Joint Metabonomic and Instrumental Analysis for the Classification of Migraine Patients with 677-MTHFR Mutations

Pierangela Giustetto[1], William Liboni[1], Ornella Mana[1], Gianni Allais[2], Chiara Benedetto[2] and Filippo Molinari[*,3]

[1]*Department of Neuroscience, Gradenigo Hospital, Torino, Italy*

[2]*Women's Headache Center, Department of Gynecology and Obstetrics, University of Torino, Torino, Italy*

[3]*Biolab, Department of Electronics, Politecnico di Torino, Torino, Italy*

Abstract: Migraine is a neurological disorder that correlates with an increased risk of cerebrovascular lesions. Genetic mutations of the MTHFR gene are correlated to migraine and to the increased risk of artery pathologies. Also, migraine patients show altered hematochemical parameters, linked to an impaired platelet aggregation mechanism. Hence, the vascular assessment of migraineurs is of primary importance.

Transcranial Doppler sonography (TCD) is used to measure cerebral blood flow velocity (CBFV) and vasomotor reactivity (by an index measured during breath-holding – BHI). Aim of this study was the metabolic profiling of migraine subjects with T/T677-MTHFR and C/T677-MTHFR mutations and its correlation with CBFV and BHI.

Metabonomic multidimensional techniques were used to describe and cluster subjects. Fifty women suffering from migraine (age: 18-64; 21 with aura) underwent TCD examination, hematochemical blood analysis, Born test, and genetic tests for MTHFR mutation. Fourteen (7 with aura) had T/T677, 18 (8 with aura) had C/T677, and 18 (6 with aura) had no mutation. The total number of variables was 24.

Unsupervised and supervised techniques showed the correlation between CBFV and BHI with mutation. Discriminant analysis allowed for classifying the subjects with 95.9% sensitivity and 89.0% specificity. Aura was not correlated to mutation or variations of instrumental data.

Our study showed that metabonomics could be effectively applied in clinical problems to show the overall correlation structure of complex systems in pathology. Specifically, our results confirmed the importance of TCD in the metabolic profiling and follow-up of migraine patients.

Keywords: Metabonomics, migraine, genetic mutations, PCA, PLS, classification, MTHFR, data mining, transcranial Doppler sonography.

1. INTRODUCTION

Migraine is a neurological disorder that correlates with an increased risk of subclinical cerebral vascular lesions [1]. Subjects suffering from migraine with aura (MwA) are particularly exposed to such a vascular risk [2]. There is a trend in considering migraine as a systemic vasculopathy [3]. The relationship between migraine with aura and cerebral lesions is not fully known, but it seems that the increased vascular risk of the migraineurs is not caused by the traditional risk factors (i.e., hypertension, hyperglycemia, hypercholesterolemia, …) [2]. However, the above-mentioned factors have implications on the possible drug therapies administered to migraine sufferers.

It has been demonstrated that migraine is associated to alterations in some hematochemical parameters related to platelet aggregation [4, 5]. The biological response to some agents acting on platelet function has been found altered in women suffering from migraine: adenosine diphosphate, collagen, ristocetin, epinephrine, prostacyclin, arachidonic acid, and 5-hydro-triptamine [6]. These alterations could increase the risk of vascular diseases on a thrombotic base.

Another important alteration that was correlated to MwA is the increase in the concentration of homocysteine. This amino acid, which is linked to the metabolism of the B6 and B12 vitamins, represents a powerful risk factor for cardiovascular diseases when highly concentrated in blood serum [7, 8]. The genetic mutation T/T 677 in homozygosis of the genotype 677-MTHFR (methylenetetrahydrofolate reductase) may cause increased levels of blood homocysteine. This mutation has been correlated to MwA [9] and to an increment in the angiopathies due to oxidative stress [10]. Recent studies proved that the T/T 677 mutation correlates to hyperhomocysteinemia, peripheral arteries pathologies [11], and cardiovascular events in presence of hypertension [12]. Therefore, due to the possible vascular implications they carry, genetic mutations in MwA are being more and more studied.

*Address correspondence to this author at the Biolab – Department of Electronics, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Torino, Italy; Tel: +39 11 564 4135; Fax: +39 11 564 4217; E-mail: filippo.molinari@polito.it

Given the correlation between migraine and vascular disorders, migraine sufferers usually undergo assessment of the cerebrovascular status. The transcranial Doppler sonography (TCD) is a non-invasive and reliable technique to measure the cerebral blood flow velocity (CBFV) in the arteries of the macro-circulation [13]. Some Authors evidenced that MwA patients present altered CBFV values compared to healthy controls, both in resting conditions [14] and in reactivity maneuvers [15]. The assessment of cerebral vasomotor reactivity (i.e., the arteries' capacity of compensating systemic blood pressure alterations) is of primary importance to assess the overall status of the artery bed. Despite this, the clinical utility of TCD in diagnosis and follow-up of migraine sufferers is still debated.

The aims of this study were: 1) to test the importance of TCD examinations in migraine sufferers by studying the correlations of CBFV with hematochemical and genetic parameters; 2) to develop a classifier for the detection of the 677-MTHFR genotype based on hematochemical parameters and instrumental data (i.e., TCD analysis).

