

## Original Article

# Potentially overlapping DNA methylation-associated genes between atherosclerosis and aging in humans

Jingyu Zhang<sup>1\*</sup>, Yue Gu<sup>2\*</sup>, Jiemei Yang<sup>3\*</sup>, Yongli Yang<sup>4</sup>, Lei Xu<sup>1</sup>, Ye Tian<sup>4</sup>, Yunming Zhang<sup>1</sup>

<sup>1</sup>Division of Gerontology, The First Affiliated Hospital of Harbin Medical University, Harbin 150081, Heilongjiang, China; <sup>2</sup>College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, Heilongjiang, China; <sup>3</sup>Division of Cardiac Ultrasound, The First Affiliated Hospital of Harbin Medical University, Harbin 150081, Heilongjiang, China; <sup>4</sup>Division of Cardiology, The First Affiliated Hospital of Harbin Medical University, Harbin 150081, Heilongjiang, China. \*Equal contributors.

Received January 17, 2019; Accepted September 2, 2019; Epub October 15, 2019; Published October 30, 2019

**Abstract:** DNA methylation is an important epigenetic modification method associated with disease and aging. In clinical practice, patients with atherosclerosis are generally elderly. We intend to use DNA methylation status as an intermediate to relate aging with atherosclerosis. We acquired DNA methylation data both in arteriosclerotic patients and normal samples ranging in age from 22 to 85 years old from Gene Expression Omnibus (GEO). By these data, the differentially expressed age-associated genes and atherosclerosis-associated genes were identified. A human protein interaction network (PPI) was constructed by the weight of DNA methylation regarding aging and atherosclerosis. The functional modules were further mined from this network. After a biological function enrichment analysis of the modules, a total of 11 DNA methylation-associated genes, including VEGFA, CBX5, EPS15, PLAT, ACTA2, FMOD, TGFB3, CTGF, LAMB2, FBLN1 and WWP2, were found to likely play overlapping roles in both aging and atherosclerosis. In conclusion, the screened DNA methylation-associated genes may be potential candidates involved in aging-related atherosclerosis.

**Keywords:** DNA methylation, atherosclerosis, aging, bioinformatics

## Introduction

Atherosclerosis, often considered as a disease of aging, is characterized by cellular senescence and is impacted by several factors. There are multiple complications associated with atherosclerosis that might be among the main causes of human death and disability over the next 20 years [1]. Therefore, it is necessary to identify age-related gene markers of atherosclerosis.

Aging is a complex process affected by many factors, including telomere shortening, the accumulation of genetic variations, oxidative stress, and cell and organ decay [2]. With age, the risk of multiple diseases, such as cancer, cardiovascular disease and neurodegenerative disease all increase. Unfortunately, due to limitations of current genetic research, human aging is still not completely understood. The emergence of epigenetics provides a new direc-

tion to solve this critical problem. Increasing evidence reveals that CpG methylation plays an important role in gene silencing [3]. Epigenetic analysis has become an important prediction tool applied not only to cancer research but also in neurological and cardiovascular diseases [4-6]. DNA methylation, one of the most frequently studied epigenetic markers, can regulate the expression of genes. Similar to atherosclerosis, DNA methylation is affected by environmental and other external factors, such as aging [7].

Currently, it is known that dynamic changes in DNA methylation exert important effects on both cancer and aging. Similarly, aging is considered one of the most prevalent risk factors for developing cancer, cardiovascular disease and neurodegenerative disease. Meanwhile, DNA methylation is considered to play a key role in atherosclerosis; in which several genes with hypo- or hyper-methylation in their respec-

**Table 1.** Database information

GEO series	Atherosclerosis related data	Age related data
GSE69270	-	184
GSE72773	-	310
GSE72775	-	335
GSE72777	-	46
GSE46394	34	15

tive promoters have been detected, compared with healthy controls [8]. Therefore, we hypothesized that DNA methylation may be associated with aging as well as the occurrence and development of atherosclerosis. The present study aimed to identify genetic markers associated with aging and atherosclerosis by DNA methylation profiling, to enhance disease prediction in clinic.

## Materials and methods

### Materials

Two Illumina Human Methylation 450 BeadChip derived DNA methylation data sets were downloaded from the Gene Expression Omnibus (GEO) database. One data set analyzed blood samples from healthy individuals; the other was comprised of samples from patients with carotid atherosclerosis, including 34 tumor and 15 normal tissue samples (Table 1).

