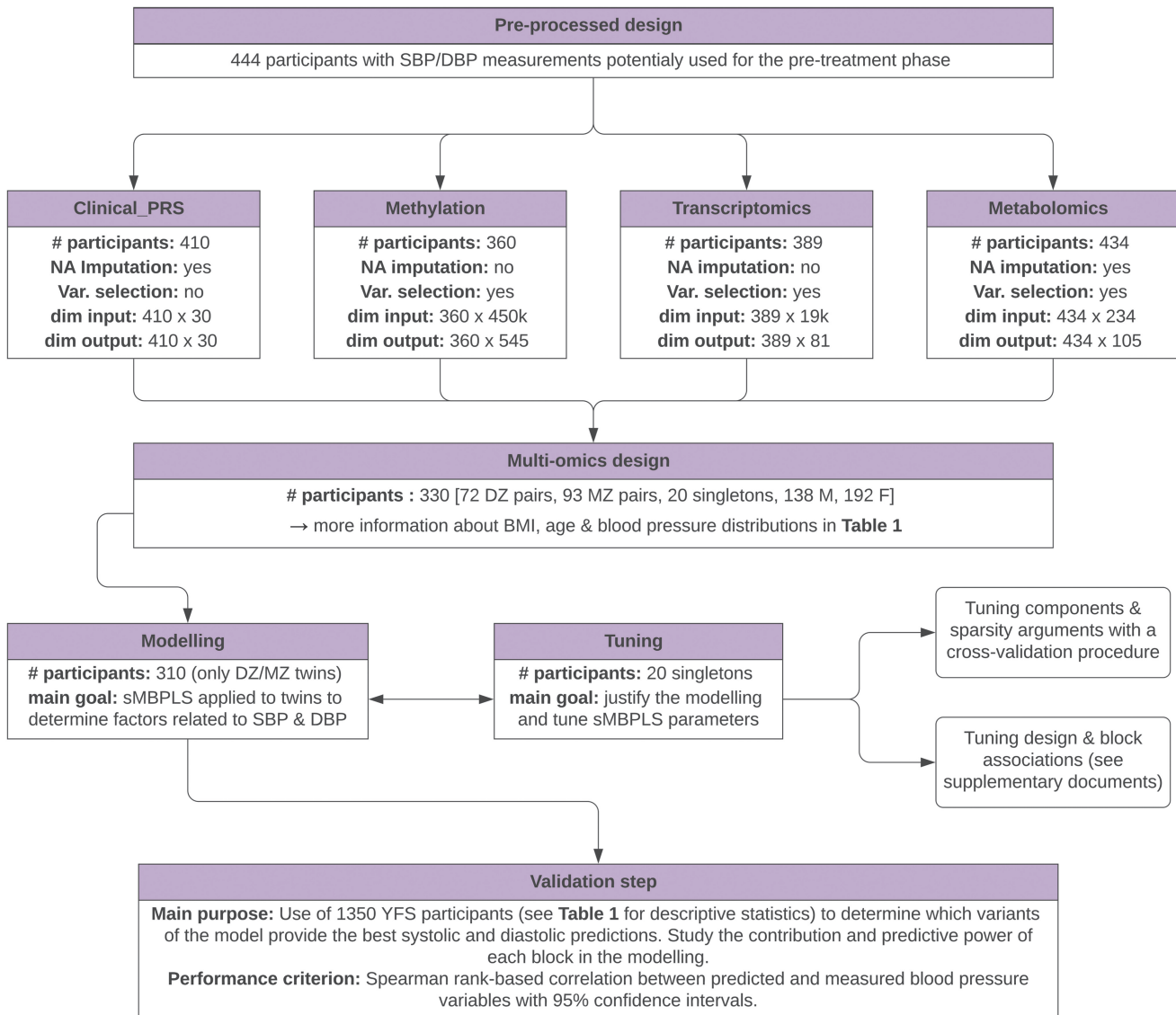
Pre-processed design

444 participants with SBP/DBP measurements potentialy used for the pre-treatment phase

Clinical_PRS	**Methylation**	**Transcriptomics**	**Metabolomics**
# participants: 410 **NA Imputation:** yes **Var. selection:** no **dim input:** 410 x 30 **dim output:** 410 x 30	**# participants:** 360 **NA imputation:** no **Var. selection:** yes **dim input:** 360 x 450k **dim output:** 360 x 545	**# participants:** 389 **NA imputation:** no **Var. selection:** yes **dim input:** 389 x 19k **dim output:** 389 x 81	**# participants:** 434 **NA imputation:** yes **Var. selection:** yes **dim input:** 434 x 234 **dim output:** 434 x 105

Multi-omics design

 # participants : 330 [72 DZ pairs, 93 MZ pairs, 20 singletons, 138 M, 192 F]

→ more information about BMI, age & blood pressure distributions in **Table 1**

Modelling

 # participants: 310 (only DZ/MZ twins)
main goal: sMBPLS applied to twins to determine factors related to SBP & DBP

Tuning

 # participants: 20 singletons
main goal: justify the modelling and tune sMBPLS parameters

Tuning components & sparsity arguments with a cross-validation procedure

Tuning design & block associations (see supplementary documents)

Validation step

Main purpose: Use of 1350 YFS participants (see **Table 1** for descriptive statistics) to determine which variants of the model provide the best systolic and diastolic predictions. Study the contribution and predictive power of each block in the modelling.
Performance criterion: Spearman rank-based correlation between predicted and measured blood pressure variables with 95% confidence intervals.

FIG. 1. Study design diagram. The study design is structured into three main phases: a preprocessing phase at the scale of each omic, a multi-omic modeling phase and a prediction phase. #, number; DBP, diastolic blood pressure; dim, dimension; DZ, dizygotic; F, female; M, male; MZ, monozygotic; NA, missing value; SBP, systolic blood pressure; Var., variable.

defined as weighted sums of the block-related components, and the latent components of the Y matrix.

This method therefore expresses Q omics block matrices $X_1,\ldots,X_Q$ as matrix products of block-related components by loading vectors ($Q=4$ in this study), and provides upscaled latent components used in our study to predict a two-dimensional Y matrix composed of the SBP and DBP variables.

The sMBPLS modeling was performed using methods implemented in the mixOmics R package (Rohart et al., 2017). In addition to the classical sMBPLS structure, the mixOmics package introduces a so-called design matrix, allowing for linking each omics block to influence the covariance maximization phase (Lê Cao and Welham, 2021). This $Q \times Q$ design matrix, commonly noted as C, associates an omics block to another omics block using a coefficient defined on the segment [0,1] (0=no link, 1=complete association). Because the choice of this matrix is based on

a priori and observational choices, we used all the participants who did not have their co-twin (Fig. 1) among the initial 330 to estimate this matrix, resulting in the selection of 20 participants, hereafter called singletons.

