



# The Methylation Pattern for Knee and Hip Osteoarthritis

Zhen Wu<sup>1</sup>, Lu Shou<sup>2</sup>, Jian Wang<sup>1</sup>, Tao Huang<sup>3\*</sup> and Xinwei Xu<sup>1\*</sup>

<sup>1</sup> Department of Orthopaedics, Tongde Hospital of Zhejiang Province, Hangzhou, China, <sup>2</sup> Department of Pneumology, Tongde Hospital of Zhejiang Province, Hangzhou, China, <sup>3</sup> Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai, China

## OPEN ACCESS

### Edited by:

Jing Zhang,  
Shanghai Jiao Tong University, China

### Reviewed by:

Jun Xu,  
Shanghai Jiao Tong University, China  
Hongting Jin,  
Zhejiang Chinese Medical University,  
China

### \*Correspondence:

Xinwei Xu  
xuxinwei011@163.com  
Tao Huang  
tohuangtao@126.com

### Specialty section:

This article was submitted to  
Epigenomics and Epigenetics,  
a section of the journal  
Frontiers in Cell and Developmental  
Biology

**Received:** 02 September 2020

**Accepted:** 22 October 2020

**Published:** 06 November 2020

### Citation:

Wu Z, Shou L, Wang J, Huang T  
and Xu X (2020) The Methylation  
Pattern for Knee and Hip  
Osteoarthritis.  
*Front. Cell Dev. Biol.* 8:602024.  
doi: 10.3389/fcell.2020.602024

Osteoarthritis is one of the most prevalent chronic joint diseases for middle-aged and elderly people. But in recent years, the number of young people suffering from the disease increases quickly. It is known that osteoarthritis is a common degenerative disease caused by the combination and interaction of many factors such as natural and environmental factors. DNA methylations reflect the effects of environmental factors. Several researches on DNA methylation at specific genes in OA cartilage indicated the great potential roles of DNA methylation in OA. To systematically investigate the methylation pattern in knee and hip osteoarthritis, we analyzed the methylation profiles in cartilage of 16 OA hip samples, 19 control hip samples and 62 OA knee samples. 12 discriminative methylation sites were identified using advanced minimal Redundancy Maximal Relevance (mRMR) and Incremental Feature Selection (IFS) methods. The SVM classifier of these 12 methylation sites from genes like MEIS1, GABRG3, RXRA, and EN1, can perfectly classify the OA hip samples, control hip samples and OA knee samples evaluated with LOOCV (Leave-One Out-Cross Validation). These 12 methylation sites can not only serve as biomarker, but also provide underlying mechanism of OA.

**Keywords:** osteoarthritis, methylation, Support Vector Machine, minimal Redundancy Maximal Relevance, Incremental Feature Selection

## INTRODUCTION

Osteoarthritis is one of the most prevalent chronic joint diseases, characterized by the loss, degeneration and calcification of articular cartilage (Im and Choi, 2013). It often occurs in middle-aged and elderly people, and the incidence in females is significantly higher than that in males (Rushton et al., 2014). Research shows that in recent years, with the quickening pace of life, the number of young people suffering from the disease began to increase, ultimately resulting in sustained growth in the social burden (den Hollander and Meulenbelt, 2015).

Despite the unremitting efforts of many scholars, the pathogenesis of osteoarthritis is still not clear yet. Current studies indicate that osteoarthritis is a common degenerative disease caused by the combination and interaction of many factors such as natural and environmental factors (Felson, 2004). The increase in age, damage of tissue and cell, obesity, the overuse of joints and genetic susceptibility are known as OA major risk factors (Hunter et al., 2002).

Osteoarthritis mainly involves the cartilage of the weight-bearing joints, followed by the synovium. At present, it is generally believed that the central part of OA mechanism is the degeneration of articular cartilage, resulting from the imbalance between anabolism and catabolism

in the cartilage extracellular matrix caused by mechanical and biological factors (Martel-Pelletier et al., 2008). Since these changes are reported to be caused by genetic changes in the chondrocyte associated with the OA epigenetic mechanism, it's assumed that the epigenetic changes of chondrocytes may be a key driver in osteoarthritis (Reynard and Loughlin, 2012).