We studied a population of female migraineurs. We applied metabonomic techniques to a database consisting of metabolic and hematochemical data, genetic data, and instrumental data coming from Doppler clinical analyses. Metabonomics, which belongs to the "-omics" sciences, quantitatively measures living systems undergoing the effects of diseases of genetic mutations. Unlike genomics and proteomics, metabonomics focuses on the multiparameter evaluation of a living complex system by studying its overall physiological profile. In this study, we combined unsupervised and supervised analysis techniques. The final result of unsupervised techniques is not guided by the maximization of externally imposed classification goals (i.e., discrimination between normal and pathologic, or placebo *vs* drug administration, or different classes of disease). Rather, unsupervised techniques maximize the overall perception of the data set feature by minimizing the number of variables used to represent them. Therefore, unsupervised techniques allow for an unbiased description of the natural correlation structure in the data set. Conversely, supervised methods aim at maximizing an externally imposed task, such as the separation of a-priori defined classes. Given the difficulty of the task and the biological variability we expected in this study, we combined unsupervised and supervised techniques: the unsupervised approach was used to drive the supervised analysis. We used the principal component analysis (PCA) as unsupervised technique, and the partial least squares - discriminant analysis (PLS-DA) as supervised approach.

2. MATERIALS AND METHODS

2.1. Description of the Database

We considered in this study 50 women suffering from migraine, diagnosed following the criteria of the International Headache Society [16], that were consecutively admitted to the observation of the Women's Headache Center of the University of Torino (Italy). The mean age of the patients was 38.5 ± 1.6 years (range: $18 - 64$ years). Twenty-one patients suffered from MwA. The patients were referred to the Neurology Division of the Gradenigo Hospital of Torino (Italy), where all the instrumental analyses (i.e.,

TCD examinations) were carried out. Exclusion criteria were the presence of headache other than migraine or of chronic metabolic diseases.

All the subjects were informed about the purpose of the study and about data acquisition and treatment. All patients signed an informed consent before entering the study. The local Institutional Review Board approved this research.

In the following we describe in detail the database. In the description, we separate instrumental data (subsection 2.1.1), hematochemical data (subsection 2.1.2), and genetic data (2.1.3). To the best of our knowledge, this is the only study that considers a mixed database consisting of metabolic, clinical, and instrumental variables.

2.1.1. TCD Data

All the subjects underwent TCD examination of the middle cerebral arteries (MCA). No subjects were discarded due to poor bone window. We used a commercially available TCD device (X4 MultiDop, DWL, Germany) equipped by two 2 MHz ultrasound probes. The probes were placed in correspondence of the temporal window and were firmly held in place by using an apposite holder. The same operator (a sonographist with more than 30 years of experience in vascular ultrasounds) performed all the examinations. We measured the CBFV in the M1 tract of the subjects' MCAs.

The TCD recording took place in a quiet and dimmed room, with the subjects lying in a supine and comfortable position, and breathing room air. All the TCD examinations were carried out in the interictal period, so that the subjects were free of pain. After 5 minutes of baseline recording, the subjects were asked to hold the breath. Breath – holding (BH) is a powerful stimulus that induces vasodilation in the cerebral arteries [17]. Vasodilation causes a drop in the resistance of the peripheral vessels, thus increasing cerebral blood flow. The measurement of the CBFV variation during BH is effective in assessing the cerebral vascular status of the subject [18]. We measured the vasomotor reactivity by using the breath – holding index (BHI), which is defined as:

$$BHI = \frac{V_{BH} - V_{BASE}}{V_{BASE}} * 100 \tag{1}$$

where:

- V_{BH} represents the CBFV measured at the end of the BH and averaged on a 60-seconds time window;

- V_{BASE} represents the CBFV measured at the beginning of the BH and averaged on a 60-second time window.

It has been shown that in the population of MwA sufferers there is an increased percentage of cardiac atrial septal defects, including patent foramen ovale (PFO) [19, 20]. PFO has been indicated as one of the possible causes of cerebrovascular accidents affecting MwA sufferers. Therefore, we considered the presence of emboli in the intracranial circulation as an index of cerebrovascular risk.

When a blood cloth flows in the insonated cerebral artery, it produces a characteristic signal called "high-intensity transient signal" (HITS). The TCD device we used was equipped by an automated HITS detection algorithm. We measured the number of the detected HITS in both the

MCAs and inserted these numbers into the instrumental variables set.

Finally, in a previous study conducted on a relatively large sample population of Italian women, we showed that CBFV decreases with age [21]. Therefore, we also considered age as a variable in our database. Table **1** summarizes the six instrumental variables (plus age) that we used in this study.

Table 1. **TCD Data. TCD Data (with Unit) that were Inserted in the Database**

Variable	Unit
Age	year
Left CBFV	cm/s
Right CBFV	cm/s
Left BHI	%
Right BHI	%
Left HITS	a.u.
Right HITS	a.u.

2.1.2. Hematochemical Data

The same laboratory made all the hematological analyses. The subjects were asked to undergo blood analysis within 2 days from the TCD examination. Since our purpose was to develop a methodology based on data that are usually acquired during migraine treatment protocols, we roughly kept the same list of blood analysis that was routinely used at the Women's Headache Center of Torino (Italy). Table **2** reports the complete list of variables we considered. The first column reports the hematochemical variables. The patients underwent the Born test for platelet aggregation under the effect of some agents [22]. The list of agents we considered is reported by the second column of Table **2**.