This exploratory approach allowed us to tune the design matrix (Supplementary Document, Section S3 and Supplementary Fig. S1 and Supplementary Table S1) by introducing a metric weighting the systolic and diastolic root mean square error (RMSE). Two nonzero omics block associations minimized this metric: a moderate association (0.4) between the Metabolomics and Clinical_PRS blocks as well as a weak association (0.1) between the methylation and transcriptomics blocks. The design matrix was therefore set accordingly. Each block X_i was also penalized with a penalty term λ_i that enables variable selection in each omics block. These $\lambda_i,\ldots,\lambda_Q$ ($Q=4$) constrain the number of variables within each block.

TABLE 1. DESCRIPTION OF THE FINNISH TWIN COHORT AND THE YOUNG FINNS STUDY PARTICIPANTS

Statistic	N	Mean	Standard deviation	Min	Pctl(25)	Pctl(75)	Max
Finnish twin cohort							
SBP	330	150.68	20.20	102.00	137.00	163.20	230.00
DBP	330	85.61	11.94	58.00	77.00	92.50	126.50
Sex	330	138M/192F					
Age	330	62.31	3.82	55.85	59.29	65.54	69.69
BMI	330	27.32	4.73	18.06	24.00	29.60	45.91
Alc	328	327.77	442.14	0.00	83.20	385.50	4928.70
Waist	330	94.65	14.39	60.00	85.00	103.00	140.00
Young Finns Study							
SBP	1350	119.21	14.31	83.00	109.30	127.30	179.00
DBP	1350	75.31	10.61	44.00	68.00	81.33	113.33
Sex	1350	733M/617F					
Age	1350	41.63	5.09	34.00	37.00	46.00	49.00
BMI	1347	26.66	5.06	16.49	23.23	29.26	58.47
Alc	1249	245.75	363.37	0.00	26.14	305.00	4357.14
Waist	1347	92.38	14.31	61.10	82.15	100.47	160.40

The distributions of BMI and waist circumference were similar between the two cohorts, but differences in alcohol consumption, age, SBP and DBP distributions were observed. Age in years.

Alc, alcohol consumption (g/month); BMI, body mass index in kg/m^2; DBP, diastolic blood pressure (mmHg); F, female; M, male; Pctl, percentile; SBP, systolic blood pressure (mmHg); Waist, waist circumference (cm).

To avoid defining sparsity arguments and the number of components (k) based on a biological *a priori*, we implemented a cross-validation procedure in a mixOmics framework to automatically select the best combination (k^*, $\lambda_1^*, \ldots, \lambda_Q^*$), minimizing a criterion called cross-validation score (CV score) (Supplementary Document, Section S4) (Li et al., 2012).

Cross-validation procedure. Links between sMBPLS and traditional methods such as principal component analysis (PCA) exist, insofar as PCA aims to summarize information from linear combinations of variables to project individuals into a reduced space built from components. Within the framework of PCA, some tools make it possible to establish an optimal number of components to be selected to optimize the explained variance wisely; one can note the use of elbow or Kaiser criteria as examples. In the sMBPLS framework, this selection is more subtle and no automatic mixOmics method exists when it comes to a quantitative Y matrix to be regressed: cross-validation is only available for the discriminant version of sMBPLS, called sMBPLS-DA. The main drawback of the sMBPLS-DA cross-validation procedure is the computational time cost, because the sparsity arguments applied to each block as well as the number of components k make rapid increase in the number of modeling combinations to be tested.

With the awareness of the potential computational shortcomings of this type of cross-validation procedure, we implemented a self-governed cross-validation tailored to sMBPLS (Li et al., 2012) in R using the features of mixOmics (Supplementary Document, Section S4). A total of $N = 310$ individuals were therefore distributed into $L = 10$ groups before training L models on $N - N/L$ individuals to derive the loadings and weight vectors. A CV score was calculated at each iteration, for each combination of sparsity arguments λ_i ($i = 1, \ldots, Q$) and number of components k. The best model combination minimizes the CV score.

Predictive methods

Data linkage. Although all blocks were overlapped in the YFS test cohort, the variables in each block were only subsets of the variables in the corresponding block in the FTC cohort. Of the clinical data, almost one-third of the variables were not retrieved in the YFS data. Lymphocytes, neutrophils, B neutrophils, B lymphocytes, and the two PRS variables for SBP and DBP were not available. The PRS for CAD risk and the PRS for BMI were obtained using a *p*-value threshold of 10^{-5} (Võsa et al., 2021). Only 5 of the 105 metabolomic variables were missing in the YFS data; the other 100 variables did not suffer from missing values.

YFS methylation data were obtained from Illumina EPIC, and the β-values were computed (Supplementary Document, Section S2). CpG site selection was carried out by name linkage with the FTC methylation data, leading to the selection of 463 methylation variables from the original 545. The selection of transcriptomic variables was more subtle, as several probes pointed to the same genes (*MYADM*, *CD97*). To match each probe obtained with FTC data and those available within the YFS data, a linkage by ProbeID was performed. A total of 66 YFS transcriptomic variables were thus retrieved, whereas there were 81 in the FTC data.

A consequence of missing variables and cohort heterogeneity may be a significant bias in predictions. The absence of a few clinical variables with strong predictive power should be avoided even if the mixOmics package allows predictions to be made from partially missing data. To reduce the discrepancies in predictions, a correction for batch effect using the Combat method (Leek et al., 2012) on transcriptomic and methylation data was carried out (Supplementary Document, Section S5 and Supplementary Fig. S2). This correction resulted in a reduction of the dimensions of the FTC transcriptomic and methylation datasets, as the batch correction imposes the same FTC and YFS variables. This operation was necessary as predictions without batch-effect correction proved unreliable because the prediction errors were particularly high.

Performance criteria. In addition to missing variable management, significant clinical heterogeneity between the two cohorts was observed and suspected to introduce prediction biases as illustrated by the age distribution of the two cohorts (Table 1). These cohort differences may bias an RMSE-type measure as the weight given to age in the modeling based on FTC participants is likely to be underestimated when using the YFS test cohort. For all these reasons, a rank-based Spearman correlation ρ was preferred as a performance measure. Besides the correlation coefficients, 95% confidence intervals were calculated as implemented in the DescTools R package (Signorell et al., 2021).

This performance measure was used both to estimate the correlation between predicted and observed blood pressure values in the YFS and FTC cohorts as well as to gauge the correlation between variables and the phenotypic traits of interest (SBP and DBP). Correlation nullity tests were also undertaken using R base implemented functions.