### Data preprocessing

First, the four age related data sets were merged, removing sites on SNPs and sex chromosomes. Next, the R package was used to fill in the missing values by the k-Nearest Neighbor (KNN) method, which finds the nearest k neighbor by using the Euclidean distance  $d(x,y)$  according to the KNN algorithm, and sets the median to fill in the missing values. We compute this test  $d(x,y)$  by using the following formula (1):

$$d(x,y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (1)$$

The R package *sva* was used to remove the batch effects of multiple data sets by identifying and estimating surrogate variables for unknown sources of variation in high-through-

put experiments. The same procedure was performed for the carotid atherosclerosis samples. The study workflow is shown in Figure 1.

### Identification of differential DNA methylated genes associated with aging and atherosclerosis

The *samr* package in R was used to screen the different DNA methylation sites between healthy individuals and patients with carotid atherosclerosis, based on the gene chip significance analysis method (SAM) [9]. Algorithm SAM used relative difference ( $d_i$ ) to judge the degree of difference between the two types of samples. We compute this test  $d_i$  by using the following formula (2):

$$d_i = \frac{\bar{x}_{1i} - \bar{x}_{2i}}{s_i + s_0} \quad (2)$$

where  $\bar{x}_{1i}$  and  $\bar{x}_{2i}$  are the mean value of the expression of gene  $i$  in two classes of samples,  $s_i$  is the sum of variance within the sample class, and is  $s_0$  a correction factor.

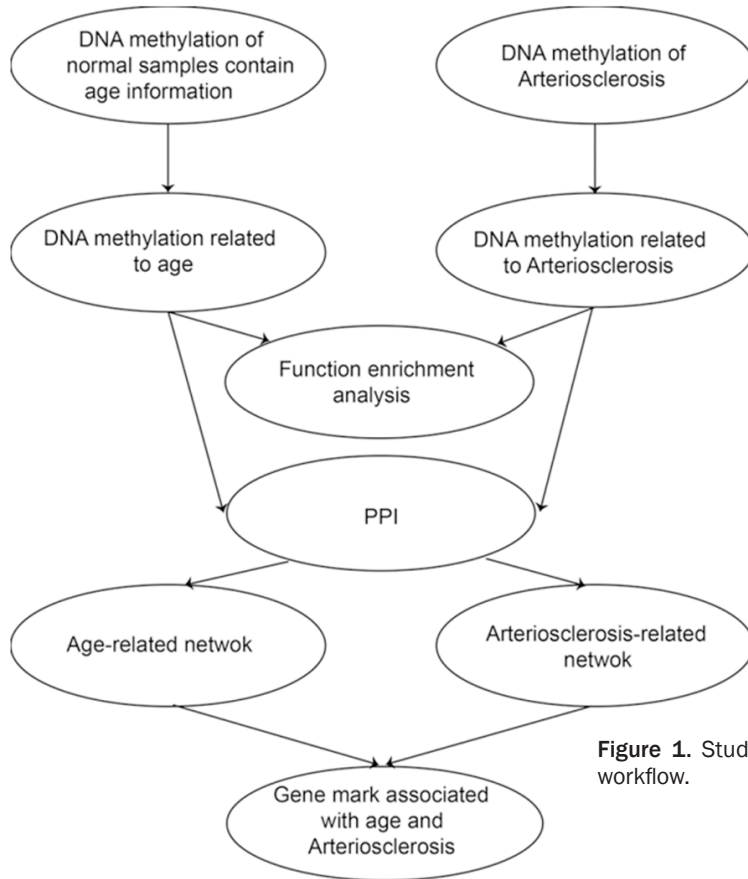
Then, the Benjamini-Hochberg (BH) method was employed for multiple testing adjustment (FDR<0.01). Fold change was used for further screening, with cutoff values greater than 2 or below 0.5. Finally, the corresponding genes with differential DNA methylation based on the GPL13534 platform were identified.

Age-related DNA methylation data show a linear distribution [10-12]. Therefore, we used the *lm* function in R, taking age as the dependent variable and the gene's DNA methylation status as the independent variable. To obtain age related genes with differential DNA methylation, the Benjamini-Hochberg (BH) method in multiple testing adjustment (FDR<0.01) was used. Then, the age related genes with differential DNA methylation were determined.