Studies on epigenetic of OA have suggested that these mechanisms is very significant during the onset and progression of disease (Barter and Young, 2013). Despite the fact that there exist many epigenetic mechanisms like DNA methylation, miRNA and histone modifications (Aref-Eshghi et al., 2015), the involvement of DNA methylation in OA pathophysiology is the most studied subject (Rice et al., 2020; Sun et al., 2020; Zhou et al., 2020). DNA methylation mainly occurs at CpG dinucleotides, selectively adding a methyl group to cytosine to form 5-methylcytosine under the catalysis of DNA methyltransferase (DNMT). DNA methylation is involved with transcriptional inhibition by preventing the binding of proteins to gene promoters and changing chromatin structure (Loeser, 2008). The changes of methylation status can accelerate the development of OA (Miranda-Duarte, 2018). Thus, further studies on the mechanisms involved in DNA methylation is another way to develop new OA therapy strategies.

There have been several researches on DNA methylation at specific genes in OA cartilage. For instance, the expression of matrix metalloproteinase genes has been reported to be upregulated in OA chondrocytes, resulting in extracellular matrix degradation (Kevorkian et al., 2004). In addition, genes associated with OA chondrocytes like GDF-5, SOX-9, DIO-2, and ADAMTS-4 were also suggested to be differentially expressed between OA cartilage and control group (Reynard and Loughlin, 2012). It has also been reported that the IL1B promoter is demethylated in articular chondrocytes as a response to inflammatory cytokine signaling (Hashimoto et al., 2009).

OA happens in knees and hips mostly. But knee OA is more common than hip OA. Although articular cartilage of hip and knee joints have substantial similar characters and functions, the disease progression and subsequent treatment may be different between the two joints (Pereira et al., 2016). Transcriptomic studies have indicated that genes dysregulated in hip and knee OA have great difference (Loughlin and Reynard, 2015), so does the methylation patterns in hip and knee OA samples (Rushton et al., 2014). These findings highlight the importance of the separation of OA researches from skeletal sites and help us understand the cartilage homeostasis.

The cure for osteoarthritis is mainly to mitigate pain, improve the function of the joints and avoid the side effects of the treatment as far as possible. However, due to the slow progress of osteoarthritis and the lack of sensitive detection methods to identify early OA changes, it is difficult to find disease-modifying drugs at present. Epigenetic markers mentioned above can detect the phenotypes of various chondrocytes, including articular cartilage homeostasis, chondrogenic differentiation and

the development of OA, which may provide new targets and strategies for drug treatment of OA.

To systematically investigate the methylation pattern in knee and hip osteoarthritis, we analyzed the methylation profiles in cartilage of 16 OA hip samples, 19 control hip samples and 62 OA knee samples. 12 discriminative methylation sites were identified based on minimal Redundancy Maximal Relevance (mRMR) and Incremental Feature Selection (IFS). mRMR is a widely used power feature selection method (Chen et al., 2017b, 2018b, 2019c; Cai et al., 2018; Li and Huang, 2018; Li et al., 2019a) which considers not only the relevance with OA status but also the redundancy among methylation status. It can identify a small but well performed methylation signature. What's more, an SVM (Support Vector Machine) (Chen et al., 2017c, 2019a,b,d; Sun et al., 2018; Li et al., 2019b; Pan et al., 2019) OA classifier was built based on these 12 methylation sites and it can perfectly classify the OA hip samples, control hip samples and OA knee samples evaluated with LOOCV (Leave-One Out-Cross Validation) (Chen et al., 2014; Zhang N. et al., 2015; Cheng et al., 2016; Li et al., 2018; Wang and Huang, 2019). Although the model needs to be validated on independent large dataset, these 12 methylation sites provided clues for the mechanisms of OA.

## MATERIALS AND METHODS

### The Cartilage DNA Methylation Profiles of OA Hips, Control Hips and OA Knees

We downloaded the cartilage DNA methylation profiles of 16 OA hip samples, 19 control hip samples and 62 OA knee samples from publicly available GEO (Transcript Expression Omnibus) database under accession number of GSE63695 (Rushton et al., 2014). The DNA methylations were measured using Illumina Human Methylation 450 Array. There were 4,82,421 probes corresponding to methylation sites. The processed beta values that were normalized with preprocess Funnorm function from R package minfi were used for further analysis.