Results

Parameter estimation and cross-validation

Under the optimal design matrix outlined in the Materials and Methods section, the number of components was set to $k = 1$ pursuant to the CV score values ($k = 1$: pooled CV score = 166,198, standard deviation [SD] = 386; $k = 2$: pooled CV score = 309,956, SD = 1082; $k = 3$: pooled CV score = 348,222, SD = 26,422). A final cross-validation procedure was performed to tune the sparsity arguments related to the Clinical_PRS and Metabolomics blocks because variable selection was already performed on the transcriptomic and methylation data (Supplementary Document, Section S1).

The CV score over 20 iterations by testing different sparsity value ranges (2×2 for the Clinical_PRS block and 4×4 for the Metabolomic block simultaneously) revealed that a nonsparse model produces the lowest CV score. This result can be explained by the fact that the weights of the Clinical_PRS and Metabolomic blocks were found to be consistent in both the integrative and predictive phases of our study. The definition of the CV score (Li et al., 2012) thus likely offered a significant weight to the variables of these two blocks in the creation of the CV score, strongly penalizing the removal of one of them.

When tuning sparsity arguments in the methylation and transcriptomic blocks, differences in CV score as a function of sparsity restriction were heterogeneous. These differences were weak for the methylation block: the CV score with all 466 methylation variables remained within 1 SD of the CV score with 100 methylation variables. In the transcriptomic block, the CV score was more sensitive to changes in sparsity: a nonexistent sparsity argument significantly minimized the CV score. In addition to showing difficulties in association with other blocks (Supplementary Document, Section S3), the cross-validation procedure pointed to the low weight of CpG sites in minimizing the CV score criterion.

Uneven predictive gains across omics blocks

To estimate the predictive contribution of each omics block within the modeling ($k = 1$; no sparsity arguments), systolic and diastolic data from the 1350 participants in the YFS cohort were predicted from block permutations.

TABLE 2. PREDICTIVE PERFORMANCE EXPRESSED AS SPEARMAN CORRELATION COEFFICIENTS BY PERMUTING OMICS BLOCKS IN THE MODEL

Model permutation	Blood pressure	ρ	95% CI
C	SBP	0.377	[0.330 to 0.422]
Me	SBP	0.051	[−0.002 to 0.104]
T	SBP	0.176	[0.124 to 0.227]
Mb	SBP	0.332	[0.284 to 0.379]
C+Me+T+Mb	SBP	0.359	[0.312 to 0.404]
C+T+Mb	SBP	0.436	[0.391 to 0.478]
C	DBP	0.448	[0.405 to 0.490]
Me	DBP	0.045	[−0.009 to 0.098]
T	DBP	0.147	[0.094 to 0.198]
Mb	DBP	0.392	[0.345 to 0.436]
C+Me+T+Mb	DBP	0.393	[0.347 to 0.438]
C+T+Mb	DBP	0.487	[0.446 to 0.527]

The three-block model achieved the best predictive performance for both SBP and DBP, highlighting the failure to integrate methylation data for which the Spearman correlation between blood pressure measurements and blood pressure predictions was not significantly non-null at the 5% threshold in a single-block context.

CI, confidence interval of ρ; C, Clinical_PRS; Mb, metabolomics; Me, methylation; PRS, polygenic risk scores; T, transcriptomics.

Spearman correlation coefficients were calculated, as described in the Materials and Methods section, to estimate the correlation between predicted and measured blood pressure values (Table 2). The performance of six models was studied, including the original four-block model (noted as C+Me+T+Mb hereafter). A three-block model excluding the methylation block (C+T+Mb) was also studied, for which only the Clinical_PRS/Metabolomics association of the design matrix was preserved. In addition to these two permuted models, four submodels corresponding to four single-block PLS regressions, that is, simple PLS regressions, were used to highlight the predictive power of each isolated block.

The omics blocks had heterogeneous predictive power (Table 2). We reported a Spearman correlation close to 5% for the methylation data, for both SBP and DBP, in a single-omic setting. The 95% confidence intervals also contained the value 0 by a small margin in both SBP and DBP; methylation data struggled to provide good predictions (Spearman correlation nullity test, p-value >5% for DBP and SBP). Integration of methylation data in the four-block modeling was also deemed to be deleterious, insofar as the Spearman coefficient ρ was 9.4% lower in the case of DBP (compared with 7.7% in the case of SBP). Once the methylation block was removed from the four-block model, the three-block model obtained the best predictive performance, with a ρ close to 50% for DBP.

Although the differences in predictive performance between the three-block and single-omics models appear to be slight, biological and technical limitations prevent particularly high correlation coefficients from being obtained and strong statistical differences from being shown. Cohort differences (age and blood pressure distributions in particular) and missing clinical predictors illustrate these limitations. Integrating multiple blocks also averages each block-related latent variable into a single latent variable, thus explaining the difficulty of significantly improving predictions although the modeling has been enriched. These block-

related components also showed consistent predictive powers compared with those obtained in single-omics predictive phases (Table 2), while embedded in a multi-omics model.

Indeed, the distributions of each of these block-related components of the first and last decile of DBP, that is, the 10% of participants with the lowest (compared with highest) DBP in each of the two cohorts, show a slight replication defect of the transcriptomic data (Fig. 2). Similar to the weaker predictions reported for the transcriptomics block in single-omic settings ($\rho = 17.6\%$ for SBP, $\rho = 14.7\%$ for DBP; Table 2) compared with those measured for the metabolomics and clinical data, we observed a greater weakness of the transcriptomic block in distinguishing the first and last DBP decile of the YFS cohort in a multi-omics framework. Projections of the first and last DBP decile of the YFS test cohort onto the Metabolomic and Clinical_PRS block-related components have been more convincing in that their distribution is markedly different along the component (Fig. 2).

Global view of the modeling

To better understand the biological relevance of a multi-omics approach in the study of blood pressure values, the loading vectors of the three-block model (C+T+Mb) were derived. These have the function, as in the case of a PCA, of showing which variables contribute most to the creation of the sMBPLS block-related components. The log p-values obtained by testing the nullity of the Spearman correlation between each transcriptomic variable and SBP or DBP corrected for age, sex, and BMI in the YFS test cohort were compared with the loading factors of these transcriptomic variables in the modeling (Fig. 3). Genes contributing little to the creation of the transcriptomic-related component, that is, having a loading factor close to 0, struggled to be replicated within the YFS cohort, whereas the key replicated genes identified in the variable screening step (Supplementary Document, Section S1) had a major role in the modeling.