### Establishment of weighted networks

Weighted networks for age and atherosclerosis were established as following. First, PPI data were downloaded from five PPI network databases, namely HPRD, IntAct, DIP, MINT and BIND, and combined to form the human protein interaction network. Secondly, with the human PPI as the background network, one step networks of genes related with age and atherosclerosis were extracted. In these two net-

## DNA methylation gene overlap as and aging



**Figure 1.** Study workflow.

works, a specific DNA methylated gene and the interaction represented the node and edge, respectively. For the DNA methylation status of the interaction gene in each sample, we found the Pearson correlation coefficient ( $r$ ) by using the following formula (3):

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} \quad (3)$$

The weights were determined by  $|r|$ , the absolute value of Pearson correlation coefficient between both nodes.

### Module mining and function enrichment analysis

Algorithm Cluster ONE is a network clustering algorithm, which can handle weighted networks and overlapping clusters, obtaining modules within a given network. Algorithm Cluster ONE (Clustering with Overlapping Neighborhood Expansion) was used to mine modules in networks related with age and atherosclerosis

[13]. For the two networks, modules with  $p < 0.05$  were selected. Enrichment analysis was performed on genes overlapping between the mined modules in the two related networks ( $p < 0.01$ ). The obtained genes were considered functional genes associated with both age and atherosclerosis.

### Results

#### Genes associated with aging based on DNA methylation profiles

The DNA methylation data contained age information, and a total of 875 normal samples from individuals aged between 22 and 85 years were reanalyzed. After preprocessing, the linear correlation model was utilized to predict age associated genes. Finally, a total of 8761 genes related to aging ( $FDR < 0.01$ ) were screened.

#### Genes associated with atherosclerosis based on DNA methylation profiles

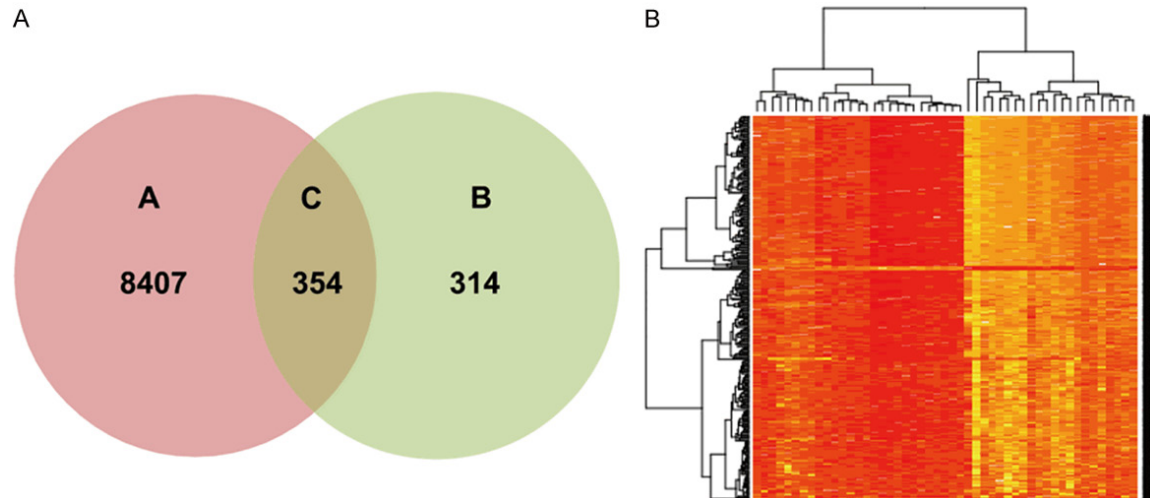
By using SAM and fold change we determined genes with differential DNA methylation status in atherosclerosis (assessing 34 diseased and 15 normal samples), 668 genes were identified (Supplementary Table 1). Then, by intersecting with genes associated with aging, a total of 354 genes were retained for further analysis (Figure 2A, Supplementary Table 2).

Using the hierarchical clustering method, we demonstrated that the DNA methylation values of age-related genes were significantly lower in samples from patients with atherosclerosis, compared with those of normal samples.

#### Functional enrichment analysis of differential genes

Functional enrichment analysis was performed on atherosclerosis-associated genes, age-associated genes, and genes related to both age and atherosclerosis in DAVID (Figure 3). Functional enrichment analysis of age-related genes showed that most genes with differential

## DNA methylation gene overlap as and aging



**Figure 2.** Selected genes associated with aging and atherosclerosis. A. Venn diagram showing the numbers of age- and atherosclerosis-associated genes. A and B are numbers of aging and atherosclerosis-associated genes, respectively. C is the number of overlapping genes. B. Heat map of age-associated differentially expressed genes in atherosclerosis. Rows and columns represent genes and samples, respectively. Yellow and orange represent low and high DNA methylation levels, respectively.