### Identify the Representative OA Methylation Sites

To identify the most discriminative features among different groups, many analysis methods have been developed (Huang et al., 2008; Cai et al., 2010; Zhang et al., 2012, 2016, 2017; Li et al., 2014; Zhang P.W. et al., 2015; Chen et al., 2018a; Wang et al., 2018). ANOVA (Analysis of Variance) is an obvious choice. But such statistical methods don't consider the relationship between features, therefore a lot of redundant features will be selected.

In our study, the number of features, i.e., methylation sites, was extremely large. Obviously, many of them were redundant. To select the representative features, we adopted mRMR method developed by Peng et al. (2005) to reduce redundancy of selected genes. This method has been widely used and proven to be very effective in handling high dimensional data (Niu et al., 2013; Zhao et al., 2013; Zhou et al., 2015; Zhang et al., 2016; Li and Huang, 2017; Liu et al., 2017). The C++ version

**Abbreviations:** mRMR, minimal Redundancy Maximal Relevance; SVM, Support Vector Machine; IFS, Incremental Feature Selection; DNMT, DNA methyltransferase; LOOCV, Leave-One Out-Cross Validation; GEO, Transcript Expression Omnibus.

mRMR program downloaded from <http://home.penglab.com/proj/mRMR/> was used in this study.

Its idea is simple and clear. Let us use  $\Omega$  to denote all the 4,82,421 methylation sites,  $\Omega_s$  to denote the selected  $m$  sites, and  $\Omega_g$  to denote the  $n$  sites to be selected.

First, we evaluated the relevance of site  $g$  from  $\Omega_g$  with three-class sample labels  $l$  (OA hips, control hips and OA knees) was calculated with mutual information ( $I$ ) equation (Sun et al., 2012; Huang and Cai, 2013):

$$I(g, l) \quad (1)$$

Meanwhile, the redundancy of site  $g$  with the  $m$  selected sites  $g_i$  ( $i = 1, 2, \dots, m$ ) in  $\Omega_s$  can also be calculated based on mutual information:

$$\frac{1}{m} \left( \sum_{g_i \in \Omega_s} I(g, g_i) \right) \quad (2)$$

The goal can be characterized as maximizing the function which balances the relevance and redundancy:

$$\max_{g_j \in \Omega_g} \left[ I(g_j, l) - \frac{1}{m} \left( \sum_{g_i \in \Omega_s} I(g_j, g_i) \right) \right] \quad (j = 1, 2, \dots, n) \quad (3)$$

Each time, one best site that maximized this function will be moved from  $\Omega_g$  to  $\Omega_s$ . Eventually, all the sites will be ranked base on their relevance with sample labels and redundancy between each other. The ranked methylation sites from large to small can be represented as:

$$S = \{g'_1, g'_2, \dots, g'_r, \dots, g'_N\} \quad (4)$$

The top ranked methylation sites have better discriminative ability.

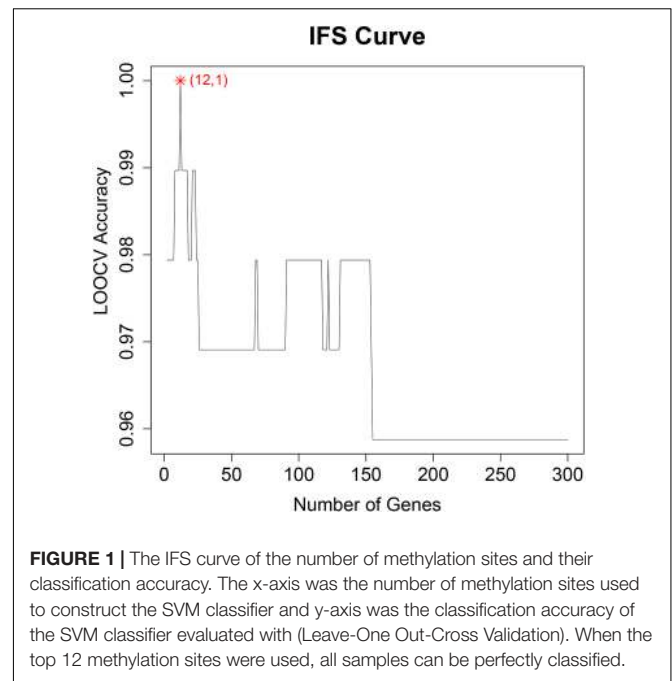
To reduce the computational complexity, we focused on the top 300 mRMR sites for further analysis which should be enough to classify the samples and suitable as biomarkers.