The transcriptomic values of the replicated *TPPP3* and *MYADM* genes (Huan et al., 2015; Zeller et al., 2017) were significantly correlated with the corrected values of SBP and DBP in the YFS cohort, as these two genes remained significantly associated even after Bonferroni correction. High loading-factor genes *TIPARP* and *SLC31A2*, replicated in the hypertension and blood pressure literature (Huan et al., 2015; Zeller et al., 2017), remained significant after Bonferroni correction for SBP, but not for DBP. Other genes with low correlation null test p-values close to 10^{-5} like *CD97*, *LMNA*, *F12*, and *AFAP1* were also found to be well represented in the hypertension literature (Kraja et al., 2017; Zeller et al., 2017). Thus, the modeling gave significant weight in the creation of the transcriptomic latent variable to genes replicated in both the YFS cohort and the hypertension literature, bridging the gap between the hypertension literature and our study dealing with unitary increases in SBP and DBP.

BMI and waist and hip circumference had particularly high loading factors (Table 3) reinforcing the clinical value of performing such measurements for predictive purposes. In addition to classical clinical variables such as lymphocyte or leukocyte counts, metabolomic variables were found to be related to BMI (e.g., branched chain amino acids [BCAAs] such as leucine and isoleucine) (Felig et al., 1969; Pietiläinen

et al., 2008) and blood lipid levels. The association between BCAAs and blood pressure was also driving the modeling, extrapolating the known link between BCAAs and hypertension (Mahbub et al., 2020) to the study of blood pressure values. Although valine, 1 of the 3 BCAAs, played a minor role in the modeling, it was found to be highly correlated with the variables leucine and isoleucine for which a Pearson correlation of >70% in both cases was measured in the 310 FTC participants included in the modeling.

Discussion

The integration of multiple datasets in multi-omics frameworks has become, in recent years, one of the leading methods to both compile knowledge in a domain and discover highly complex relationships between omics (Olivier et al., 2019). We conducted this study to extend the use of such integrative approaches in the study of blood pressure values. Metabolomic, clinical, and transcriptomic risk factors highlighted in the blood pressure modeling were widely replicated in the hypertension literature at the single omic level, proving the robustness of our approach to recover results usually obtained in single-omics and discriminative approaches.

In particular, the *CD97*, *MYADM*, *TIPARP*, *SLC31A2*, and *TPPP3* genes strongly contributed in creating the transcriptomic latent variable. Their significant contribution corroborated the previous results in hypertension and blood pressure settings (Huan et al., 2015; Huang et al., 2018; Zeller et al., 2017) while also showing that the connection between blood pressure and hypertension remains tight when studying the transcriptome.

Metabolomic and clinical factors replicated in the hypertension literature have been highlighted as playing a key role in understanding blood pressure, such as BCAAs (Mahbub et al., 2020) and obesity-related measures (Tanaka, 2020) while spotlighting the link connecting BCAAs to obesity measures in the study of blood pressure values. The multi-omics approach thus allowed overlapping with replicated results in the hypertension and blood pressure literature, while providing new multi-omics insights and readout in understanding the biological mechanisms underlying blood pressure unit variations.

The findings of our study go beyond novel biological contributions: they are part of a clinical and public health context and perspective. An in-depth understanding of the blood pressure-related mechanisms is of definite clinical and public health importance. Numerous studies have focused on blood pressure fluctuations in longitudinal frameworks, showing associations between high blood pressure variability over time and increased risks of cardiovascular or coronary heart diseases (Parati et al., 2018; Stevens et al., 2016). In addition, it is recently known that some diseases, such as cardiovascular disease, are associated with linear or nonlinear increases in blood pressure (Arvanitis et al., 2021; Wan et al., 2021), demonstrating the value of the present multi-omics integrative findings in considering blood pressure in its continuous, nondiscriminatory form.