DNA methylation were involved in cell cycle regulation and system development. Enriched genes associated with atherosclerosis were involved in extracellular matrix organization, blood circulation and angiogenesis. The enriched genes associated with both age and atherosclerosis are shown in **Figure 3**. They were mainly involved in skeletal system development, blood vessel growth, angiogenesis and vasculature development. All these genes were closely linked to age and atherosclerosis.

### *Establishment of the weighted network*

The background network was merged from five human protein databases, with 80,980 node pairs interacting with each other. Age-related differential genes were introduced into the network to form an age-dependent network containing 4,638 nodes and 19,820 edges. Atherosclerosis-associated differential genes were introduced into the network to form the atherosclerosis network with 120 nodes and 107 edges. The weight was determined by the absolute value ( $|r|$ ) of the Pearson correlation coefficient between two nodes. Then, selecting edges with nodes showing  $p < 0.05$ , weighted networks associated with age and atherosclerosis were formed (**Figure 4**). The mean degree and clustering coefficient of the weighted networks associated with age were 4.264 and 0.338, respectively; those associated with ath-

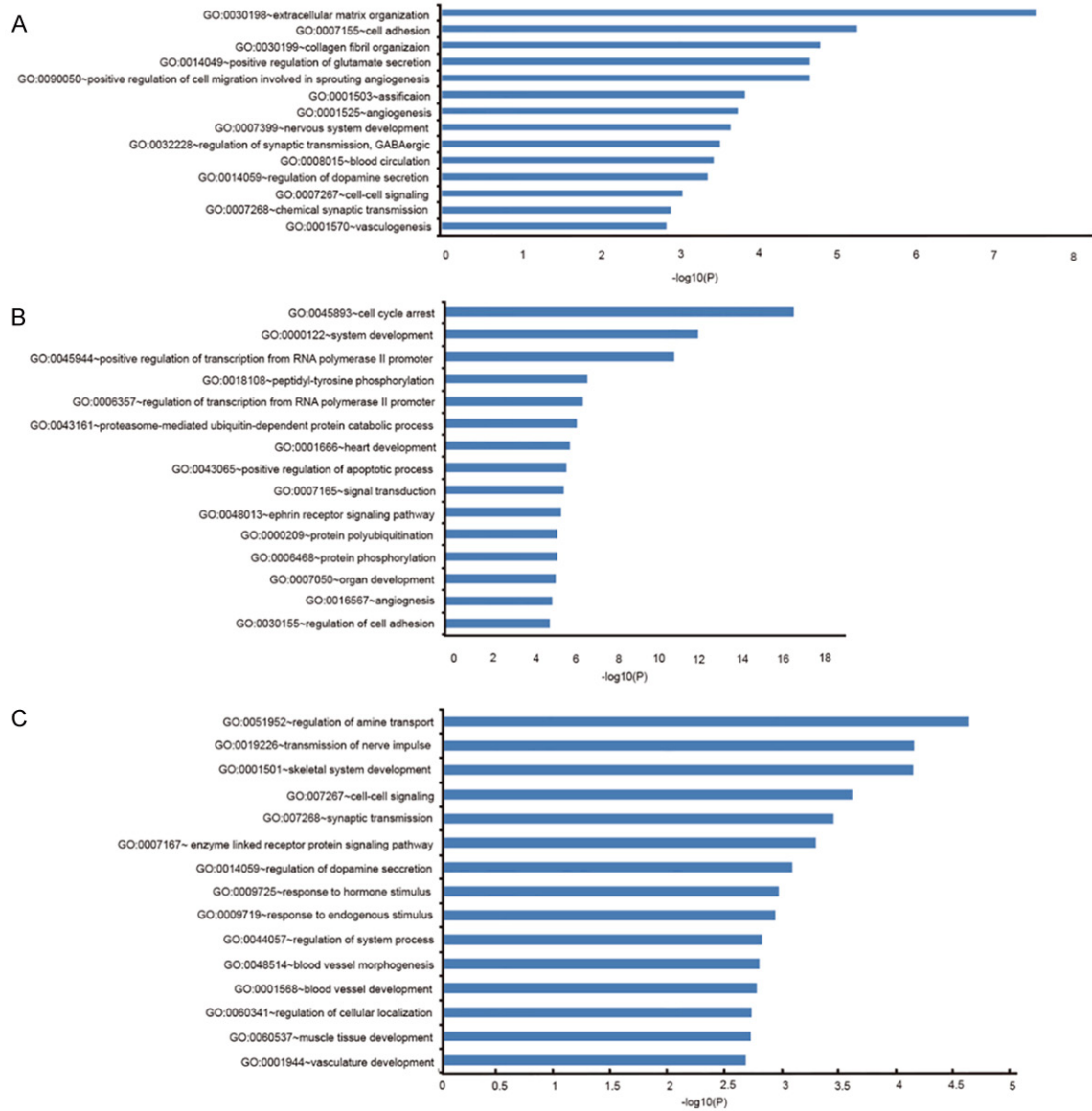
erosclerosis had values of 2.053 and 0.029, respectively. In these two weighted networks associated with age and atherosclerosis, the majority of the nodes had a low degree, and only a small number showed a large degree; namely, the two weighted networks followed the power-law distribution model.