2.1.3. Genetic Data

All the patients were asked to undergo a genetic test to discover possible mutations of the 677-MTHFR genotype. The C/T677-MTHFR variant (leading to an alanine to valine substitution) results in a thermolabile enzyme and decreased production of folate, which is a cofactor required for homocysteine remethylation. Homozygosity for the 677-MTHFR variant is associated with hyperhomocysteinemia up to 50% higher than normal. The testing methodology was Direct Mutation Analysis**:** an automated high-throughput system incorporating DNA amplification (PCR), primer extension, and allele resolution by Matrix Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF) Mass Spectrometry was used to test for variant̲s in the MTHFR gene.

The genetic test revealed 14 subjects with T/T677-MTHRF mutation, 18 with C/T677-MTHFR mutation, and 18 with no mutation. Seven patients with T/T677 mutation, eight with C/T677 mutation, and six with no mutation suffered from MwA.

Table 2. **Data from Hematochemical Analysis and Born Test (Biochemical Data). The First Column of the Table Reports the Hematochemical Data (with Unit) that were Inserted in the Database. The Second Column of the Table Reports the Agents we Considered for the Platelet Aggregation Response (Born Test) (the Response is Given in Percentage)**

Hematochemical Variable	Agents for Platelet Aggregation Test (Born Test)
Antithrombin (%)	Arachidonic Acid (%)
Activated partial thromboplastin time (APTT) (s)	Adenosine diphosphate (ADP) (1 micromolar) – ADP1 (%)
Hemoglobin (g/dl)	Collagen (2 micromolar) (%)
Erythrocytes (x1000000/μ)	Ristocetine (%)
Fibrinogen (mg/dl)	Arachidonic Acid (%)
Leucocytes (x1000/μl)	
Platelets (x1000/μl)	
C Protein (%)	
S Protein (%)	
Prothrombin (%)	
Prothrombin time in International Normalized Ratio – Ptinr (a.u.)	
Activated partial thromboplastin time (APTT) (s)	

2.2. Statistical Analysis

Multivariate analysis was applied to the data set constituted by the above-described variables. The raw data set was a matrix having as rows (statistical units) 50 patients and as columns (variables) the 24 values relative to the hematochemical, instrumental, and genetic exams.

We used an analysis strategy combining multidimensional ANOVA analysis, unsupervised analysis, supervised analysis and classification. The entire database was first analyzed by using ANOVA and unsupervised techniques, in order to extract the correlation structure that constitutes the inner texture of the data. We considered the mutation type and the presence or absence of aura as principal factors. The right and left CBFVs were considered as dependent variables. We used ANOVA to detect the most significant variables for the dependent variables. This helped us in reducing the problem complexity and representation. Subsequent unsupervised and supervised techniques were then conducted on a reduced variables set, in order to better detail the inner correlation structure of the data set.

To build the classifier, the raw data set was then split in two parts. This procedure is known as cross-validation. We decided to use it in order to avoid possible over fitting of the data set given by the relatively small number of patients.

All the statistical analyses were carried out by using StatGraphics Centurion XV. In the following, we briefly explain the analysis techniques we used in this research.

2.2.1. Multidimensional Analysis and Outlier Rejection

Prior of performing unsupervised and supervised clustering, we performed a multivariate analysis of the raw data. Specifically, we aimed at evidencing the variables that are linked to genetic mutation in migraine.

In a previous study, we demonstrated that the left and right CBFV of healthy women are highly correlated [21]. Right and left CBFV as well as right and left BHI were correlated also for migraine subjects (Pearson's correlation coefficient always greater than 0.9, *p < 0.0001*).

We considered as dependent variables the left and right CBFV. Multidimensional ANOVA tests were performed in order to extract the variables that influence the dependent variables. Mutation and the presence of aura were considered as factors. Among the variables (biochemical data) we removed all the variables with a correlation coefficient greater than 0.6.

We performed an outlier analysis using the Q and T^2 statistics. In processing the hematochemical and platelet-related data, we found six outliers. However, we decided to keep them since they could be explanatory of pathology. When processing the instrumental data, we found four outliers relative to right CBFV (1 subject), left CBFV (2 subjects), and right BHI (1 subject). Again, we preferred to keep such subjects in the database and note them in subsequent statistical analysis.

2.2.2. Principal Component Analysis (PCA)

PCA is an unsupervised technique that effectively represents the information embedded in multidimensional data sets [23]. The raw data are represented in a transformed domain with lower dimensionality. The correlation structure of the raw data generates few latent variables (or principal components - PCs). PCs form an orthogonal basis and are sorted in order of decreasing explained data variance. Each PC can be expressed as $PC = aX_1 + bX_2 + cX_3 + \dots$ where X_1, X_2, X_3, … represent the measured features and a, b, c, … represent numerical weights. Each statistical unit is assigned a score relative to each extracted component, whereas the correlation coefficients between the original variables and the extracted components (loadings) furnish the significance of the specific PC.

2.2.3. Partial Least Squares (PLS)

PLS is a supervised technique that aims at finding the better linear combination explaining experimental data. With reference to our study, we can define $\mathbf{X}$ as the matrix containing the measured variables and $\mathbf{Y}$ the matrix of the genetic mutation. Hence, $\mathbf{X}$ contains the independent and $\mathbf{Y}$ the dependent variables. The goal of PLS is to find the linear combination of $\mathbf{X}$ that better explains $\mathbf{Y}$. The initial linear model can be expressed as:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{2}$$

where $\mathbf{B}$ is the regression coefficient matrix and $\mathbf{E}$ is the error matrix (same dimensions as $\mathbf{Y}$). Given the difference in the numerical values of our data, we centered $\mathbf{X}$ and $\mathbf{Y}$ by subtracting their mean value and by normalizing with respect to their standard deviations.