The predictive contribution of each omic block on the test cohort showed a strong predictive potential, especially for clinical and metabolomic data. The best predictions were obtained with a three-block model discarding the methylation data, although a slight defect in replication of the
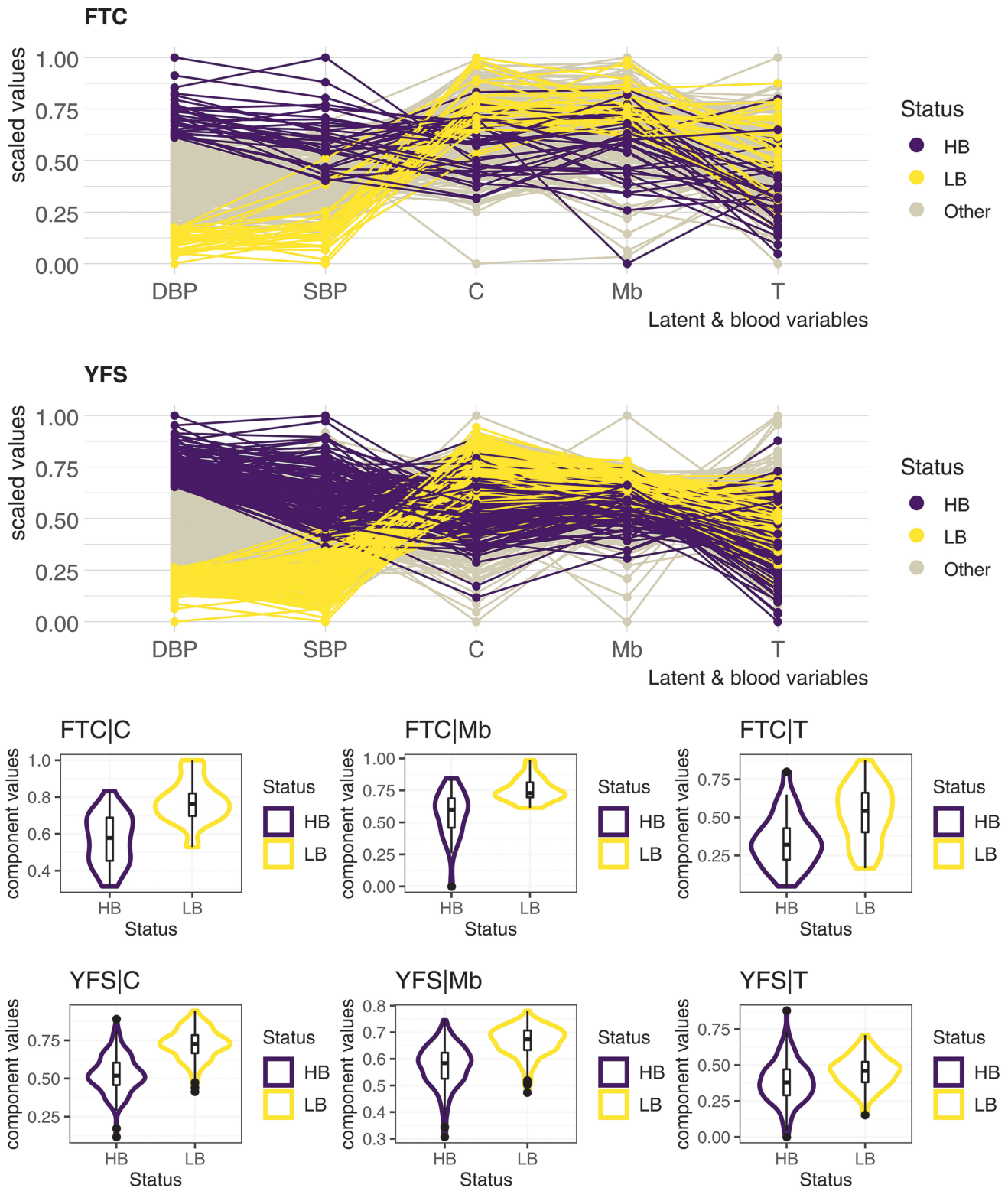
FIG. 2. Projection of participants of both cohorts on each block-related component. Despite strong differences in the distribution of diastolic (and systolic) blood pressure between the two cohorts (Table 1 and Supplementary Document, Section S4), the three-block model distributed the first and last decile participants fairly distinctly over its block-related components. The transcriptomic component, however, lost some of its strength in that the distributions of the first and last decile on the YFS cohort are considerably closer. Blood measures and block-related components were scaled in each of the two cohorts to obtain this figure. C, clinical_PRS; HB, last decile; LB, first decile; Mb, metabolomics; PRS, polygenic risk scores; T, transcriptomics; YFS, Young Finns Study.
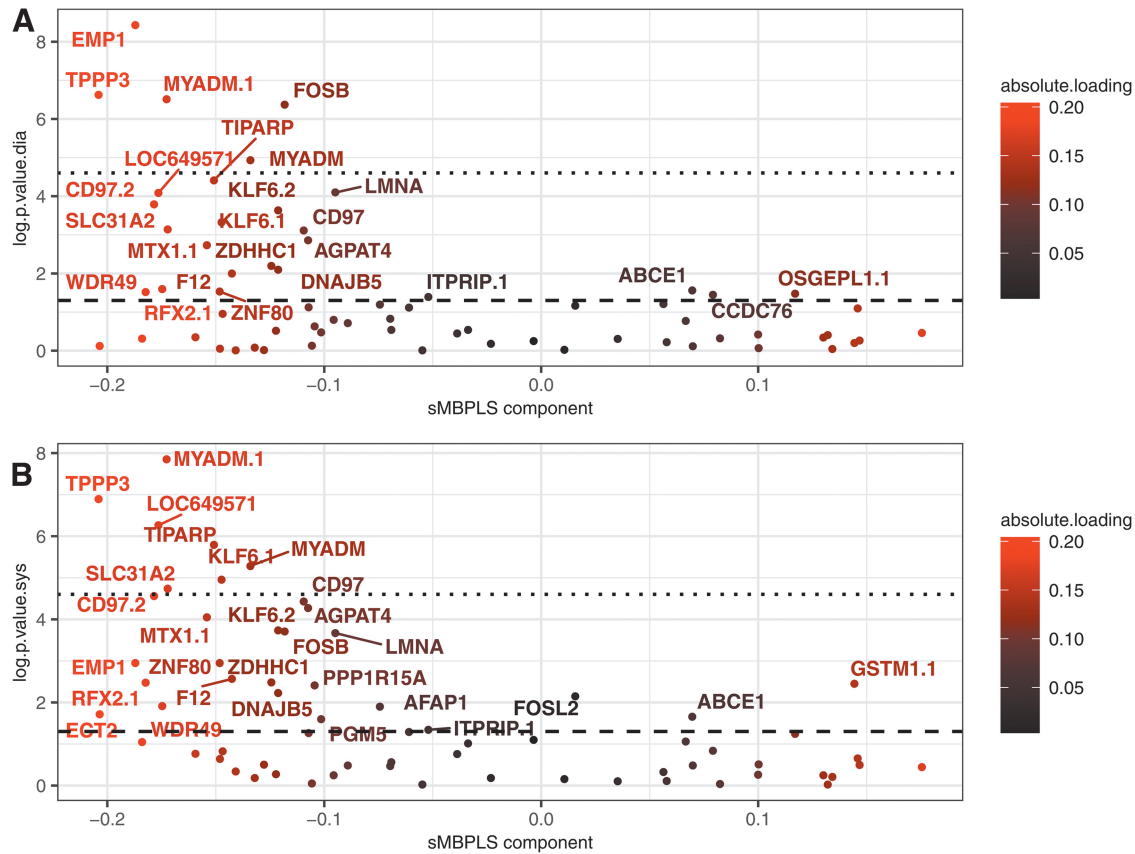
FIG. 3. Transcriptomic loadings compared with *p*-values in Spearman's correlation nullity test in corrected SBP and DBP applied on YFS participants. Genes contributing the most to the creation of the transcriptomic component, that is, having high loading factors in absolute value, tended to have low Spearman's correlation nullity test *p*-values compared with SBP and DBP controlled by age, sex, and BMI. The axis log.*p*.value.sys on plot (**B**) (resp. log.*p*.value.dia on plot (**A**)) refers, respectively, to the negative logarithm to base 10 of the *p*-value obtained in the Spearman correlation nullity test between each transcriptomic variable and systolic (resp. diastolic) blood pressure controlled by age, sex, and BMI in the YFS test cohort. The absolute.loading coloring refers to the absolute factor loading value of each gene in the modeling. The semi-full line refers to the negative logarithm to base 10 of the 5% *p*-value threshold while the dashed line refers to the Bonferroni threshold. BMI, body mass index.

