

Transcriptomic analysis of elderly women with low muscle mass: association with immune system pathway

Levi H. Jales Neto^{1,*}, Bidossessi W. Hounkpe^{1,*}, Georgea H. Fernandes¹, Liliam Takayama¹, Valéria F. Caparbo¹, Neuza H.M. Lopes², Alexandre C. Pereira³, Rosa M.R. Pereira¹

¹Bone Metabolism Laboratory, Rheumatology Division Faculdade de Medicina FMUSP, Universidade de Sao Paulo, Sao Paulo, SP, Brazil

²Instituto do Coracao (InCor), Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, SP, Brazil

³Laboratory of Genetics and Molecular Cardiology, Instituto do Coração (InCor), Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, SP, Brazil

*Equal contribution

Correspondence to: Rosa M.R. Pereira; email: rosa.pereira@hc.fm.usp.br

Keywords: low muscle mass, muscle weakness, immune system, vitamin D, aging

Received: May 24, 2021

Accepted: August 24, 2021

Published: September 7, 2021

Copyright: © 2021 Neto et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Despite the well-established association of gene expression deregulation with low muscle mass (LMM), the associated biological mechanisms remain unclear. Transcriptomic studies are capable to identify key mediators in complex diseases. We aimed to identify relevant mediators and biological mechanisms associated with age-related LMM. LMM-associated genes were detected by logistic regression using microarray data of 20 elderly women with LMM and 20 age and race-matched controls extracted from our SPAH Study (GSE152073). We performed weighted gene co-expression analysis (WGCNA) that correlated the identified gene modules with laboratorial characteristics. Gene enrichment analysis was performed and an LMM predictive model was constructed using Support Vector Machine (SVM). Overall, 821 discriminating transcripts clusters were identified ($|\beta \text{ coefficient}| > 1$; $p\text{-value} < 0.01$). From this list, 45 predictors of LMM were detected by SVM and validated with 0.7 of accuracy. Our results revealed that the well-described association of inflammation, immunity and metabolic alterations is also relevant at transcriptomic level. WGCNA highlighted a correlation of genes modules involved in immunity pathways with vitamin D level ($R = 0.63$, $p = 0.004$) and the Agatston score ($R = 0.51$, $p = 0.02$). Our study generated a predicted regulatory network and revealed significant metabolic pathways related to aging processes, showing key mediators that warrant further investigation.

INTRODUCTION

Accelerated loss of muscle mass is one of the dramatic changes associated with aging and reflects muscle remodeling in elderly population [1, 2]. Age-related low muscle mass leads to lower mobility, disability and increased mortality in elderly population [1, 2]. The diagnosis is particularly based on the evaluation of appendicular lean mass (ALM) by dual X-ray

absorptiometry (DXA). Among the methods used to identify LMM, the criterion defined by Newman et al. in 2003 [1] is based on ALM measurement adjusted for total fat mass [3, 4]. Therefore, it seems to identify more individual with sarcopenia, particularly those with sarcopenic obesity [3, 4]. In a previous study, we define the Newman residual specific to our population and demonstrated that LMM is a risk factor for increased mortality in elderly men and women [4]. In our

Brazilian community a residual that yielded the best performance of sensitivity and specificity to define LMM in elderly women was -1.32 [4].

Multiple factors are involved in the development of age-related low muscle mass, including several known age-related factors, such as neuromuscular degeneration, abnormal protein turnover, decreased hormone levels and sensitivity, chronic inflammation, oxidative stress, and lifestyle (nutritional status and physical inactivity) [5]. Despite the well-described contribution of these factors, current therapies are only limited to lifestyle changes by mainly targeting resistance training and/or protein supplementation [6, 7]. There is no drug specifically approved for sarcopenia treatment [8]. Owing to the high prevalence of the clinical complications of LMM in elderly women, further studies are needed to identify possible effective therapies. Thus, defining key cellular and molecular mechanism underlying this condition is a crucial step to identify new relevant therapeutic targets. However, these mechanisms are not fully elucidated. We believe a global transcriptomic portrait of the systemic implications of LMM would advance our understanding of the biological processes associated with this condition, thereby paving the way to more efficient clinical intervention.

So far, the pathophysiological mechanisms of muscle mass loss in elderly women are not fully understood. Several studies have already pointed the contribution of genetic associations to low muscle mass process and its complications [9]. Application of transcriptomic technologies to muscle of sarcopenia patients has already identified key alterations of biological processes specific to human and murine muscle tissue [10, 11]. However, most studies have focused on local perturbations, rather than systemic implications of LMM at transcriptomic level.

In this study, we aim to identify gene sets and biological processes that differentiate elderly women with LMM from age-matched controls. Our study identified predictors of LMM at transcriptomic level and revealed the molecular mechanisms underlying the disease. Gene co-expression analysis revealed gene modules correlated with relevant laboratorial and clinical data of the patients, providing rather than a classical transcriptomic profile analysis, relevant clinical information associated with LMM gene expression profiles.

RESULTS

Sample characteristics

Using the definition of Newman et al. [1] and the residual previously defined for our population, we

included forty elderly women in the study, of which 20 had LMM (residual <-1.32) and 20 age and race-matched controls without LMM. Lower weight, total and appendicular lean mass and lower BMI were observed in the LMM group (Table 1). No difference in abdominal circumference was found between the groups. A higher percentage of fat was observed in the group with LMM (median 41.63 (3.94) vs. 38.61 (4.18); $p = 0.02$ for LMM and controls respectively). Regarding associated chronic diseases and medications for chronic use, no differences were found between groups. Comparing the biochemistry between the groups, a significantly lower serum level of vitamin D (25OHD) (18.75 (7.06) vs. 24.15 (6.74) ng/ml; $p = 0.018$) and serum insulin (11.6 (5.7) vs. 17 (14.57) UI/mL; $p = 0.02$) was observed in the LMM group. The participants included in this study are homogeneous in terms of lifestyle and habits. They reside in the same geographic region, and have the same socioeconomic, educational level and access to the same health system. As part of the SPAH cohort (São Paulo Ageing and Health Study) they were followed up in the city of São Paulo, and are routinely guided by medical doctors, medical students, and other health professionals from the University of São Paulo (USP), School of Medicine. Serum albumin levels were similar in both groups (4.65 ± 0.2 vs. 4.65 ± 0.22 g/dL; $p = 0.30$ for LMM and control respectively) and were in the normal range (3.5–5.2 g/dL) suggesting a reasonable nutritional level. Regarding consumption of milk and derivative products, no difference was observed between the groups with LMM and control (median 527.1 vs. 553.2 mg/day; $p = 0.75$ for LMM and control respectively). All other characteristics are presented in Table 1.

Discriminating transcript clusters and differential expression in LMM

In this analysis, for simplicity purposes, transcript clusters are sometimes identified by their corresponding genes symbols. The logistic regression model adjustment identified 821 unique transcript clusters of which 565 with positive coefficients (PPR) and 255 with negative coefficients (PNR) (Supplementary Table 1 and Figure 1A). Differential expression analysis identified 56 and 79 up- and down-regulated transcript clusters respectively (Supplementary Table 2). Among these transcript clusters, fifteen are also identified in the logistic regression model (Table 2). We then used unsupervised clustering of the normalized expression of the top 40 predictors to show their expression pattern between samples. This analysis stratifies the control and LMM groups in two different clusters (Figure 1B). The heat map of the expression profile shows a high homogeneity in each group when we consider this subset of 40 discriminating transcript clusters (Figure 1B).

Table 1. Clinical, anthropometric, biochemical and calcium score characteristics of the participants.