### *Modules in weighted networks associated with age and atherosclerosis, and enrichment*

Cluster ONE was used to mine modules in the two networks that were associated with age and atherosclerosis. In these two networks, modules with  $p < 0.05$  were selected, i.e. 22 and 7 modules in age- and atherosclerosis-related networks, respectively. Then, the degree of enrichment of the modules in both weighted networks was determined by hypergeometric distribution ( $p < 0.01$ ). After identifying the overlapping modules, the associated genes were considered as the functional genes related to both age and atherosclerosis.

Four genes (VEGFA, CBX5, EPS15 and PLAT) in two or more module pairs were found (green points in **Figure 4**); while 7 others (ACTA2, FMOD, TGFB3, CTGF, LAMB2, FBLN1 and WWP2, **Figure 4**) were identified in one overlapping module. As shown in **Figure 5A**, EPS15, CBX5, PLAT, VEGFA, WWP2 and ACTA had lower methylation levels in normal samples com-

## DNA methylation gene overlap as and aging



**Figure 3.** Enrichment analysis of age- and atherosclerosis-associated genes. A. Age-associated genes. B. Atherosclerosis-associated genes. C. Genes associated with both age and atherosclerosis.

pared with those from patients with atherosclerosis. Meanwhile, FBLN1, CTGF, LAMB2, FMOD and TGFB3 in normal and atherosclerotic samples all had low methylation levels.