TABLE 3. TEN CLINICAL AND METABOLOMIC VARIABLES WITH THE HIGHEST ABSOLUTE LOADING FACTORS

Variable name	Biological meaning	Block	Loading
Waist	Waist circumference	Clinical_PRS	−0.420
BMI	BMI (kg/m^2)	Clinical_PRS	−0.389
HIP	Hip circumference	Clinical_PRS	−0.306
FB LEUK	Leucocytes	Clinical_PRS	−0.304
B MONOS	Monocytes	Clinical_PRS	−0.271
B NEUT	Neutrophils	Clinical_PRS	−0.250
B HB	Hemoglobin	Clinical_PRS	−0.250
SEX	Sex (M/F)	Clinical_PRS	0.241
B LYMF	Lymphocytes	Clinical_PRS	−0.229
B HKR	Haematocrit	Clinical_PRS	−0.229
Gp	Glycoprotein acetylation	Metabolomics	−0.197
Ile	Isoleucine	Metabolomics	−0.194
Leu	Leucine	Metabolomics	−0.188
LHDLFC	Free cholesterol in large HDL	Metabolomics	0.184
XLHDLP	c^o of very large HDL particles	Metabolomics	0.179
TGPG	r^o triglycerides/phosphoglycerides	Metabolomics	−0.169
PUFAFA	r^o polyunsaturated f.a/total f.a	Metabolomics	0.168
LHDLPL	Phospholipids/total lipids r^o (LHDL)	Metabolomics	0.161
IDLFC	Free cholesterol/total lipids r^o (IDL)	Metabolomics	0.160
LHDLPL	Phospholipids in LHDL	Metabolomics	−0.157

c^o, concentration; f.a, fatty acid; LHDL, large high-density lipoprotein; r^o, ratio.

transcriptomic block in the test cohort was observed. This three-block model was able to order participants according to their SBP and DBP in the test cohort, despite particularly different SBP and DBP distributions between the training and test cohorts (Supplementary Document, Section S6 and Supplementary Fig. S3 and Table 1). The rejection of methylation data in the modeling was motivated by its deleterious role in acquiring good predictions. The preselection of CpG sites by elastic-net (Supplementary Document, Section S1) could be one of the sources of this integration failure as there was a lack of statistical power.

The study of blood pressure values in its quantitative form could also play an important role in this failure as studying unit increases in SBP and DBP is probably too ambitious in light of the sample size. However, these may not be the only reasons for this failure and beyond the purely technical aspect, it is the predictive robustness of the methylation data that seems to be problematic when using an external replication cohort. An additional study (Supplementary Document, Section S7 and Supplementary Table S2) using a different methylation preprocessing method (van Dongen et al., 2021) and considering a selection of replicated CpG sites (Richard et al., 2017) in the modeling showed that the predictive power of the methylation block remained particularly low.

Thus, the choices made in our study do not seem to be the major cause of this integration failure. Because the epigenome is strongly sensitive to age and a large number of confounders such as smoking (Bollepalli et al., 2019; Martin and Fry, 2018), the difficulty in obtaining satisfactory quality predictions may mainly be explained by differences between training and test cohorts as well as a lack of finesse in controlling for blood variables. The use of methylation data for predictive purposes is therefore challenging in the context of blood pressure and would require further studies. The use of multi-omics methods for nonpredictive exploratory purposes could, however, be relevant and has already been demonstrated in a wide variety of contexts (Kolenc et al., 2021).

The achievement of better predictions of blood pressure values is also conditioned on other factors. The democratized use of deep learning (DL) methods to predict complex phenotypes (Cao et al., 2018) could also be suitable for the study of blood pressure values: the high volumes of blood pressure-related data and the growing knowledge in the field could allow the acquisition of excellent quality predictions. As the black box effect is difficult to counter with DL methods, the use of the sMBPLS method is all the more justified to derive biological and clinical interpretations easily. However, the sMBPLS method still needs to be used more extensively to understand its full value, as has already been carried out with discriminative versions of latent-based methods (Singh et al., 2019).

Recent work tends to gain interpretability with DL methods by forming connections with traditional PLS methods, such as in the context of metabolomic data (Mendez et al., 2020): increased methodological developments should, in the coming years, make it possible to reconcile interpretability and predictive performance. Adding data to feed the modeling could also easily improve these predictions, in addition to uncovering important biological mechanisms. Proteomics could fulfill both these tasks as some blood pressure-related proteomic species are already identified (Arnett and Claas,

2018; Carty et al., 2013) and their predictive potential in a discriminatory context has already been demonstrated (Gajjala et al., 2017). Associations between proteomics and other omics such as transcriptomic data are also common (Kolenc et al., 2021), making their use in the study of blood pressure-related phenotypes encouraging. Other omics could also be suitable for multi-omics integration, but more exploratory studies need to be conducted for this purpose.

Complementary approaches can also significantly improve the quality of modeling and predictions, such as multi-omics imputation methods. Although multiple imputation has been used judiciously to impute a reasonable proportion of missing clinical and metabolomic values (Supplementary Document, Section S1), the use of new emerging methods specifically designed for multi-omics contexts may allow for easier imputation with at least as good quality (Song et al., 2020). The increasing use of multi-omics approaches therefore induces the development of auxiliary methods making its use easier, more efficient, and more relevant. The massive use of multi-omics approaches in the understanding of complex phenotypes can only be encouraged because, in addition to its biological and predictive interest, it contributes to the methodological expansion of the multi-omics field.

Data Availability

The YFS dataset comprises health-related participant data and their use is therefore restricted under the regulations on professional secrecy (Act on the Openness of Government Activities, 612/1999) and on sensitive personal data (Personal Data Act, 523/1999, implementing the EU data protection directive 95/46/EC). Owing to these legal restrictions, the Ethics Committee of the Hospital District of Southwest Finland has in 2016 stated that individual-level data cannot be stored in public repositories or otherwise made publicly available. Data sharing outside the group is carried out in collaboration with the YFS group and requires a data-sharing agreement with the understanding that collaborators will protect the data and not share it with any other parties.

The list of all investigators that collaborate with the YFS group is displayed at the website of the YFS (http:// youngfinnsstudy.utu.fi/). Investigators can submit an expression of interest to the chairperson of the data sharing and publication committee, professor Mika Kähönen (Tampere University) and for genomics information to professor Terho Lehtimäki (Tampere University).

The Finnish Twin Cohort data used in the analysis is deposited in the Biobank of the Finnish Institute for Health and Welfare (https://thl.fi/en/web/thl-biobank/for-researchers). It is available to researchers after written application and following the relevant Finnish legislation.

Authors' Contributions

G.D. conducted this study and performed the analyses. J.K. supervised this work. G.D. wrote the first draft of the article with editing assistance from J.K., O.M., J.M., and P.M. The revision of the article was carried out by G.D., J.K., M.O. and J.M. M.O., O.R., T.L., M.K., X.W., and J.K. collected the data used in this article. J.M. handled the transfer and preparation of the YFS data. All authors had a substantial role in the completion of this study. All authors read and approved the final version of the article.

Acknowledgments

Author Disclosure Statement

The authors declare they have no conflicting financial interests.