Variables	LMM <i>n</i> = 20	Control <i>n</i> = 20	<i>p</i> -value
Clinical variables			
Age, years, mean (SD)	80.45 (4.44)	79.65 (3.76)	0.54
Ancestry, <i>n</i> (%)			
Caucasian	10 (50)	10 (50)	1
Black	10 (50)	10 (50)	
BMI, kg/m ² , mean (SD)	28.17 (4.26)	32.68 (5.08)	0.004
Physical activity, <i>n</i> (%)			
Low	5 (25)	0 (0)	
Moderate	10 (50)	16 (80)	0.047
High	5 (25)	4 (20)	
Hypertension, <i>n</i> (%)	16 (80)	15 (75)	1.0
Diabetes Mellitus, <i>n</i> (%)	6 (30)	3 (15)	0.45
Dyslipidemia, <i>n</i> (%)	9 (45)	12 (60)	0.53
Heart attack, <i>n</i> (%)	1 (5)	2 (10)	1.0
Stroke, <i>n</i> (%)	2 (10)	1 (5)	1.0
Osteoporosis, <i>n</i> (%)	8 (0.4)	11 (0.55)	0.53
Hypothyroidism, <i>n</i> (%)	4 (20)	4 (20)	1.0
Alcohol intake, <i>n</i> (%)	2 (10)	3 (15)	1.0
Smoking, <i>n</i> (%)	1 (5)	1 (5)	1.0
Medication			
AAS, <i>n</i> (%)	7 (35)	7 (35)	1.0
ACE inhibitors, <i>n</i> (%)	5 (25)	8 (40)	0.50
Beta blocker, <i>n</i> (%)	6 (30)	8 (40)	0.74
Thiazide diuretic, <i>n</i> (%)	10 (50)	6 (30)	0.33
Calcium antagonists, <i>n</i> (%)	2 (10)	6 (30)	0.23
Statins, <i>n</i> (%)	7 (35)	11 (55)	0.34
Antidepressants, <i>n</i> (%)	4 (20)	0 (0)	0.10
Bisphosphonates, <i>n</i> (%)	8 (40)	8 (40)	1.00
Corticoid, <i>n</i> (%)	1 (5)	0 (0%)	1.00
Vitamin D supplement, <i>n</i> (%)	9 (45)	12 (60)	0.53
Lean and Fat Mass			
Total lean mass, g, mean (SD)	34294.25 (4738.27)	42698.73 (5344.11)	<0.0001
Appendicular lean mass, Kg, mean (SD)	13.13 (2.25)	18.26 (2.53)	<0.0001
Fat mass, g (SD)	26192.1 (5900.05)	27641.25 (9434.72)	0.31
Fat, % (SD)	41.63 (3.94)	38.61 (4.18)	0.02
Newman Residual, median (SD)	-2.77 (1.47)	1.55 (1.88)	<0.0001
Laboratory variables			
25OHD, ng/mL, mean (SD)	18.75 (7.06)	24.15 (6.74)	0.018
Total calcium, mg/dL, mean (SD)	9.86 (0.54)	9.61 (0.36)	0.09

iPTH, pg/dL, median (IQR)	61 (29.75)	50.5 (17)	0.48
Alkaline phosphatase, U/L, median (IQR)	78 (31.25)	70 (37.25)	0.43
Total phosphorus, mg/dL, mean (SD)	3.48 (0.32)	3.51 (0.37)	0.78
Creatinine clearance, mL/min/1.73m ² , median (IQR)	54.45 (16.65)	58 (37.82)	0.50
Total cholesterol, mg/dL, mean (SD)	210.95 (39.86)	206.40 (41.34)	0.72
LDL cholesterol, mg/dL, mean (SD)	123.35 (39.62)	127.30 (41.74)	0.76
HDL cholesterol, mg/dL, mean (SD)	59.75 (14.88)	52.60 (16.53)	0.17
Triglycerides, mg/dL, mean (SD)	139.20 (42.13)	150.75 (64.66)	0.51
Lipoprotein A, mg/dL, median (IQR)	67 (97)	50 (67.5)	0.93
Blood glucose, mg/dL, median (IQR)	101 (29.25)	93 (15.5)	0.62
Serum insulin, UI/mL, median (IQR)	11.6 (5.7)	17 (14.57)	0.02
C Reactive Protein, mg/L, median (IQR)	1.9 (3.55)	2.4 (3.05)	0.52
Albumin, g/dL, mean (SD)	4.65 (0.2)	4.65 (0.22)	0.30
Ferritin, ng/mL, median (IQR)	131.65 (161.44)	192.6 (136.7)	0.2
Serum uric acid, mg/dL, mean (SD)	4.89 (1.15)	5.61 (1.71)	0.13
Agatston method			
0, <i>n</i> (%)	8 (0.4)	4 (0.2)	
1–100, <i>n</i> (%)	4 (0.2)	6 (0.3)	0.5
>100, <i>n</i> (%)	6 (0.3)	9 (0.45)	
ND	2 (0.1)	1 (0.05)	
Dietary intake			
Milk and derivatives*, mg/day, median (IQR)	527.1 (433.35)	553.2 (518.25)	0.75

The Shapiro-Wilk test was used to evaluate the normality. The values of *p* are for comparisons of means (Student's *t*-test), median (Mann-Whitney test) or proportions (Chi-square test or Fisher's test). Values of *p* < 0.05 were considered significant. Abbreviations: 25OHD: 25-hydroxyvitamin D; BMI: body mass index; iPTH: intact parathyroid hormone, IQR: interquartile range; SD: standard deviation; ND: non-determined. *Milk and derivatives are presented as an estimated daily calcium ingested quantity.

Support vector machine-based LMM classification model

RFE algorithm discriminates a list of the 45 most informative variables from the discriminating transcripts clusters identified in the logistic regression adjustment. This subset of transcript clusters was used to train a SVM model with a polynomial kernel of degree 1. Model performance was evaluated in the test cohort and shows an accuracy of 0.70. This list of 45 classifiers is enriched in differentially expressed transcript clusters based on their fold changes (Supplementary Table 3).

Functional gene set enrichment highlights an alteration of muscle development and metabolism pathways in LMM

Genes associated with the subset of transcript clusters with positive and negative regression coefficients were

then used to perform to different gene set analyses with ClueGO in Cytoscape software. Using the lists of annotated genes, ClueGO identifies and clusters biological pathways and gene ontology terms that participate in the same biological function, thereby removing redundancy. The top significant and non-redundant biological pathways and gene ontology terms associated with PPR genes were shown in Figure 2A. Muscle development pathways and metabolism ontology terms were overrepresented in this analysis. ClueGO shows few gene interactions between the different enriched gene set groups, suggesting an absence of hub genes involved in the simultaneous regulation of different pathways (Figure 2A). Genes with negative regression coefficients were associated with metabolism deregulation (Supplementary Figure 1). Even though these analyses provide us interesting information about the mechanisms underlying the LMM, they did not reveal the human phenotypes that were associated with the observed alterations.

Human phenotype ontology (HPO) analysis reveals the involvement of immune and inflammatory alterations

To gain additional insights into the phenotypes previously associated with the LMM discriminating genes (identified in the logistic regression analysis), a

HPO analysis was performed with the full list of predictors using GenePattern [12]. Figure 2B shows several relevant phenotypes up-regulated in LMM. Phenotypes associated with immune system deregulation, inflammation and pathologies involving muscle disfunction were overrepresented in the HPO network. Interestingly, these phenotypes are highly