When in presence of many collinear variables, PCA may fail in properly extracting the covariance structure between the predictors. Conversely, PLS only extracts the covariance structure between the predictors and the response variables, thus leading to a more reliable system description. The PLS defines a matrix $\mathbf{T}$, so that $\mathbf{T} = \mathbf{XW}$; where $\mathbf{W}$ is an appropriate weight matrix and $\mathbf{T}$ is the factor score matrix. Then, the considered linear regression model becomes $\mathbf{Y} = \mathbf{TQ} + \mathbf{E}$; where $\mathbf{Q}$ is the matrix of regression coefficients (loadings) for $\mathbf{T}$, $\mathbf{E}$ is the error matrix. Once $\mathbf{Q}$ is calculated, the overall system is equivalent to the one in eq. (2), where $\mathbf{B} = \mathbf{WQ}$. This can now be used as a linear predictive regression model. Roughly, PLS can be thought as a PCA applied to the $\mathbf{X}$ matrix, a PCA applied to the $\mathbf{Y}$ matrix and the correlation analysis of the two sets of obtained PCs.

Using the standard NIPALS algorithm we performed all PLS numerical computations.

2.2.4. Discriminant Analysis (DA)

Discriminant function analysis (DA) is used to determine which variables better discriminate between two or more occurring groups in the sample population. In our study, there are three natural groups, each one corresponding to genetic classification (no mutation, C/T677, and T/T677 mutation). Computationally, DA determines a set of weight multiplying the $\mathbf{X}$ variables, so that the assignment error of each statistical unit to the correct $\mathbf{Y}$ class is minimized. We used PLS-DA to build a supervised model that could predict genetic mutation on the basis of patients' instrumental and hematochemical data.

To build the classifier, the raw data set was split in two parts. Cross – validation was carried out it in order to avoid possible over fitting of the dataset given by the relatively small number of patients. We randomly selected 40 patients to build the classifier and we used the remaining 10 subjects for validation. We tested 100 different possible combinations of 40 (classifier set) and 10 (validation set) subjects, with the condition that in the validation subset there were at least 2 subjects with C/T677 mutation, 2 with T/T677, and 2 with no mutation. We forced this condition since a preliminary study (results not reported in this paper) revealed that the classifier did not perform correctly when build on a heterogeneous subset and validated on a homogeneous one. Therefore, we only considered the random combinations leading to subsets in which all the three classes of subjects (no mutation, C/T677, and T/T677 mutation) were present.

3. RESULTS

Table **3** reports the result of the multidimensional ANOVA analysis performed on the hematochemical and platelet-related data of Table **2** when considering the left and right CBFV as dependent variables. The principal factors were considered the presence of mutation and of aura. The CBFVs are influenced by response to ADP3 and collagen (p < 0.001). The left CBFV is also related to the respone to ristocetin (p < 0.001), whereas right CBFV is not. Between the principal factors, only mutation is related to blood velocity, whereas the presence of aura is not. The significant variables were used in the following unsupervised step to confirm relations between instrumental and biochemical data.

Table 3.   **ANOVA Results. Covariable Significance as Computed by ANOVA Analysis. Principal Factors are "Mutation" and "Aura" (Reported in Italics as Last Row of Each Dependent Variable). The Dependent Variables are the Left and Right CBFV (Listed in the First Column). The Second Column Reports the Name of the Significant Covariables and the Third Column Reports the Associated P Value**

Dependent Variable	*Covariable*	*P Value*
Left CBFV	ADP3	0.0040
	Collagen	0.0207
	Ristocetin	0.0340
	Mutation	*0.0424*
	Aura	*0.5307*
Right CBFV	ADP3	0.0021
	Collagen	0.0057
	Mutation	*0.0230*
	Aura	*0.1553*

PCA was performed on a data set consisting of the 50 subjects (rows) and the following 7 observations: ADP3, collagen, ristocetin, right CBFV, left CBFV, right BHI and left BHI. All the variables were standardized. We extracted 3 PCs; their combination explained 79.9% of the total variance of the data. Fig. (**1**) reports the graph of the seven eigenvalues. The eigenvalues are proportional to the percent of variability in the data they explain. We chose the eigenvalues greater than 1 (horizontal line in Fig. **1**), since eigenvalues lower than 1 explain the same variance of the original variables, so they are not useful for data analysis in a reduced domain.
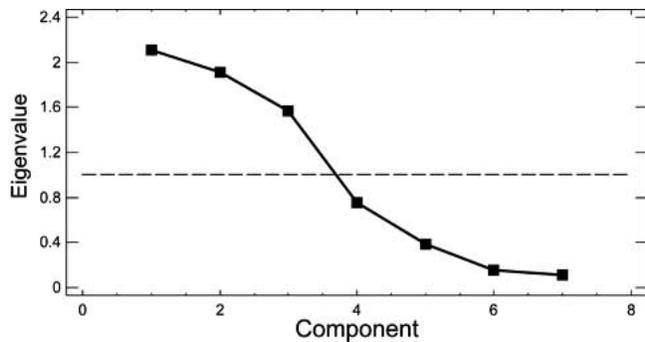

Fig. (1). Eigenvalues of the PCA applied to the data set consisting of 7 variables (ADP3, collagen, ristocetin, right and left CBFV, right and left BHI). The first three eigenvalues are higher than 1 (horizontal dashed line). Therefore, these three eigenvalues were used to represent the subjects in a reduced domain of three principal components.