Funding Information

Supplementary Material

Supplementary Data
Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Table S1
Supplementary Table S2

References

Abayomi K, Gelman A, and Levy M. (2005). Diagnostics for multivariate imputations. J R Stat Soc Ser C Appl Stat 57, 273–291.

Abdi H, and Williams LJ. (2013). Partial least squares methods: Partial least squares correlation and partial least square regression. Methods Mol Biol 930, 549–579.

Ahola-Olli AV, Mustelin L, Kalimeri M, et al. (2019). Circulating metabolites and the risk of type 2 diabetes: A prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia 62, 2298–2309.

Arnett DK, and Claas SA. (2018). Omics of blood pressure and hypertension. Circ Res 122, 1409–1419.

Arvanitis M, Qi G, Bhatt DL, et al. (2021). Linear and nonlinear Mendelian randomization analyses of the association between diastolic blood pressure and cardiovascular events: The J-curve revisited. Circulation 143, 895–906.

Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. (2014). Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30, 1363–1369.

Baek S, Jang J, Cho SH, Choi JM, and Yoon S. (2020). Blood pressure prediction by a smartphone sensor using fully convolutional networks. Annu Int Conf IEEE Eng Med Biol Soc 2020, 188–191.

Benton MC, Sutherland HG, Macartney-Coxson D, Haupt LM, Lea RA, and Griffiths LR. (2017). Methylome-wide association study of whole blood DNA in the Norfolk Island isolate identifies robust loci associated with age. Aging (Albany NY) 9, 753–768.

Boks MP, Derks EM, Weisenberger DJ, et al. (2009). The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. PLoS One 4, e6767.

Bollepalli S, Korhonen T, Kaprio J, Anders S, and Ollikainen M. (2019). EpiSmokEr: A robust classifier to determine smoking status from DNA methylation data. Epigenomics 11, 1469–1486.

Cao C, Liu F, Tan H, et al. (2018). Deep learning and its applications in biomedicine. Genom Proteom Bioinform 16, 17–32.

Carty DM, Schiffer E, and Delles C. (2013). Proteomics in hypertension. J Hum Hypertens 27, 211–216.

Cazaly E, Thomson R, Marthick JR, Holloway AF, Charlesworth J, and Dickinson JL. (2016). Comparison of pre-processing methodologies for Illumina 450k methylation array data in familial analyses. Clin Epigenetics 8, 75.

Domingo-Relloso A, Huan T, Haack K, et al. (2021). DNA methylation and cancer incidence: Lymphatic-hematopoietic versus solid cancers in the Strong Heart Study. Clin Epigenetics 13, 43.

Du P, Kibbe W, and Lin S. (2008). Lumi: A pipeline for processing Illumina microarray. Bioinformatics 24, 1547–1548.

Elovainio M, Taipale T, Seppälä I, et al. (2015). Activated immune-inflammatory pathways are associated with long-standing depressive symptoms: Evidence from gene-set enrichment analyses in the Young Finns Study. J Psychiatr Res 71, 120–125.

Felig P, Marliss E, and Cahill GF Jr. (1969). Plasma amino acid levels and insulin secretion in obesity. N Engl J Med 281, 811–816.

Friedman J, Hastie T, and Tibshirani R. (2010). Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33, 1–22.

Gajjala PR, Jankowski V, Heinze G, et al. (2017). Proteomic-Biostatistic integrated approach for finding the underlying molecular determinants of hypertension in human plasma. Hypertension 70, 412–419.

Hayati Rezvan P, Lee KJ, and Simpson JA. (2015). The rise of multiple imputation: A review of the reporting and implementation of the method in medical research. BMC Med Res Methodol 15, 30.

Honaker J, King G, and Blackwell M. (2011). Amelia II: A program for missing data. J Stat Softw 45, 1–47.

Huan T, Esko T, Peters MJ, et al. (2015). A meta-analysis of gene expression signatures of blood pressure and hypertension. PLoS Genet 11, e1005035.

Huang Y, Ollikainen M, Muniandy M, et al. (2020). Identification, heritability, and relation with gene expression of novel DNA methylation loci for blood pressure. Hypertension 76, 195–205.

Huang Y, Ollikainen M, Sipilä P, et al. (2018). Genetic and environmental effects on gene expression signatures of blood pressure: A transcriptome-wide twin study. Hypertension 71, 457–464.

Irvin MR, Jones AC, Claas SA, and Arnett DK. (2021). DNA methylation and blood pressure phenotypes: A review of the literature. Am J Hypertens 34, 267–273.

Jordan J, Kurschat C, and Reuter H. (2018). Arterial hypertension. Dtsch Arztebl Int 115, 557–568.

Kaprio J, Bollepalli S, Buchwald J, et al. (2019). The older Finnish Twin Cohort—45 years of follow-up. Twin Res Hum Genet 22, 240–254.

Kaprio J, Koskenvuo M, Langinvainio H, Romanov K, Sarna S, and Rose RJ. (1987). Genetic influences on use and abuse of alcohol: A study of 5638 adult Finnish twin brothers. Alcohol Clin Exp Res 11, 349–356.

Keil U, Chambless L, Filipiak B, and Härtel U. (1991). Alcohol and blood pressure and its interaction with smoking and other behavioural variables: Results from the MONICA Augsburg Survey 1984–1985. J Hypertens 9, 491–498.

Kelly DM, and Rothwell PM. (2020). Blood pressure and the brain: The neurology of hypertension. Pract Neurol 20, 100–108.

Kraja AT, Cook JP, Warren HR, et al. (2017). New blood pressure-associated loci identified in meta-analyses of 475 000 individuals. Circ Cardiovasc Genet 10, e001778.

Kolenc Ž, Pirih N, Gretic P, and Kunej T. (2021). Top trends in multiomics research: Evaluation of 52 published studies and new ways of thinking terminology and visual displays. OMICS 25, 681–692.

Ku E, Lee BJ, Wei J, and Weir MR. (2019). Hypertension in CKD: Core curriculum 2019. Am J Kidney Dis 74, 120–131.

Kwong EW, Wu H, and Pang GK. (2018). A prediction model of blood pressure for telemedicine. Health Informatics J 24, 227–244.

Lê Cao KA, and Welham Z. (2021). *Multivariate Data Integration Using R: Methods and Applications with the mixOmics Package,* 1st ed. Chapman and Hall/CRC, London, United Kingdom.

Leek JT, Johnson WE, Parker HS, Jaffe AE, and Storey JD. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28, 882–883.

Li W, Zhang S, Liu CC, and Zhou X. (2012). Identifying multi-layer gene regulatory modules from multi-dimensional genomic data. Bioinformatics 28, 2458–2466.