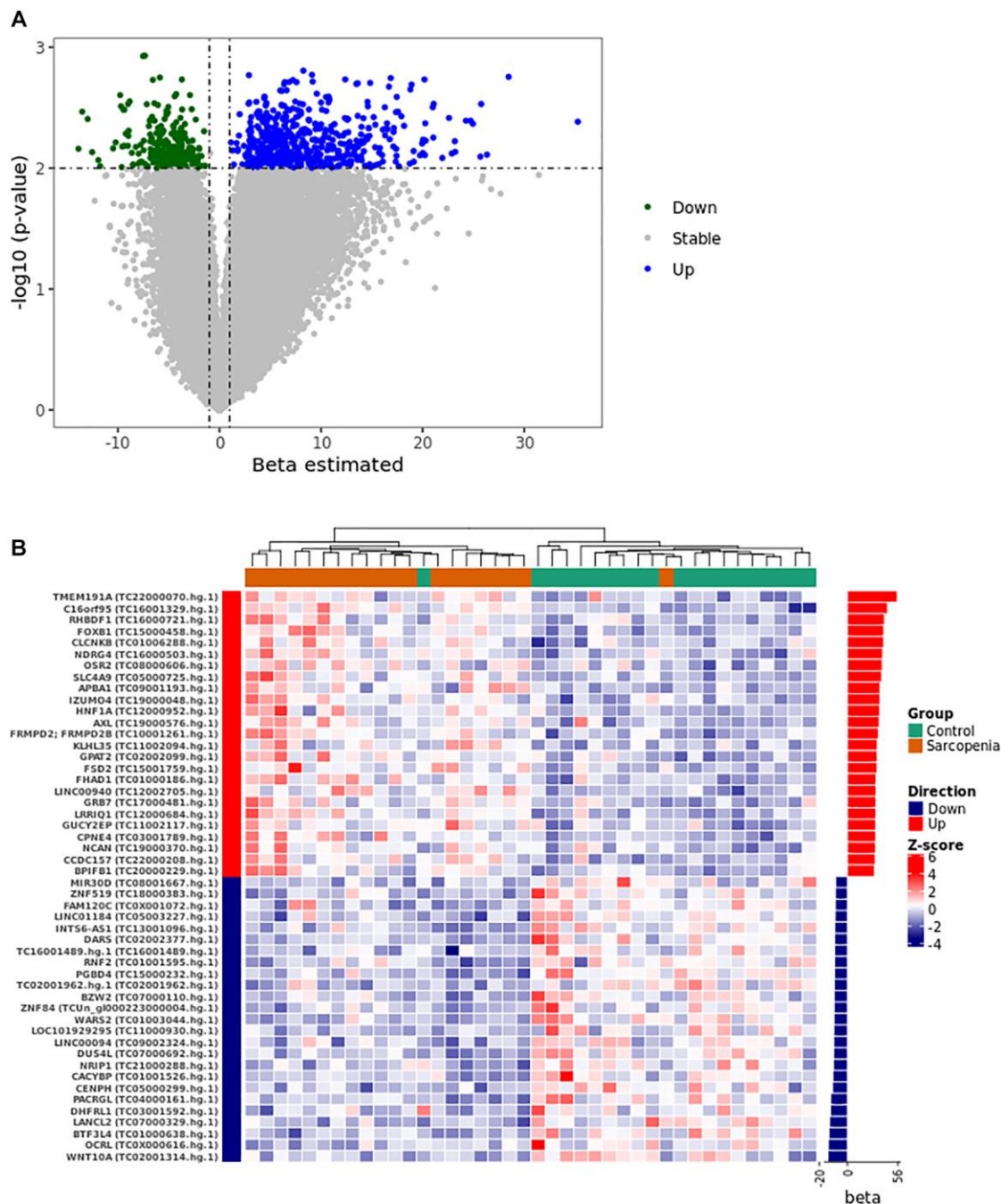


Figure 1. Gene expression pattern from the logistical regression analysis. (A) The volcano plot was constructed using the full list of 67528 transcript clusters analyzed. The top 821 transcripts were highlighted in blue (556 PPR) and green (255 PNR). A p -value of <0.01 and $|\text{beta}| >1$ was considered statistically significant. (B) The heat map shows the unsupervised clustering of the normalized expression pattern. The dendrogram indicates two clusters that stratified LMM and control group in distinct clusters. The right panel of the heat map shows the bar graph of beta coefficient values. Abbreviations: Beta: Estimated logistic regression coefficient; PPR: Predictor with Positive Relationship also called Up in panel A and B; PNR: Predictor with Negative Relationship also called Down in panel A and B.

Table 2. List of 15 differentially expressed transcript clusters identified in logistic regression model.

Transcript Cluster	Ensembl Genes	Gene Symbol	Biotype	Beta	FC
TC0X001828.hg.1	ENSG00000157600	TMEM164	Protein coding	1.29	1.82
TC0X000538.hg.1	ENSG00000157600	TMEM164	Protein coding	1.69	1.64
TC0X000540.hg.1	ENSG00000265584	MIR3978	miRNA	1.81	1.63
TC03001967.hg.1	ENSG00000200052	Y RNA	misc RNA	1.79	1.57
TC04001133.hg.1	ENSG00000252975	Y RNA	misc RNA	2.9	1.56
TC09000563.hg.1	ENSG00000200261	Y RNA	misc RNA	1.41	1.54
TC11000956.hg.1	ENSG00000282373	BIRC3	Protein coding	-2.21	0.64
TC08001180.hg.1	ENSG00000254165	AC090739.1	lncRNA	-2.17	0.62
TC01005703.hg.1	ENSG00000260948	AL390195.2	lncRNA	-2.91	0.61
TC0X002160.hg.1	ENSG00000166432	ZMAT1	Protein coding	-2.33	0.6
TC0X002160.hg.1	ENSG00000150347	ZMAT1	Processed Transcript	-2.33	0.6
TC10002092.hg.1	ENSG00000023445	ARID5B	Protein coding	-1.75	0.59
TC15000394.hg.1	ENSG00000244879; ENSG00000284284	GABPB1-AS1; MIR4712	LncRNA; miRNA	-1.52	0.56
TC14002234.hg.1	ENSG00000211923	IGHD3-10	IG D gene	-0.9	0.3

Transcript clusters were ranked according to the fold change. Abbreviations: FC: Fold-change; Beta: Estimated logistic regression coefficient.

connected, suggesting a tight relationship between them and highlighting a probable association of their common mediators with the physiological mechanisms that underlie LMM development. In addition, based on network connectivity analysis, the following genes were predicted as hub genes of the HPO network: (i) *CLCNKB* and *KDSR* (abnormality of endocrine system, glucose intolerance, abnormality of immune system physiology, abnormal inflammatory response, decreased body weight); (ii) *KIAA0586* and *OCRL* (abnormality of endocrine system, abnormality of immune system physiology, abnormal inflammatory response, decreased body weight, abnormal phalanx morphology); (iii) *FANCM* (abnormality of endocrine system, abnormality of immune system physiology, decreased body weight, abnormal phalanx morphology); (iv) *TGDS* (abnormality of immune system physiology, abnormal inflammatory response, decreased body weight, abnormal phalanx morphology); (v) *NDN* (abnormality of endocrine system, glucose intolerance, abnormality of immune system physiology, decreased body weight). These genes are highly connected with most of the phenotypes and might play prominent role in the regulation of LMM.

Prediction of the regulatory networks

The regulatory network system upstream to the expression of the discriminating genes (with positive regression coefficients) was reconstructed using

Expression2Kinase tool. This prediction tool allows us to identify in Figure 3A the top 20 predicted transcription factors (TFs) that can regulate the expression of the genes detected in this study (Figure 3A). The circos diagram in Figure 3B represents the relationship between the top 2 TFs that target the largest number of genes identified in our study (49 and 42 genes for *FOXLI* and *TCF4* respectively). We then analyzed the subnetworks of regulation and observed a common regulatory network pattern associated with *TCF4* and *HNF1A* that was identified as a PPR (estimated coefficient $\beta = 22.87$; p -value = 0.007). The regulatory network in Figure 3C presents these TFs (red nodes); (ii) the intermediate proteins likely to be involved in the formation of a regulatory complex with these TFs (green nodes) and (iii) the kinases that are expected to be the activators of the regulatory complexes (dark blue nodes). This predicted regulatory network connects *HNF1A* and *TCF4* and suggests a possibility of a co-regulatory system involving both TFs (Figure 3C).