Table **4** reports the weights of the seven variables on the three components. Both the CBFVs have a high weight in the first component, whereas the BHIs in second and third. Also, ristocetin, collagen, and ADP3 responses have a negative high weight in the third component.

Table 4.   **PCA Components. Weights of the Seven Variables with Respect to the First Three Principal Components of the PCA Analysis**

Variable	*Component 1*	*Component 2*	*Component 3*
ADP3	-0.27	0.40	-0.45
Collagen	0.13	0.18	-0.57
Ristocetin	-0.04	0.37	-0.41
Left CBFV	0.60	-0.18	-0.21
Right CBFV	0.59	-0.21	-0.22
Right BHI	0.34	0.54	0.28
Left BHI	0.27	0.55	0.35

Fig. (**2**) reports the subjects' distribution on the hyperplanes formed by component 1 and 2 (uppermost graph), component 2 and 3 (central graph), and component 1 and 3 (bottom graph). Black circles represent the subjects in the T/T677 group, gray circles those in the C/T677 group, and white circles those without mutation. The black continuous lines represent the projection of the original variables on the hyperplanes. The first component separates the subjects with T/T677 mutation from the subjects with C/T677 mutation or no mutation (Fig. **2**, upper panel in the top right quadrant and bottom panel in the upper part of the chart). Component 2 separates the T/T677 mutation in the upper part of the chart (Fig. **2**, middle panel). Table **5** reports the result of the *t-test* performed on the mutation subgroups. Table **5** confirms that Component 1 differentiates T/T677 from C/T677, whereas Component 2 separates T/T677 from C/T677 and from the subjects without mutation.

Table 5.   ***t-Test* Comparing Different Mutations. Results of the *t-Test* Comparing the Three Mutation Subgroups. The Significance Threshold is 0.05. Asterisks Mark the Statistically Significant Differences. In Parenthesis, the Explained Variance by the Specific PCA Component**

Patient Groups	**Component 1 (30.13)**	**Component 2 (27.31)**	**Component 3 (22.46)**
No mutation *vs* C/T677	0.805	-0.946	0.783
No mutation *vs* T/T677	-1.459	-6.623*	-5.075
T/T677 *vs* C/T677	2.581*	-5.075*	-1.387

The overall result of PCA was that subjects with T/T677 mutation might be distinguished from the other subjects because of their PC values. However, such PCs are mainly correlated to the CBFVs (first component), BHI (second and third component), collagen, ADP3, and ristocetin (third component). This result confirms the importance of the vascular parameters alterations carried by genetic mutations in migraine.

The analysis of the subjects with respect to mutation was refined by using PLS as supervised approach. Independent variables were same as PCA, dependent variables were
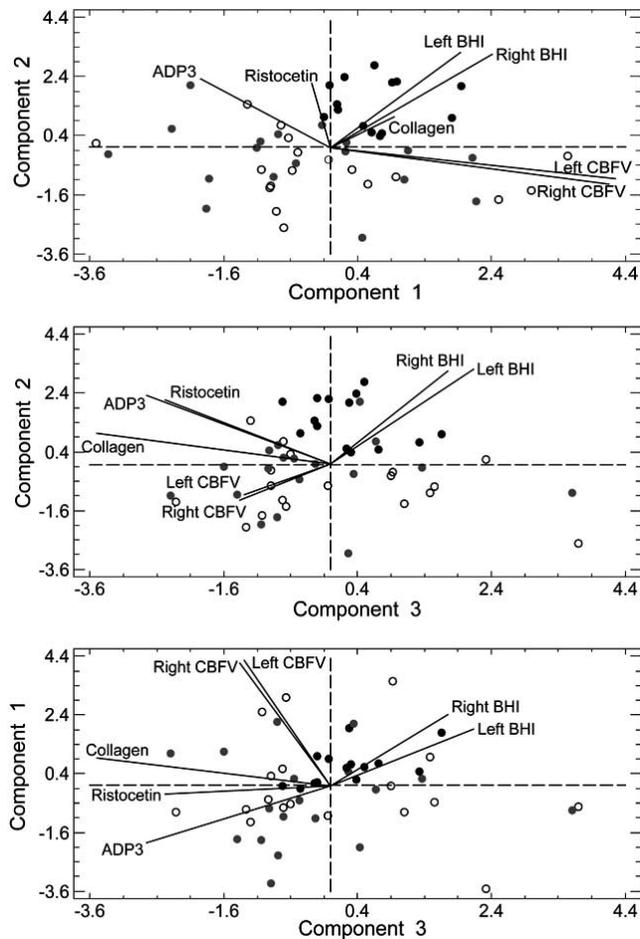
Fig. (2). PCA representation of the subjects in the three hyperplanes. Black circles represent the subjects in the T/T677 group, gray circles those in the C/T677 group, and white circles those without mutation. The dashed lines represent the zero level of the principal components. The black lines represent the projection of the original variables on the hyperplanes. Upper panel: PC1 *vs* PC2. Central panel: PC3 *vs* PC2. Lower panel: PC3 *vs* PC1.

mutation and aura. Table **6** summarizes the weights of the dependent and independent components. Fig. (**3**) reports the PLS discrimination of the subjects in the hyperplane (factor 2, factor 4). Factor 2, in particular, is discriminant of the T/T677 (black circles – left part of the graph) and C/T677 mutations (gray circles – right part of the graph). In particular, collagen responseis higher in the subjects without mutation with respect to T/T677 (*t-test, p < 0.001*) and in the subjects with C/T677 mutation with respect to T/T677 (*t-test, p < 0.002*).