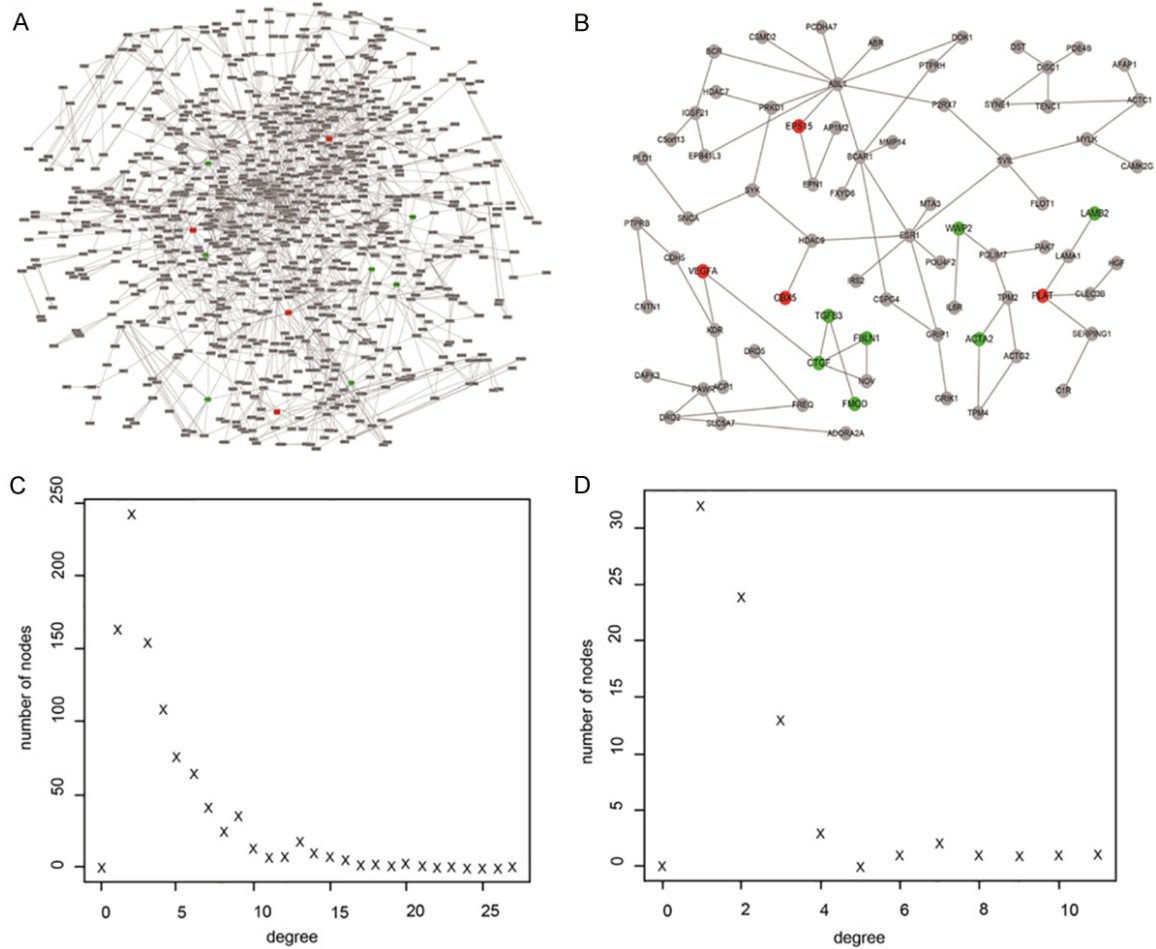
### Discussion

Aging is the most acknowledged cause of atherosclerosis. Age-related atherosclerosis is highly prevalent in the elderly. We identified overlapping genes with similar DNA methylation between aging and atherosclerosis using stepwise biostatistics. A total of 11 methylated genes, including VEGFA, CBX5, EPS15,

PLAT, ACTA2, FMOD, TGFB3, CTGF, LAMB2, FBLN1 and WWP2, were associated with both atherosclerosis and aging.

The present study showed that most age-associated gene clusters with DNA methylation were involved in cell cycle regulation and system development. It is well documented that extracellular matrix components are involved in cardiovascular atherosclerosis [14]. Meanwhile, the extracellular matrix has been found to be closely related to aging [15]. Therefore, this study indicated that aging-accumulated atherosclerosis was entangled with cellular DNA

## DNA methylation gene overlap as and aging



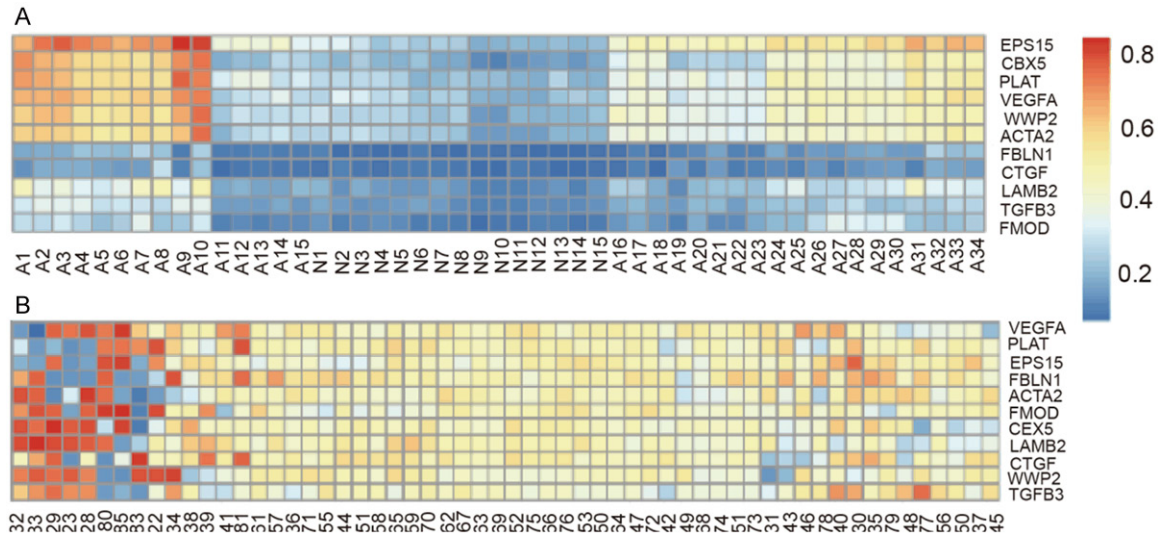
**Figure 4.** Age- and atherosclerosis-associated weight-networks. A. Age-associated weighted network. B. Atherosclerosis-associated weighted network. Green points represent genes in two or more pairs of modules at the same time (VEGFA, CBX5, EPS15 and PLAT). Red points are genes in one overlapping module (ACTA2, FMOD, TGFB3, CTGF, LAMB2, FBLN1 and WWP2). C. Degree distribution in the age-associated weighted network. D. Degree distribution in the atherosclerosis-associated weighted network.