Gene modules detected by weighted gene co-expression network analysis are correlated with patient's characteristics

Fourteen co-expression modules were identified in the weighted co-expression analysis performed exclusively with the patient group. Correlation of clinical and laboratorial data with gene modules shows tight

relationship with relevant parameters (Figure 4A). Pink module is moderately correlated with both serum levels of vitamin D ($r = 0.63$, p -value = 0.004) and Agatston score ($r = 0.51$, p -value = 0.02). This module regroups genes involved in immune system regulation including:

(I) innate immunity, phagocytosis, and neutrophil degranulation, (ii) B cell receptor and IL1 signaling pathways (Figure 4B). In contrast, green-yellow and light-green modules are negatively correlated with plasma level of vitamin D and body fat percentage

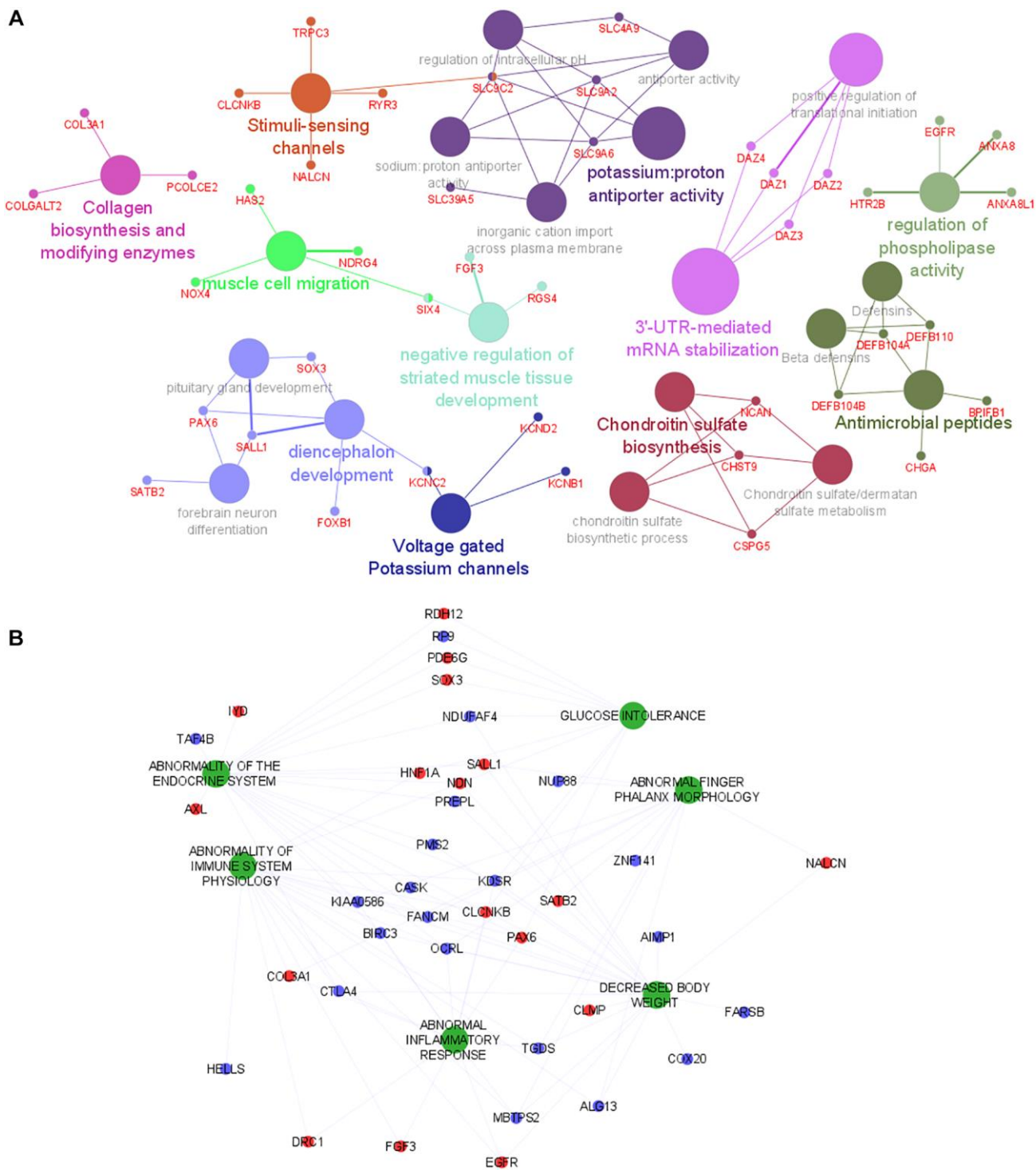


Figure 2. Functional analysis of genes and human phenotypes associated with LMM. (A) Pathways that are associated with the gene discriminated in the logistic regression (with positive coefficient) are indicated by colored nodes. GSA terms are interconnected with their associated genes. Related GSA terms are indicated by the same color. (B) The human phenotype ontologies are presented in green and shared genes are represented as red (positive coefficient) and blue (negative coefficient) dots. Abbreviation: GSA: Gene Set Analysis.

($r = -0.58$, p -value = 0.009 and -0.51 , p -value = 0.05 respectively). Both green-yellow and light-green modules are associated with ubiquitination mechanisms

(Figure 4C). In addition, genes of green-yellow are involved in erythrocyte specific metabolism pathways and oxidoreduction processes (Supplementary Figure 1).