The supervised PLS-DA classifier was built by coding by a number (0 = no mutation; 1 = C/T677 mutation; 2 = T/T677 mutation) the dummy variable **Y**. To build the PLS-DA classifier we relied on the previously performed ANOVA analysis and removed all the variables with a correlation coefficient higher than 0.6. This lead to remove the following four variables for the **X** matrix: antithrombin, PTinr, hematocrit, and erythrocytes. The classifier was developed on a set of 40 subjects (corpus set). Fig. (**4**) shows a sample of the discrimination of a corpus set performed by the first two functions. The black "+" (plus sign) mark the

centroids of the clusters. In this specific case, sensitivity is 100% and specificity is 100% (on the corpus set); the Wilks' lambda parameter is equal to 0.071, leading to *p < 0.003*.
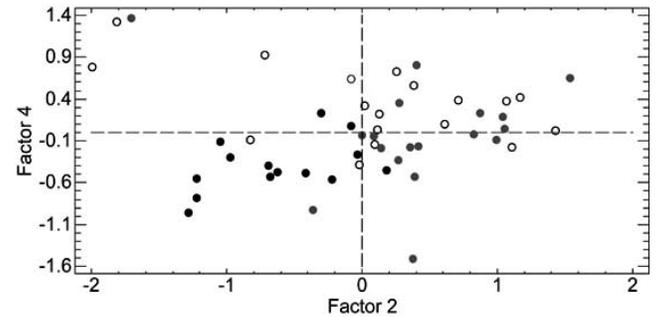

Fig. (3). Sample PLS representation of the subjects in the hyperplane defined by Factor 2 and Factor 4 (factors are numerically reported by table VI). Black circles represent the subjects in the T/T677 group, gray circles those in the C/T677 group, and white circles those without mutation. The dashed lines represent the zero level of the principal components. It can be observed that Factor 2 separates the genetic mutation T/T677 (left part of the chart – black circles) from the C/T677 (right part of the chart – gray circles).

Table 6. PLS Factors. Weights for the Dependent and Independent Variables for the PLS Analysis

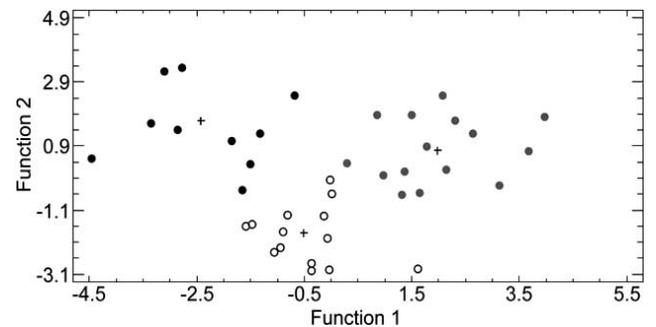| Independent Variables	Factor 1	Factor 2	Factor 3	Factor 4
Aura	0.36	-0.15	0.41	0.07
ADP3	-0.72	-0.41	-0.28	-0.18
Collagen	0.51	0.54	-0.03	-0.29
Ristocetin	-0.04	0.40	0.40	-0.05
Left BHI	0.01	-0.60	-0.29	-0.75
Right BHI	0.32	0.02	0.72	0.55
Dependent Variables	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 4*
Left CBFV	0.50	0.29	0.24	0.22
Right CBFV	0.52	0.28	0.10	0.22

Fig. (4). Sample PLS-DA representation of the 40 subjects of a corpus set in the hyperplane defined by the discriminant Functions 1 and 2. Black circles represent the subjects in the T/T677 group, gray circles those in the C/T677 group, and white circles those without mutation. The black "+" (plus sign) mark the centroids of the clusters.

The PLS-DA classifier was then used to cluster the remaining 10 subjects (validation set). Corpus and validation sets were randomly selected. We tested 100 combinations leading to corpus and validation sets containing at least 2 subjects for each group. Hence, we classified 1000 subjects, 363 of which belonged to the T/T677 group, 364 to the C/T677 group, and 273 to the no mutation group.

Each element of the validation set was assigned to a group on the basis of its distance from the centroids. For all the considered classifiers, the discriminant function p-value were always lower than 0.01. The classification performance were (between brackets we indicated the range of variation):

- 100% sensitivity (95.6% - 100%)and 97.6% specificity (94.1% - 98.3%) for the T/T677 mutation;

- 91.8% sensitivity (87.5% - 95.5%) sensitivity and 97.6% specificity (92.3% - 99.0%) for the C/T677 mutation;

- 89.0% sensitivity (84.7% - 92.4%) and 95.9% specificity (90.2% - 98.8%) for the group without mutation.

Overall classifier performance was 95.9% sensitivity, 89.0% specificity, and 94.0% efficiency.

4. DISCUSSION

Migraine is a complex neurological disorder involving different aspects, ranging from metabolic imbalances to neuronal and vascular anomalies. In this study, we profiled different groups of women suffering from migraine. The independent observation we considered in this study was the genetic mutation of the 677-MTHFR gene. We believe that, though preliminary, this study could be important in the clinical management of migraineurs.