methylation. Functional clustering analysis indicated that candidate genes of atherosclerosis were related to extracellular matrix organization, blood circulation and angiogenesis. Cellular changes of degenerative epithelial or endothelial vascular cells in atherosclerosis are influenced by epigenetic modifications, telomerase shortening, and mitochondrial dysfunction [16, 17]. Therefore, DNA damage is considered one of the cellular bases of atherosclerotic phenotypes [18]. Moreover, the identified methylation-related genes were associated with cardiac development and/or degeneration according to the above functional analysis. Intriguingly, disparity of DNA methylation status was shown between elderly and atherosclerotic patients. Six genes (EPS15, CBX5, PLAT, VEGFA, WWP2

and ACTA2) showed hyper-methylation only in patients with atherosclerosis; whereas five genes (FBLN1, CTGF, LAMB2, FMOD and TGFB3) were hypo-methylated in both atherosclerotic patients and healthy controls. This phenomenon indicated that baseline DNA methylation may occur during general ageing.

Among the identified overlapping genes that were altered both in aging and atherosclerosis, VEGFA was one of the most remarkable, intensively focusing on vascular performance [19]. As the top molecular determinant of vascular growth and hemostasis, VEGFA was methylated during aging-related atherosclerosis. Moreover, seven genes, including CBX5, PLAT, EPS15, TGFB3, CTGF, LAMB2 and FBLN1, have been

## DNA methylation gene overlap as and aging



**Figure 5.** Heat map of DNA methylation levels of 11 genes associated with aging and atherosclerosis. A. Heat map showing DNA methylation levels of 11 genes associated with aging and atherosclerosis in age-associated samples. B. Heat map of DNA methylation levels of 11 genes associated with aging and atherosclerosis in atherosclerosis-associated samples.

previously shown to play distinct roles in vascular differentiation and aging [14, 20-25]. How these methylation-involved genes interact in an integrative network requires cellular validation in further investigation.

The present study was not without limitations. First, patients with atherosclerosis and aging, had enormous heterogeneity, and were not further stratified, which resulted in loose representation. Secondly, functional validation at the cellular level has not been carried out.

In conclusion, eleven genes, including VEGFA, CBX5, EPS15, PLAT, ACTA2, FMOD, TGFB3, CTGF, LAMB2, FBLN1 and WWP2, profoundly involved in DNA methylation, and were differentially expressed in both atherosclerosis and aging. The potential molecular network integrating the overlapping genes could help decipher the mechanism of methylation-based aging-related atherosclerosis.

### Acknowledgements

The study was supported by the Heilongjiang Provincial High School Basic scientific research business fees research projects (2017LCZX34) and the Heilongjiang Provincial Health and Family Planning Commission Scientific Research Project (2016-019).

### Disclosure of conflict of interest

None.

**Address correspondence to:** Ye Tian, Division of Cardiology, The First Affiliated Hospital of Harbin Medical University, No. 23 Postal Street, Nangang District, Harbin 150081, Heilongjiang, China. Tel: +86-13766815997; Fax: +86-21-57643271; E-mail: yetian1@hotmail.com; Yunming Zhang, Division of Gerontology, The First Affiliated Hospital of Harbin Medical University, No. 23 Postal Street, Nangang District, Harbin 150081, Heilongjiang, China. E-mail: zhangyunming@hrbmu.edu.cn

### References

- [1] Mathers CD and Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006; 3: e442.
- [2] Portela A and Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol* 2010; 28: 1057-1068.
- [3] Egger G, Liang G, Aparicio A and Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 2004; 429: 457-463.
- [4] Maeda T, Horiuchi T and Makino N. Epigenetic status of subtelomere of peripheral leukocytes corresponds to cardiographic parameters with a sex association. *Geriatr Gerontol Int* 2018; 18: 1415-1419.
- [5] Huang Q, Xu S, Mo M, Yang Q, Li J, Zhong Y, Zhang J, Zhang L, Ye X, Liu Z and Cai C. Quanti-