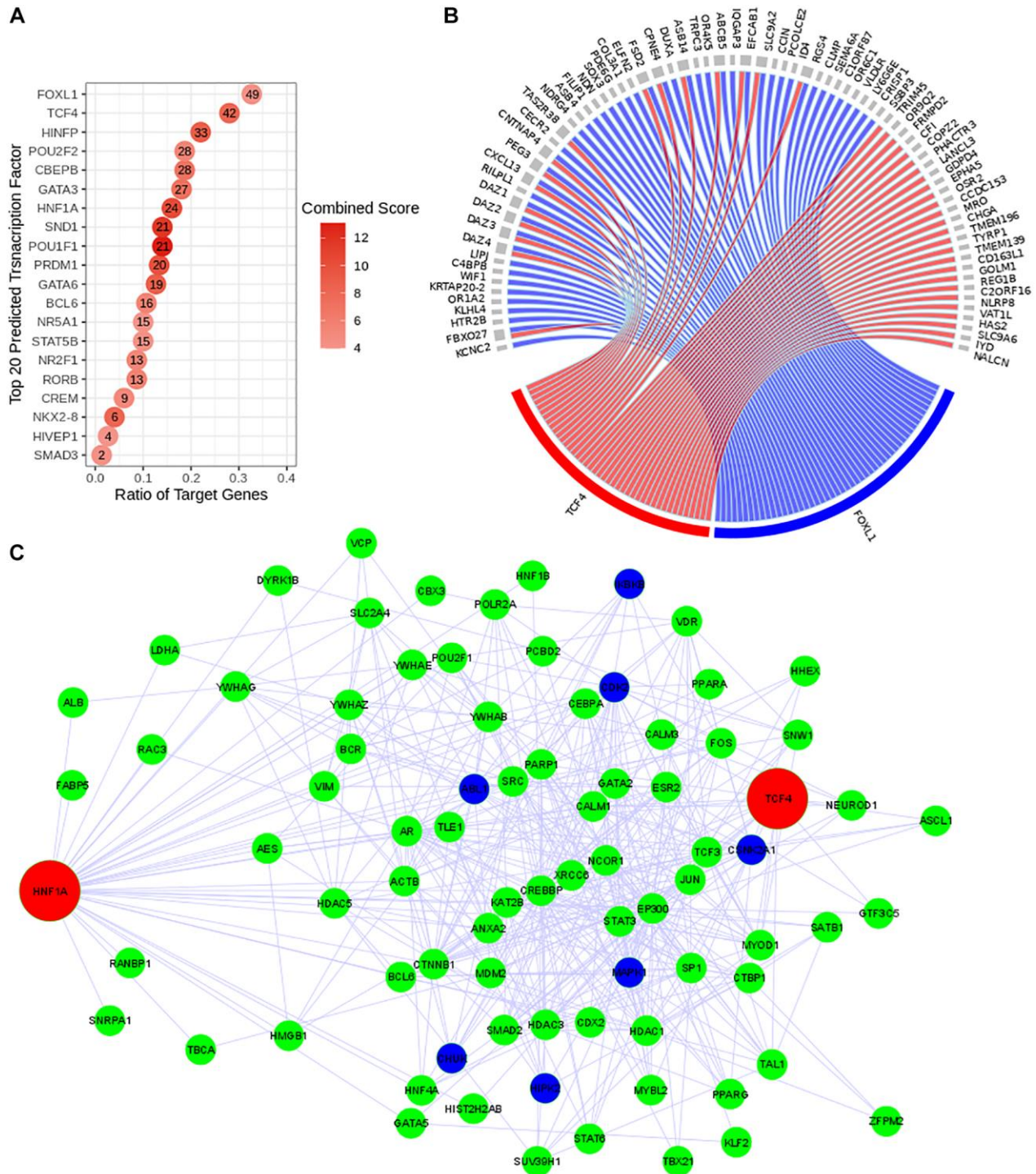


Figure 3. Reconstruction of the predicted transcriptional regulatory network. TF regulatory network was reconstructed using Expression2Kinase tool. (A) Top 20 predicted TFs were presented and ranked based on their combined scores. The intensity of the red coloration is proportional to the combined scores. The ratio of the target genes (x axis) indicates the proportion of genes targeted by a determined TF. (B) The circos diagram shows the interaction between the top two predicted TFs (*TCF4* in red and *FOXL1* in blue), ranked by the ratio of target genes) and their targeted genes (gray). Both TFs target 16 common genes. (C) Co-regulatory network of *TCF4* and *HNF1A*. The transcriptional regulatory network is presented with the transcription factors (TFs) in red, the intermediate protein that are predicted to interact with these TFs in green and the kinases in dark blue.

DISCUSSION

Age-related LMM is a multiple factorial geriatric condition that is associated with known age-related conditions [5]. Despite its prevalence in older population and its impact on patients' quality of life, cellular and biological processes associated with LMM are not fully investigated. This highlights the need for studies addressing in more detail the systemic molecular alterations in this condition to identify possible therapeutic targets. By performing transcriptomic analysis in elderly women with LMM, we were able to identify genes that discriminate LMM from matched controls and showed that alterations of muscle development, protein metabolism pathways, immune system and inflammation are involved in the pathogenesis of LMM. In addition, weighted gene co-expression network analysis revealed a correlation of gene module involved in immunity pathways.

Using human transcriptome microarray and a bioinformatic approach based on logistic regression

analysis, we identified 821 transcript clusters (565 with positive coefficients (β coefficient >1 and p -value <0.01) and 255 with negative coefficients (β coefficient <-1 and p -value <0.01)) that discriminate LMM patients from age and race-matched controls. We took advantage of the performance of a machine learning method (Support Vector Machine) to further filter this list of predictors. Support Vector Machine identified 45 classifiers of LMM with an accuracy 0.7. This approach has already been used in other scenario to classifier complex diseases [13, 14] based on gene expression profile due to its effectiveness and best performance with high dimensional data such as transcriptomic data.

We observed that the identified genes by logistic regression are predominantly associated with muscle development and metabolic pathways (Figure 2A). In addition, human phenotypes associated with immune alteration and inflammation are also overrepresented (Figure 2B). These results reflect well-described perturbations associated with aging [15, 16] and have already associated with sarcopenia process and progress

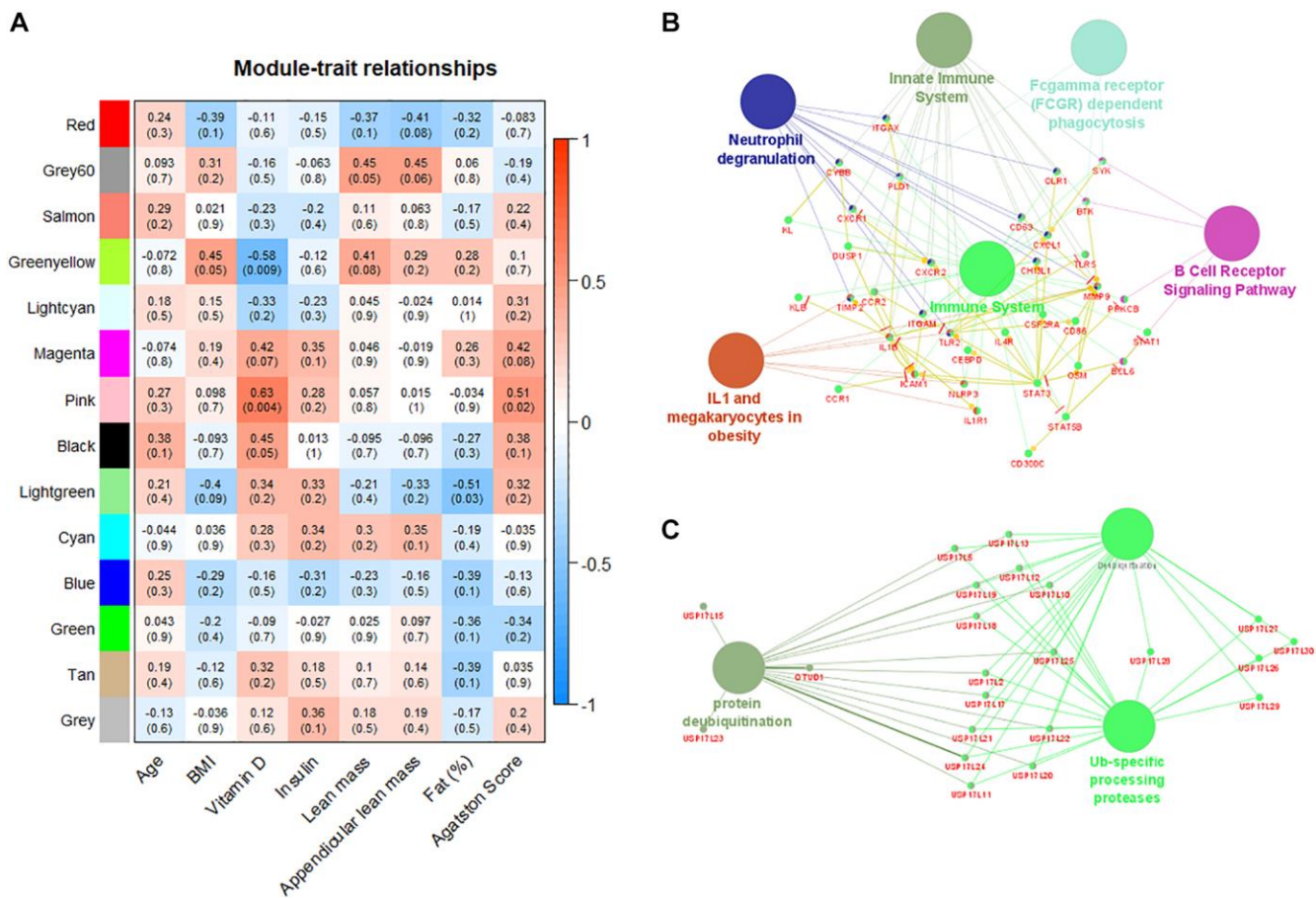


Figure 4. Weighted gene co-expression network analysis reveals modules correlated with patient's characteristics. (A) Correlation matrix of detected gene modules with the characteristics of patients. Red color indicates positive correlation and blue color indicates negative correlation. Pink module is positively correlated with plasma levels of vitamin D and Agatston score. **(B)** Pink module is enriched with immune system pathways. **(C)** Green-yellow module is enriched with pathways associated with ubiquitination processes.

in several recent studies [6, 17]. Among the top ranked genes, there are two transcription factors (TF) *FOXB1* and *HNF1A* that can be involved in the regulation of key mechanisms underlying the process of LMM. Of note, using a regulatory network reconstruction approach, we identified a co-regulation network of *HNF1A* (Figure 3C) that is a TF with a wide range of functions [18] and *TCF4* (a predicted TF). Together these TFs target 16 common genes identified in our study. In pancreas β -islet cells, *HNF1A* contributes to insulin secretion in response to glucose and is associated with maturity-onset diabetes of the young-like diabetes [19]. In our study we observed a low serum level of insulin in LMM group (11.6 vs. 17; $p = 0.02$). *TCF4* has been described as a key regulator of myogenesis [20]. However, their role in peripheral blood cells is uncertain. Thus, further detailed studies are necessary to investigate its contribution to LMM processes and progress.