We built a mixed database incorporating instrumental data (the CBFVs in the middle cerebral arteries and the cerebral reactivity index to BH), genetic data (the mutation type: C/T677 or T/T677, or the absence of mutation), and some biochemical parameters (including hematochemical variables and data related to platelet aggregation. The complete list is reported by Table **2**).

PCA and multivariate ANOVA essentially pointed out that, in our sample population of migraine sufferers, 677-MTHFR mutations are associated to altered platelet response to collagen, ristocetin, and ADP3. However, we also documented that the MTHFR mutations in migraine correlate with the instrumental measures of CBFVs and BHI. PLS results confirmed the importa-nce of the CBFVs and BHI patterns in the different subgroups of the genetic mutations (last four rows of Table **6**). When compared to C/T677 subgroup, T/T677 subjects showed increased BHIs (t-test, $p < 5 \times 10{-6}$ for left and $p < 6 \times 10{-7}$ for the right side).

The PLS-DA offered satisfactory classification performance when classifying a population of migraineurs on the basis of their genetic mutation. The overall classification performance was 95.9% sensitivity and 89.0% specificity. The most important variables for classification were the left and right CBFVs, the BHI, and, again, the platelet response to ADP3, collagen, and ristocetin. This finding confirms the importance of assessing the vasculature-related variables in migraine. Therefore, we

believe that TCD could be of primary importance also in the follow-up of migraine sufferers.

We found that the BHI is altered in subjects with genetic mutations (t-test, $p < 5 \times 10{-6}$). This result confirms that vascular reactivity may be impaired in migraineurs [24]. In 2008, Vernieri *et al*. showed that subjects suffering from MwA presented an increased BHI in the predominant migraine side with respect to non predominant and to control subjects [25], suggesting a possible deficit in the cerebrovascular autonomic control.

Our results are confirmatory of previously published studies. Novelty resides in two results that we obtained thanks to the applied multidimensional approach. First, we found that aura is not a major significant factor when related to mutation. In other words, the presence of aura was not correlated to the overall profile of the genetic mutation. This result is somehow new, since many studies pointed out that migraine with and without aura are characterized by different pathogenesis. Also, some pathologies have a greater prevalence in MwA than in migraine without aura (i.e., the patent foramen ovale [26]. We did not find any significant correlation between aura and the instrumental or biochemical variables. PCA and PLS showed that the metabolic profile of subjects with C/T677and T/T677 mutations is significantly different than subjects without mutation, even if aura is discarded among the classification variables.

The second important and novel result is that biochemical data alone are insufficient to perform classification. We tested a PLS-DA classifier based only on the variables of Table **2**. Basically, we removed the instrumental data from the data matrix X. Performance dropped to 54% sensitivity and 58% specificity ($p > 0.05$). Mutation, therefore, induces a global alteration of the metabolic profile of the subject that is better described by the joint analysis of hematochemical and vascular parameters.

In developing the PLS-DA classifier, we removed all the collinear variables with correlation coefficient higher than 0.6. We used this as stopping criterion for PLS regression. We tried to remove the variables with correlation coefficient higher than 0.5. In this case we deleted from the database five variables: antithrombin, PTinr, hematocrit, erythrocytes, and C protein. Classification performance was lower than in the former version: sensitivity was 90.1% and specificity was 85.2%. Therefore, we adopted 0.6 as threshold for assessing collinearity and for removing the co-variables. This result demonstrates the importance of a complete metabolic profiling for clustering MTHFR genetic mutations in migraine sufferers.

Finally, we did not perform outlier rejection. Being they outliers either for the biochemical variables or for the instrumental variables, outliers were correctly classified in all the tests we did. Therefore, we assumed that outliers might be representative of the variability of the data that can be obtained when working with a complex pathology like migraine.

The principal limitation of this study relies in the relative paucity of the sample database. However, results seem very promising since even the first multivariate analysis we performed (unsupervised PCA, Fig. **1**) showed a natural tendency in clustering T/T677 mutations. The application of this

method corresponds to depicting a sort of 'natural normalization' of the studied data set, given that principal components correspond to the eigenvectors of the correlation matrix, which in turn corresponds by definition to the covariance matrix of the standardized variables. This is particularly convenient when dealing with heterogeneous variables defined by completely different measurement units, ruling out all questionable a priori defined standardization processes. Another limitation of the study is the small number of hematochemical data inserted into the database. We decided to develop a study based on exams that are routinely done. Therefore, we did not specifically test any hematochemical variable outside the list in Table **2** (first column), which somehow reflects is the list of blood examinations requested by the Institution where the patients are followed.

To the very best of our knowledge, this is the first attempt to use metabonomic statistical techniques applied to a clinical data set consisting of metabolic and instrumental data. We believe that this approach could constitute a pilot experience to bring metabolomic (and, in general, data mining) techniques into clinical practice, when tuned for specific and well-defined groups of patients.

5. CONCLUSION

By using a metabonomic approach, we developed a classifier to detect genetic mutations of the 677-MTHFR gene in a population of women suffering from migraine. Our database consisted biochemical and instrumental data. Transcranial Doppler ultrasonography was used to measure the baseline values of the cerebral blood flow and their variation during voluntary breath – holding. Unsupervised and supervised approaches were used to extract the variables correlating to mutation. Our classifier showed overall satisfactory performance: 95.9% sensitivity and 89.0% specificity. We found that cerebral blood flow velocities and response to breath – holding are important variables in characterizing mutations in migraine.