Lin SM, Du P, Huber W, and Kibbe WA. (2008). Model-based variance-stabilizing transformation for Illumina microarray data. Nucleic Acids Res 36, e11.

Mahbub MH, Yamaguchi N, Hase R, et al. (2020). Plasma branched-chain and aromatic amino acids in relation to hypertension. Nutrients 12, 3791.

Martin EM, and Fry RC. (2018). Environmental influences on the epigenome: Exposure-associated DNA methylation in human populations. Annu Rev Public Health 39, 309–333.

McCartney DL, Min JL, Richmond RC, et al. (2021). Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging. Genome Biol 22, 194.

Mendez KM, Broadhurst DI, and Reinke SN. (2020). Migrating from partial least squares discriminant analysis to artificial neural networks: A comparison of functionally equivalent visualisation and feature contribution tools using jupyter notebooks. Metabolomics 16, 17.

Nikpay M, Goel A, Won HH, et al. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 47, 1121–1130.

Olivier M, Asmis R, Hawkins GA, Howard TD, and Cox LA. (2019). The need for multi-omics biomarker signatures in precision medicine. Int J Mol Sci 20, 4781.

Parati G, Stergiou GS, Dolan E, and Bilo G. (2018). Blood pressure variability: Clinical relevance and application. J Clin Hypertens (Greenwich) 20, 1133–1137.

Pietiläinen KH, Naukkarinen J, Rissanen A, et al. (2008). Global transcript profiles of fat in monozygotic twins discordant for BMI: Pathways behind acquired obesity. PLoS Med 5, e51.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, and Reich D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904–909.

Puddey IB, Mori TA, Barden AE, and Beilin LJ. (2019). Alcohol and hypertension-new insights and lingering controversies. Curr Hypertens Rep 21, 79.

Raitakari OT, Juonala M, Rönnemaa T, et al. (2008). Cohort profile: The cardiovascular risk in Young Finns Study. Int J Epidemiol 37, 1220–1226.

Richard MA, Huan T, Ligthart S, et al. (2017). DNA methylation analysis identifies loci for blood pressure regulation. Am J Hum Genet 101, 888–902.

Rohart F, Gautier B, Singh A, and Lê Cao KA. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol 13, e1005752.

Salvador MR, Cunha Gonçalves S, Quinaz Romana G, et al. (2019). Effect of lifestyle on blood pressure in patients under antihypertensive medication: An analysis from the Portuguese Health Examination Survey. Rev Port Cardiol (Engl Ed) 38, 697–705.

Schwingshackl L, Schwedhelm C, Hoffmann G, et al. (2017). Food groups and risk of hypertension: A systematic review and dose-response meta-analysis of prospective studies. Adv Nutr 8, 793–803.

Signorell A, Aho K, Alfons A, et al. (2021). DescTools: Tools for Descriptive Statistics. R package version 0.99.43. https://cran.r-project.org/package=DescTools Last viewed on October 29, 2021.

Singh A, Shannon CP, Gautier B, et al. (2019). DIABLO: An integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics 35, 3055–3062.

Soininen P, Kangas AJ, Würtz P, Suna T, and Ala-Korpela M. (2015). Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. Circ Cardiovasc Genet 8, 192–206.

Song M, Greenbaum J, Luttrell J 4th, et al. (2020). A review of integrative imputation for multi-omics datasets. Front Genet 11, 570255.

Stevens SL, Wood S, Koshiaris C, et al. (2016). Blood pressure variability and cardiovascular disease: Systematic review and meta-analysis. BMJ 354, i4098.

Surendran P, Feofanova EV, Lahrouchi N, et al. (2020). Discovery of rare variants associated with blood pressure regulation through meta-analysis of 1.3 million individuals. Nat Genet 52, 1314–1332.

Tanaka M. (2020). Improving obesity and blood pressure. Hypertens Res 43, 79–89.

Triche TJ Jr., Weisenberger DJ, Van Den Berg D, Laird PW, and Siegmund KD. (2013). Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res 41, e90.

Tuomela J, Kaprio J, Sipilä PN, et al. (2019). Accuracy of self-reported anthropometric measures—Findings from the Finnish Twin Study. Obes Res Clin Pract 13, 522–528.

van Buuren S, and Groothuis-Oudshoorn K. (2011). mice: Multivariate imputation by chained equations in R. J Stat Softw 45, 1–67.

van Dongen J, Gordon SD, McRae AF, et al. (2021). Identical twins carry a persistent epigenetic signature of early genome programming. Nat Commun 12, 5618.

Vilaplana JM. (2006). Blood pressure measurement. J Ren Care 32, 210–213.

Võsa U, Claringbould A, Westra HJ, et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet 53, 1300–1310.

Waldmann P, Mészáros G, Gredler B, Fuerst C, and Sölkner J. (2013). Evaluation of the lasso and the elastic net in genome-wide association studies. Front Genet 4, 270.

Wan EYF, Fung WT, Schooling CM, et al. (2021). Blood pressure and risk of cardiovascular disease in UK Biobank: A Mendelian randomization study. Hypertension 77, 367–375.

Wang Q, Xu Y, Zeng G, and Sun M. (2018). Continuous blood pressure estimation based on two-domain fusion model. Comput Math Methods Med 2018, 1981627.

Yengo L, Sidorenko J, Kemper KE, et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Hum Mol Genet 27, 3641–3649.

Zeller T, Schurmann C, Schramm K, et al. (2017). Transcriptome-wide analysis identifies novel associations with blood pressure. Hypertension 70, 743–750.

Zou H, and Hastie T. (2005). Regularization and variable selection via the elastic nets. J Royal Stat Soc B 67, 301–320.

Address correspondence to:
Gabin Drouard, MSc
Institute for Molecular Medicine Finland (FIMM)
HiLIFE
University of Helsinki
Tukholmankatu 8
Helsinki 00014
Finland

E-mail: gabin.drouard@helsinki.fi

Abbreviations Used

BCAA = branched chain amino acid
BMI = body mass index
CAD = coronary artery disease
c^o = concentration
CI = confidence interval
CV score = cross-validation score
DBP = diastolic blood pressure
dim = dimension
DL = deep learning
DZ = dizygotic
F = female
f.a = fatty acid
FTC = Finnish Twin Cohort
LHDL = large high-density lipoprotein
M = male
Mb = metabolomics
Me = methylation
MZ = monozygotic
NA = missing value
PCA = principal component analysis
Pctl = percentile
PLS = partial least square
PRS = polygenic risk scores
RMSE = root mean square error
r^o = ratio
SBP = systolic blood pressure
SD = standard deviation
sMBPLS = sparse multi-block partial least square
T = Transcriptomics
Var. = variable
YFS = Young Finns Study