Although cellular and molecular mechanisms associated with LMM are not understood, inflammatory alterations, immune senescence and protein metabolism have been suggested to be involved in this process [17, 21]. Ageing is generally associated with altered inflammation with increased levels of several pro-inflammatory markers such as CRP, TNF- α , IL-6 and IL-1 [22, 23]. Moreover, cross sectional studies have suggested the increase of IL-6 and TNF- α in sarcopenia [24]. In our study, the analysis of human phenotypes associated with LMM reveals great contribution of immune system and inflammation perturbations (Figure 2B). In addition, network analysis identified key hub genes (*CLCNKB*, *KDSR*, *KIAA0586*, *OCRL*, *FANCM*, *TGDS*, *NDN*) that make an inter-connection between most of the identified human phenotypes, suggesting their prominent role in the progression of LMM and a cross talk between immune-inflammation and muscle mass decline.

Moreover, we identified by weighted gene co-expression network analysis gene modules correlated with patients' relevant clinical and laboratory characteristics. This approach is complementary to the classical differential expression analysis and have the advantage to associate gene expression profiles with patients data, improving the interpretability of the finding [25, 26]. The weighted gene co-expression analysis identified 14 correlated modules (Figure 4A). A gene module involving immune system alterations was positively correlated with serum levels of vitamin D ($r = 0.63$, p -value = 0.004) and Agatston score ($r = 0.51$, p -value = 0.02). Epidemiological studies have shown that a poor vitamin D status is associated with an increased risk of several diseases, including autoimmune diseases [27]. Of note, a lower serum level vitamin D (25-OH) was observed in

the group of LMM women (18.75 vs. 24.15 ng/ml, $p = 0.018$), supporting the deterioration of the interplay between skeletal muscle and the immune system in this population [28]. An animal model suggests that vitamin D deficiency induces muscle waste and muscle protein degradation in ubiquitin proteasome pathway-dependent manner [29]. Interestingly, two gene modules negatively correlated with vitamin D serum levels and body fat percentage ($r = -0.58$, p -value = 0.009 and -0.51 , p -value = 0.05 respectively) are also associated with ubiquitination mechanisms, confirming the critical participation of protein metabolism pathways in the progression of sarcopenia [30, 31].

This study has a main limitation worth noting. Our analysis included a relatively small sample, limiting the extrapolation of the obtained results to the general population of older women. However these participants were recruited from a highly admixed population [32] suggesting the presence of a diversified genetic background.

Our study provides new insights into the biological mechanisms and key mediators that can be involved in the development and/or progression of LMM in elderly women and highlights the contribution of vitamin D deficiency to age-related conditions. We identified lists of discriminating transcripts and weighted co-expression gene modules that can be involved in the pathogenesis of LMM by combining logistic regression analysis, a machine learning approach and network analysis. Finally, our results revealed that the well characterized association between inflammation, immunity and metabolic alterations and age-related muscle waste is also consistently observed at transcriptomic level. Noteworthy, our predicted models do not intend to highlight a relation of causality between the identified molecular interactions/mechanisms and the development of LMM. Instead, they provide a global view and new hypothesis-generating dataset, that could be explored in future investigations. To gain more insights into the mechanisms and regulatory interactions upstream to LMM development and progression, multiomics integrative approaches must also be considered in future studies.

METHODS

Subjects and microarray dataset

The subjects of this study were selected from the population-based survey (São Paulo Ageing and Health [SPAH] study) followed by the bone metabolism outpatient clinic of the Faculty of Medicine of the University of Sao Paulo (FMUSP). Transcriptome analysis of a total of 90 elderly women from this

population were performed and described in our previous studies [33, 34] using microarray technology. From this microarray dataset publicly deposited in the GEO NCBI database under the accession number GSE152073, we selected 40 elderly women, of which 20 were with LMM (defined by a Newman's residual <-1.32) and 20 age and race-matched controls (residual >-1.32). Clinical and laboratory characteristics of this cohort were extracted from our database. Race was defined based on the self-reported race of second-generation ancestors, an approach previously used for the Brazilian population [35]. Physical activity was classified as (a) low, when not even housework is performed; (b) moderate, for regular housework, non-regular walking, gardening; and (c) high, for regular housework and daily regular physical activity at least twice a week for 30 min [36]. This research was approved by the local Ethics in Research Committee of the Medicine Faculty of São Paulo University/Brazil.

Transcriptomic analysis

Microarray pre-processing and analysis

Microarray dataset was extracted from the repository São Paulo Aging and Health Study (SPAH), our previously published database [33, 34]. Raw microarray datasets were deposited in GEO NCBI repository at the accession number GSE152073. Normalization was performed by the Robust Multichip Average (RMA) method [37] using Transcriptome Analysis Console (TAC) software. Probes are summarized into a single probe set corresponding to a single transcript cluster. Then probe sets were annotated using biomaRt package [38] in R environment. Considering the binary variable Low of Muscle Mass (LMM), we adjusted a logistic regression model with age as a covariate in R to discriminate genes that affect this dependent variable [34]. Genes with regression coefficients $\beta \geq 1$ or $\beta \leq -1$ and $p \leq 0.01$ were considered significant. In this approach a positive coefficient reflects an association with LMM. These discriminating variables were called predictor with positive relationship (PPR) in this study). Negative value reflects an association with LMM absence (predictor with negative relationship [PNR]). Additionally, a differential expression analysis was also performed using limma package to identify differentially expressed (DE) genes (transcript clusters) in the comparison of LMM vs. control group. We also applied the Empirical Bayes framework to estimate the more precise expression of each transcript cluster. All analyses were performed with R (version 3.7).

Functional gene set analysis

To facilitate the interpretation of the biological relevance of discriminating genes selected from the

logistic regression adjustment, a functional gene set analysis (GSA) was performed using ClueGO [39]. ClueGO, a plug-in of Cytoscape [40], predicts the functional gene ontology terms and biological pathways associated with our genes list. Then, it organizes the detected pathways in functionally grouped networks to highlight the relevant biological relationship between them. We applied a fusion criterion to reduce the redundancy of the terms that have similar associated gene sets. To further filter the most important terms and pathways, two-sided hyper-geometric distribution tests was used and terms at a significant level Bonferroni adjusted p -value of ≤ 0.05 were selected. To gain more insights into the human phenotype previously associated with the gene sets, a single-sample GSEA (ssGSEA) was also performed with GenePattern [12, 41] using the Human Phenotype Ontology (HPO) database [42]. Differentially activated HPO were then identified by comparing ssGSA score between LMM and control using RankProd package [43].

Regulatory gene network prediction

To refine the exploratory analysis, the prediction of the regulatory networks of transcription factors (TFs) and kinases underlying gene expression of the PPR gene list was performed using Expression2Kinases (X2K) tool [44]. Only the top 20 list of human TFs and kinases ranked based on p -value were selected. Transcriptional regulatory network was visualized using Cytoscape software [40].

Predictive model of LMM using support vector machine (SVM)

We submitted the discriminating transcript clusters identified by logistic regression to Recursive Feature Elimination (RFE) algorithm using e1071 and caret package. RFE was applied to the same subset of microarray dataset included in the logistic regression analysis (20 LMM and 20 controls) to select the most informative features (transcript clusters) for the SVM model training. The Training cohort was consisted of 70% of the SPAH dataset randomly selected and the remaining sample (test cohort) was used to validate the model. SVM model was trained with a polynomial kernel of degree 1. Model performance was evaluated based on its accuracy.

Weighted gene co-expression network analysis (WGCNA)

Gene co-expression analysis was performed using WGCNA method [45] as previously described elsewhere [26]. We included in this analysis the 20 LMM patients previously analyzed in the logistic regression analysis. To

construct the weighted gene co-expression networks, WGCNA determines the connection strengths between nodes (protein coding genes) by computing the pairwise correlations between their expression profiles and using a soft thresholding of the Pearson correlation. Modules of the weighted gene co-expression network constituted of groups of highly correlated genes were identified and correlated with relevant laboratorial and clinical data of the patients. Module highly correlated with patient's data were selected for further gene set enrichment analysis using ClueGO [39] plugin in Cytoscape [40].