This pilot study adapting metabonomic techniques to the profiling of pathologic subjects could be a basis for the development of a clinically applicable methodology devoted to the profiling, classification, and evolutionary modeling of migraine patients.

REFERENCES

[1]   Kruit MC, van Buchem MA, Hofman PA, *et al*. Migraine as a risk factor for subclinical brain lesions. JAMA 2004; 291(4): 427-34.

[2]   Scher AI, Terwindt GM, Picavet HS, Verschuren WM, Ferrari MD, Launer LJ. Cardiovascular risk factors and migraine: the GEM population-based study. Neurology 2005; 64(4): 614-20.

[3]   Tietjen GE. Migraine as a systemic vasculopathy. Cephalalgia 2009; 29(9): 987-96.

[4]   Kozubski W, Walkowiak B, Pawlowska Z, Prusinski A, Cierniewski CS. Blood platelet fibrinogen receptors in migraine and related headaches. Neurol Neurochir Pol 1996; 30 (Suppl 2): 25-33.

[5]   Tietjen GE, Khubchandani J. Platelet dysfunction and stroke in the female migraineur. Curr Pain Headache Rep 2009; 13(5): 386-91.

[6]   Allais G, Facco G, Ciochetto D, De Lorenzo C, Fiore M, Benedetto C. Patterns of platelet aggregation in menstrual migraine. Cephalalgia 1997; 17 (Suppl 20): 39-41.

[7]   Moschiano F, D'Amico D, Usai S, *et al*. Homocysteine plasma levels in patients with migraine with aura. Neurol Sci 2008; 29 (Suppl 1): S173-5.

[8]   Bottini F, Celle ME, Calevo MG, *et al*. Metabolic and genetic risk factors for migraine in children. Cephalalgia 2006; 26(6): 731-7.

[9]   Scher AI, Terwindt GM, Verschuren WM, *et al*. Migraine and MTHFR C677T genotype in a population-based sample. Ann Neurol 2006; 59(2): 372-5.

[10] Sucker C, Kurschat C, Farokhzad F, *et al*. The TT genotype of the C677T polymorphism in the methylentetrahydrofolate reductase as a risk factor in thrombotic microangiopathies: results from a pilot study. Clin Appl Thromb Hemost 2009; 15(3): 283-8.

[11] Khandanpour N, Willis G, Meyer FJ, *et al*. Peripheral arterial disease and methylenetetrahydrofolate reductase (MTHFR) C677T mutations: A case-control study and meta-analysis. J Vasc Surg 2009; 49(3): 711-8.

[12] Durga J, Verhoef P, Bots ML, Schouten E. Homocysteine and carotid intima-media thickness: a critical appraisal of the evidence. Atherosclerosis 2004; 176(1): 1-19.

[13] Newell DW, Aaslid R. Transcranial Doppler: clinical and experimental uses. Cerebrovasc Brain Metab Rev 1992; 4(2): 122-43.

[14] Nowak A, Kacinski M. Transcranial Doppler evaluation in migraineurs. Neurol Neurochir Pol 2009; 43(2): 162-72.

[15] Anzola GP. Clinical impact of patent foramen ovale diagnosis with transcranial Doppler. Eur J Ultrasound 2002; 16(1-2): 11-20.

[16] Classification and diagnostic criteria for headache disorders, cranial neuralgias and facial pain. Headache Classification Committee of the International Headache Society. Cephalalgia 1988; 8 (Suppl 7): 1-96.

[17] Molinari F, Liboni W, Grippi G, Negri E. Relationship between oxygen supply and cerebral blood flow assessed by transcranial Doppler and near-infrared spectroscopy in healthy subjects during breath-holding. J Neuroeng Rehabil 2006; 3: 16.

[18] Silvestrini M, Baruffaldi R, Bartolini M, *et al*. Basilar and middle cerebral artery reactivity in patients with migraine. Headache 2004; 44(1): 29-34.

[19] Rothrock JF. Patent foramen ovale (PFO) and migraine. Headache 2008; 48(7): 1153.

[20] Anzola GP, Del Sette M, Rozzini L, *et al*. The migraine-PFO connection is independent of sex. Cerebrovasc Dis 2000; 10(2): 163.

[21] Liboni W, Allais G, Mana O, *et al*. Transcranial Doppler for monitoring the cerebral blood flow dynamics: normal ranges in the Italian female population. Panminerva Med 2006; 48(3): 187-91.

[22] Born GV. Aggregation of blood platelets by adenosine diphosphate and its reversal. Nature 1962; 194: 927-9.

[23] Benigni R, Giuliani A. Quantitative modeling and biology: the multivariate approach. Am J Physiol 1994; 266(5 Pt 2): 1697-704.

[24] Liboni W, Molinari F, Allais G, *et al*. Why do we need NIRS in migraine?. Neurol Sci 2007; 28 (Suppl 2): 222-4.

[25] Vernieri F, Tibuzzi F, Pasqualetti P, *et al*. Increased cerebral vasomotor reactivity in migraine with aura: an autoregulation disorder? A transcranial Doppler and near-infrared spectroscopy study. Cephalalgia 2008; 28(7): 689-95.

[26] Diener HC, Kurth T, Dodick D. Patent foramen ovale and migraine. Curr Pain Headache Rep 2007; 11(3): 236-40.