## DNA methylation gene overlap as and aging

- fication of DNA methylation and hydroxymethylation in Alzheimer's disease mouse model using LC-MS/MS. *J Mass Spectrom* 2018; 53: 590-594.
- [6] Ge Y, Zhu J, Wang X, Zheng N, Tu C, Qu J and Ren X. Mapping dynamic histone modification patterns during arsenic-induced malignant transformation of human bladder cells. *Toxicol Appl Pharmacol* 2018; 355: 164-173.
- [7] Ihara K, Fuchikami M, Hashizume M, Okada S, Kawai H, Obuchi S, Hirano H, Fujiwara Y, Hachisu M, Hongyong K and Morinobu S. The influence of aging on the methylation status of brain-derived neurotrophic factor gene in blood. *Int J Geriatr Psychiatry* 2018; 33: 1312-1318.
- [8] Xia Z, Gu M, Jia X, Wang X, Wu C, Guo J, Zhang L, Du Y and Wang J. Integrated DNA methylation and gene expression analysis identifies SLAMF7 as a key regulator of atherosclerosis. *Aging (Albany NY)* 2018; 355: 164-173.
- [9] Kasen S, Ouellette R and Cohen P. Mainstreaming and postsecondary educational and employment status of a rubella cohort. *Am Ann Deaf* 1990; 135: 22-26.
- [10] Shirodkar AV and Marsden PA. Epigenetics in cardiovascular disease. *Curr Opin Cardiol* 2011; 26: 209-215.
- [11] Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der Brug M, Chong S, Moore M, Longo DL, Cookson MR, Traynor BJ and Singleton AB. Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum Mol Genet* 2011; 20: 1164-1172.
- [12] Koch CM and Wagner W. Epigenetic-aging-signature to determine age in different tissues. *Aging (Albany NY)* 2011; 3: 1018-1027.
- [13] Nepusz T, Yu H and Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods* 2012; 9: 471-472.
- [14] Chistiakov DA, Sobenin IA and Orekhov AN. Vascular extracellular matrix in atherosclerosis. *Cardiol Rev* 2013; 21: 270-288.
- [15] Rubio-Ruiz ME, Perez-Torres I, Soto ME, Pastelin G and Guarner-Lans V. Aging in blood vessels. Medicinal agents FOR systemic arterial hypertension in the elderly. *Ageing Res Rev* 2014; 18: 132-147.
- [16] Mahmoudi M, Mercer J and Bennett M. DNA damage and repair in atherosclerosis. *Cardiovasc Res* 2006; 71: 259-268.
- [17] Sobenin IA, Zhelankin AV, Sinyov VV, Bobryshev YV and Orekhov AN. Mitochondrial aging: focus on mitochondrial DNA damage in atherosclerosis-a mini-review. *Gerontology* 2015; 61: 343-349.
- [18] Ishida T, Ishida M, Tashiro S, Yoshizumi M and Kihara Y. Role of DNA damage in cardiovascular disease. *Circ J* 2014; 78: 42-50.
- [19] Buroker NE. VEGFA rSNPs, transcriptional factor binding sites and human disease. *J Physiol Sci* 2014; 64: 73-76.
- [20] Zhang X, Nie S, Si X, Luo Y and Tang W. Association between the CTGF -945C/G polymorphism and systemic sclerosis: a meta-analysis. *Gene* 2012; 509: 1-6.
- [21] Regalado ES, Guo DC, Prakash S, Benseid TA, Flynn K, Estrera A, Safi H, Liang D, Hyland J, Child A, Arno G, Boileau C, Jondeau G, Braverman A, Moran R, Morisaki T, Morisaki H, Montalcino Aortic C, Pyeritz R, Coselli J, LeMaire S and Milewicz DM. Aortic disease presentation and outcome associated with ACTA2 mutations. *Circ Cardiovasc Genet* 2015; 8: 457-464.
- [22] Nataatmadja M, West J and West M. Overexpression of transforming growth factor-beta is associated with increased hyaluronan content and impairment of repair in Marfan syndrome aortic aneurysm. *Circulation* 2006; 114: 1371-377.
- [23] Maeng YS, Kwon JY, Kim EK and Kwon YG. Heterochromatin protein 1 alpha (HP1alpha: CBX5) is a key regulator in differentiation of endothelial progenitor cells to endothelial cells. *Stem Cells* 2015; 33: 1512-1522.
- [24] Corella D and Ordovas JM. Aging and cardiovascular diseases: the role of gene-diet interactions. *Ageing Res Rev* 2014; 18: 53-73.
- [25] Adams JN, Cox AJ, Freedman BI, Langefeld CD, Carr JJ and Bowden DW. Genetic analysis of haptoglobin polymorphisms with cardiovascular disease and type 2 diabetes in the Diabetes Heart Study. *Cardiovasc Diabetol* 2013; 12: 31.