AUTHOR CONTRIBUTIONS

Study design: LHJN, BWB, NHML, ACP and RMRP; Study conduct: LHJN, BWB and RMRP; Data collection: LHJN, GHF, VFC and LT; Bioinformatics analysis: BWB. Data interpretation: LHJN, BWB, ACP and RMRP; Drafting manuscript: LHJN, BWB, ACP and RMRP; Revising manuscript: LHJN, BWB, ACP and RMRP; Approving final version of manuscript: all authors.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

This study was financially supported by the Sao Paulo Research Foundation (FAPESP) grant #2016/00006-7 and #2020/14223-5; Conselho Nacional de Ciencia e Tecnologia (CNPQ) grant #45759020130; and Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior (CAPES) Finance Code001.

REFERENCES

1. Newman AB, Kupelian V, Visser M, Simonsick E, Goodpaster B, Nevitt M, Kritchevsky SB, Tylavsky FA, Rubin SM, Harris TB, and Health ABC Study Investigators. Sarcopenia: alternative definitions and associations with lower extremity function. *J Am Geriatr Soc.* 2003; 51:1602–09.
<https://doi.org/10.1046/j.1532-5415.2003.51534.x>
PMID:14687390
2. Cruz-Jentoft AJ, Sayer AA. Sarcopenia. *Lancet.* 2019; 393:2636–46.
[https://doi.org/10.1016/S0140-6736\(19\)31138-9](https://doi.org/10.1016/S0140-6736(19)31138-9)
PMID:31171417
3. Domiciano DS, Figueiredo CP, Lopes JB, Caparbo VF, Takayama L, Menezes PR, Bonfa E, Pereira RM. Discriminating sarcopenia in community-dwelling older women with high frequency of overweight/obesity: the

São Paulo Ageing & Health Study (SPAH). *Osteoporos Int.* 2013; 24:595–603.

<https://doi.org/10.1007/s00198-012-2002-1>
PMID:22588184

4. de Santana FM, Domiciano DS, Gonçalves MA, Machado LG, Figueiredo CP, Lopes JB, Caparbo VF, Takayama L, Menezes PR, Pereira RM. Association of Appendicular Lean Mass, and Subcutaneous and Visceral Adipose Tissue With Mortality in Older Brazilians: The São Paulo Ageing & Health Study. *J Bone Miner Res.* 2019; 34:1264–74.
<https://doi.org/10.1002/jbmr.3710>
PMID:30866105
5. Liguori I, Russo G, Bulli G, Curcio F, Flocco V, Galizia G, Della-Morte D, Gargiulo G, Testa G, Cacciatore F, Bonaduce D, Abete P. Validation of "(fr)AGILE": a quick tool to identify multidimensional frailty in the elderly. *BMC Geriatr.* 2020; 20:375.
<https://doi.org/10.1186/s12877-020-01788-1>
PMID:32993569
6. Dalle S, Rossmeislova L, Koppo K. The Role of Inflammation in Age-Related Sarcopenia. *Front Physiol.* 2017; 8:1045.
<https://doi.org/10.3389/fphys.2017.01045>
PMID:29311975
7. Morton RW, Murphy KT, McKellar SR, Schoenfeld BJ, Henselmans M, Helms E, Aragon AA, Devries MC, Banfield L, Krieger JW, Phillips SM. A systematic review, meta-analysis and meta-regression of the effect of protein supplementation on resistance training-induced gains in muscle mass and strength in healthy adults. *Br J Sports Med.* 2018; 52:376–84.
<https://doi.org/10.1136/bjsports-2017-097608>
PMID:28698222
8. Dhillon RJ, Hasni S. Pathogenesis and Management of Sarcopenia. *Clin Geriatr Med.* 2017; 33:17–26.
<https://doi.org/10.1016/j.cger.2016.08.002>
PMID:27886695
9. Pratt J, Boreham C, Ennis S, Ryan AW, De Vito G. Genetic Associations with Aging Muscle: A Systematic Review. *Cells.* 2019; 9:12.
<https://doi.org/10.3390/cells9010012>
PMID:31861518
10. Hunt LC, Schadeberg B, Stover J, Haugen B, Pagala V, Wang YD, Puglise J, Barton ER, Peng J, Demontis F. Antagonistic control of myofiber size and muscle protein quality control by the ubiquitin ligase UBR4 during aging. *Nat Commun.* 2021; 12:1418.
<https://doi.org/10.1038/s41467-021-21738-8>
PMID:33658508
11. Migliavacca E, Tay SKH, Patel HP, Sonntag T, Civiletto G, McFarlane C, Forrester T, Barton SJ, Leow MK,

- Antoun E, Charpagne A, Seng Chong Y, Descombes P, et al. Mitochondrial oxidative capacity and NAD⁺ biosynthesis are reduced in human sarcopenia across ethnicities. *Nat Commun.* 2019; 10:5808.
<https://doi.org/10.1038/s41467-019-13694-1>
PMID:[31862890](https://pubmed.ncbi.nlm.nih.gov/31862890/)
12. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet.* 2006; 38:500–01.
<https://doi.org/10.1038/ng0506-500>
PMID:[16642009](https://pubmed.ncbi.nlm.nih.gov/16642009/)
13. Hounkpe BW, Benatti RO, Carvalho BS, De Paula EV. Identification of common and divergent gene expression signatures in patients with venous and arterial thrombosis using data from public repositories. *PLoS One.* 2020; 15:e0235501.
<https://doi.org/10.1371/journal.pone.0235501>
PMID:[32780732](https://pubmed.ncbi.nlm.nih.gov/32780732/)
14. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer Genomics Proteomics.* 2018; 15:41–51.
<https://doi.org/10.21873/cgp.20063>
PMID:[29275361](https://pubmed.ncbi.nlm.nih.gov/29275361/)
15. Finkel T. The metabolic regulation of aging. *Nat Med.* 2015; 21:1416–23.
<https://doi.org/10.1038/nm.3998>
PMID:[26646498](https://pubmed.ncbi.nlm.nih.gov/26646498/)
16. Catic A. Cellular Metabolism and Aging. *Prog Mol Biol Transl Sci.* 2018; 155:85–107.
<https://doi.org/10.1016/bs.pmbts.2017.12.003>
PMID:[29653684](https://pubmed.ncbi.nlm.nih.gov/29653684/)
17. Ogawa S, Yakabe M, Akishita M. Age-related sarcopenia and its pathophysiological bases. *Inflamm Regen.* 2016; 36:17.
<https://doi.org/10.1186/s41232-016-0022-5>
PMID:[29259690](https://pubmed.ncbi.nlm.nih.gov/29259690/)
18. Armendariz AD, Krauss RM. Hepatic nuclear factor 1-alpha: inflammation, genetics, and atherosclerosis. *Curr Opin Lipidol.* 2009; 20:106–11.
<https://doi.org/10.1097/mol.0b013e3283295ee9>
PMID:[19280766](https://pubmed.ncbi.nlm.nih.gov/19280766/)
19. Beyssel S, Eyerci N, Pinarli FA, Kizilgul M, Ozelik O, Caliskan M, Cakal E. HNF1A gene p.I27L is associated with early-onset, maturity-onset diabetes of the young-like diabetes in Turkey. *BMC Endocr Disord.* 2019; 19:51.
<https://doi.org/10.1186/s12902-019-0375-2>
PMID:[31109344](https://pubmed.ncbi.nlm.nih.gov/31109344/)
20. Mathew SJ, Hansen JM, Merrell AJ, Murphy MM, Lawson JA, Hutcheson DA, Hansen MS, Angus-Hill M, Kardon G. Connective tissue fibroblasts and Tcf4 regulate myogenesis. *Development.* 2011; 138:371–84.
<https://doi.org/10.1242/dev.057463>
PMID:[21177349](https://pubmed.ncbi.nlm.nih.gov/21177349/)
21. Kim TN, Choi KM. Sarcopenia: definition, epidemiology, and pathophysiology. *J Bone Metab.* 2013; 20:1–10.
<https://doi.org/10.11005/jbm.2013.20.1.1>
PMID:[24524049](https://pubmed.ncbi.nlm.nih.gov/24524049/)
22. Bruunsgaard H, Skinhøj P, Pedersen AN, Schroll M, Pedersen BK. Ageing, tumour necrosis factor-alpha (TNF-alpha) and atherosclerosis. *Clin Exp Immunol.* 2000; 121:255–60.
<https://doi.org/10.1046/j.1365-2249.2000.01281.x>
PMID:[10931139](https://pubmed.ncbi.nlm.nih.gov/10931139/)
23. Ferrucci L, Corsi A, Lauretani F, Bandinelli S, Bartali B, Taub DD, Guralnik JM, Longo DL. The origins of age-related proinflammatory state. *Blood.* 2005; 105:2294–99.
<https://doi.org/10.1182/blood-2004-07-2599>
PMID:[15572589](https://pubmed.ncbi.nlm.nih.gov/15572589/)
24. Bian AL, Hu HY, Rong YD, Wang J, Wang JX, Zhou XZ. A study on relationship between elderly sarcopenia and inflammatory factors IL-6 and TNF-α. *Eur J Med Res.* 2017; 22:25.
<https://doi.org/10.1186/s40001-017-0266-9>
PMID:[28701179](https://pubmed.ncbi.nlm.nih.gov/28701179/)
25. Ma C, Lv Q, Teng S, Yu Y, Niu K, Yi C. Identifying key genes in rheumatoid arthritis by weighted gene co-expression network analysis. *Int J Rheum Dis.* 2017; 20:971–79.
<https://doi.org/10.1111/1756-185X.13063>
PMID:[28440025](https://pubmed.ncbi.nlm.nih.gov/28440025/)
26. Zhang B, Horvath S. Ridge regression based hybrid genetic algorithms for multi-locus quantitative trait mapping. *Int J Bioinform Res Appl.* 2005; 1:261–72.
<https://doi.org/10.1504/IJBRA.2005.007905>
PMID:[18048135](https://pubmed.ncbi.nlm.nih.gov/18048135/)
27. Remelli F, Vitali A, Zurlo A, Volpato S. Vitamin D Deficiency and Sarcopenia in Older Persons. *Nutrients.* 2019; 11:2861.
<https://doi.org/10.3390/nu11122861>
PMID:[31766576](https://pubmed.ncbi.nlm.nih.gov/31766576/)
28. Nelke C, Dziewas R, Minnerup J, Meuth SG, Ruck T. Skeletal muscle as potential central link between sarcopenia and immune senescence. *EBioMedicine.* 2019; 49:381–88.
<https://doi.org/10.1016/j.ebiom.2019.10.034>
PMID:[31662290](https://pubmed.ncbi.nlm.nih.gov/31662290/)
29. Bhat M, Kalam R, Qadri SS, Madabushi S, Ismail A. Vitamin D deficiency-induced muscle wasting occurs through the ubiquitin proteasome pathway and is

- partially corrected by calcium in male rats. *Endocrinology*. 2013; 154:4018–29.
<https://doi.org/10.1210/en.2013-1369>
PMID:23928374
30. Altun M, Besche HC, Overkleeft HS, Piccirillo R, Edelmann MJ, Kessler BM, Goldberg AL, Ulfhake B. Muscle wasting in aged, sarcopenic rats is associated with enhanced activity of the ubiquitin proteasome pathway. *J Biol Chem*. 2010; 285:39597–608.
<https://doi.org/10.1074/jbc.M110.129718>
PMID:20940294
31. Xia Z, Cholewa J, Zhao Y, Shang HY, Yang YQ, Araújo Pessôa K, Su QS, Lima-Soares F, Zanchi NE. Targeting Inflammation and Downstream Protein Metabolism in Sarcopenia: A Brief Up-Dated Description of Concurrent Exercise and Leucine-Based Multimodal Intervention. *Front Physiol*. 2017; 8:434.
<https://doi.org/10.3389/fphys.2017.00434>
PMID:28690550
32. Naslavsky MS, Scliar MO, Nunes K, Wang JYT, Yamamoto GL, Guio H, Tarazona-Santos E, Duarte YAO, Passos-Bueno MR, Meyer D, Zatz M. Biased pathogenic assertions of loss of function variants challenge molecular diagnosis of admixed individuals. *Am J Med Genet C Semin Med Genet*. 2021. [Epub ahead of print].
<https://doi.org/10.1002/ajmg.c.31931>
PMID:34189818
33. Wicik Z, Jales Neto LH, Guzman LEF, Pavão R, Takayama L, Caparbo VF, Lopes NHM, Pereira AC, Pereira RMR. The crosstalk between bone metabolism, lncRNAs, microRNAs and mRNAs in coronary artery calcification. *Genomics*. 2021; 113:503–13.
<https://doi.org/10.1016/j.ygeno.2020.09.041>
PMID:32971215
34. Jales Neto LH, Wicik Z, Torres GHF, Takayama L, Caparbo VF, Lopes NHM, Pereira AC, Pereira RMR. Overexpression of SNTG2, TRAF3IP2, and ITGA6 transcripts is associated with osteoporotic vertebral fracture in elderly women from community. *Mol Genet Genomic Med*. 2020; 8:e1391.
<https://doi.org/10.1002/mgg3.1391>
PMID:32602654
35. Lopes JB, Figueiredo CP, Caparbo VF, Takayama L, Menezes PR, Scazufca M, Pereira RM. Osteoporotic fractures in the Brazilian community-dwelling elderly: prevalence and risk factors. *J Clin Densitom*. 2011; 14:359–66.
<https://doi.org/10.1016/j.jocd.2011.04.006>
PMID:21652251
36. Lopes JB, Danilevicius CF, Takayama L, Caparbo VF, Menezes PR, Scazufca M, Kuroishi ME, Pereira RM. Prevalence and risk factors of radiographic vertebral fracture in Brazilian community-dwelling elderly. *Osteoporos Int*. 2011; 22:711–19.
<https://doi.org/10.1007/s00198-010-1258-6>
PMID:20442985
37. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003; 19:185–93.
<https://doi.org/10.1093/bioinformatics/19.2.185>
PMID:12538238
38. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*. 2009; 4:1184–91.
<https://doi.org/10.1038/nprot.2009.97>
PMID:19617889
39. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009; 25:1091–93.
<https://doi.org/10.1093/bioinformatics/btp101>
PMID:19237447
40. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003; 13:2498–504.
<https://doi.org/10.1101/gr.1239303>
PMID:14597658
41. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011; 27:1739–40.
<https://doi.org/10.1093/bioinformatics/btr260>
PMID:21546393
42. Shefchek KA, Harris NL, Gargano M, Matentzoglou N, Unni D, Brush M, Keith D, Conlin T, Vasilevsky N, Zhang XA, Balhoff JP, Babb L, Bello SM, et al. The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res*. 2020; 48:D704–15.
<https://doi.org/10.1093/nar/gkz997>
PMID:31701156
43. Del Carratore F, Jankevics A, Eisinga R, Heskes T, Hong F, Breitling R. RankProd 2.0: a refactored bioconductor package for detecting differentially expressed features in molecular profiling datasets. *Bioinformatics*. 2017; 33:2774–75.
<https://doi.org/10.1093/bioinformatics/btx292>

PMID:[28481966](#)

44. Chen EY, Xu H, Gordonov S, Lim MP, Perkins MH, Ma'ayan A. Expression2Kinases: mRNA profiling linked to multiple upstream regulatory layers. *Bioinformatics*. 2012; 28:105–11.

<https://doi.org/10.1093/bioinformatics/btr625>

PMID:[22080467](#)

45. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008; 9:559.

<https://doi.org/10.1186/1471-2105-9-559>

PMID:[19114008](#)

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1 to 3.

Supplementary Table 1. Transcript clusters discriminated by logistic regression.

Supplementary Table 2. Differentially expressed transcript clusters.

Supplementary Table 3. Classifiers identified by support vector machine.