## Optimize the Discriminative Methylation Sites for OA

Although we identified the non-redundant OA sites using mRMR method, we still wanted to obtain the methylation site combinations which can classify the OA hips, control hips and OA knees. To do so, we applied a widely used optimization method, IFS (Jiang et al., 2013; Li et al., 2014; Shu et al., 2014; Zhang et al., 2014; Huang et al., 2015; Zhang P.W. et al., 2015; Chen et al., 2017a).

Based on the ranked mRMR site list, each time, the top  $r$  sites  $\{g'_1, g'_2, \dots, g'_r\}$  were chosen to construct a SVM (Support Vector Machine) classifier and its accuracy evaluated with LOOCV (Chen et al., 2014; Zhang N. et al., 2015; Cheng et al., 2016; Li et al., 2018; Wang and Huang, 2019) was recorded. SVM is classical machine learning classifier with a wide range of applications in biomedicine (Chen et al., 2017c, 2019b; Li et al., 2018, 2019b; Sun et al., 2018; Pan et al., 2019). The R function `svm` in package `e1071`<sup>1</sup> was used to apply the SVM method. LOOCV,

<sup>1</sup><https://CRAN.R-project.org/package=e1071>



**FIGURE 1 |** The IFS curve of the number of methylation sites and their classification accuracy. The x-axis was the number of methylation sites used to construct the SVM classifier and y-axis was the classification accuracy of the SVM classifier evaluated with (Leave-One Out-Cross Validation). When the top 12 methylation sites were used, all samples can be perfectly classified.

as known as Jackknife test, is a widely used objective method to evaluate the prediction performance of classifiers (Chou, 2011). Each time, one sample was treated as test sample while the other samples were used to train the model. At last, each sample had been tested for once and the classes of all samples were predicted. By comparing the predicted classes and actual classes, we can calculate the LOOCV accuracy.

After 300 rounds, the performances of the 300 methylation subsets were tested. By analyzing the number of used methylation sites and the performance of corresponding classifier, we can easily find the best methylation sites.

## RESULTS AND DISCUSSION

### The Representative OA Methylation Sites

In the OA methylation dataset, there were 4,82,421 features. On one hand, the number of features was large; on the other hand, there were redundancy among these methylation sites. To reduce the number of features to a reasonable number for further study, we adopted the mRMR method.

The methylation sites were ranked based on both their relevance with the sample labels, i.e., OA hips, control hips or OA knees, and their redundancy with each other. In other words, only the methylation site that exhibited different pattern with already selected sites will be selected from the candidate methylation site pool. With the mRMR analysis, we identified the top 300 most representative OA methylation sites.

### The Discriminative Methylation Sites for OA

Although the mRMR method considered the relevance between features and sample labels, but it was balanced with redundancy.

**TABLE 1** | The 12 discriminative methylation sites for OA.

Rank	Probe	Chromosome	Coordinate	Gene	mRMR score
1	cg09462924	2	66666470	MEIS1	0.943
2	cg22118147	5	172144013	-	0.263
3	cg18576667	15	27597409	GABRG3	0.328
4	cg07533951	12	114879558	-	0.284
5	cg14545975	9	137297213	FXRA	0.245
6	cg05877497	2	66667946	MEIS1	0.264
7	cg21811143	2	119599748	EN1	0.219
8	cg09989996	1	753376	FAM87B	0.235
9	cg19738283	2	176976802	HOXD10	0.244
10	cg02824888	12	115129011	-	0.226
11	cg04288999	2	66667852	MEIS1	0.238
12	cg00995986	2	66665428	MEIS1	0.214

We would like to find the methylation sites that can best classify different OA samples. These sites not only should be concise, but also have great discriminative ability.

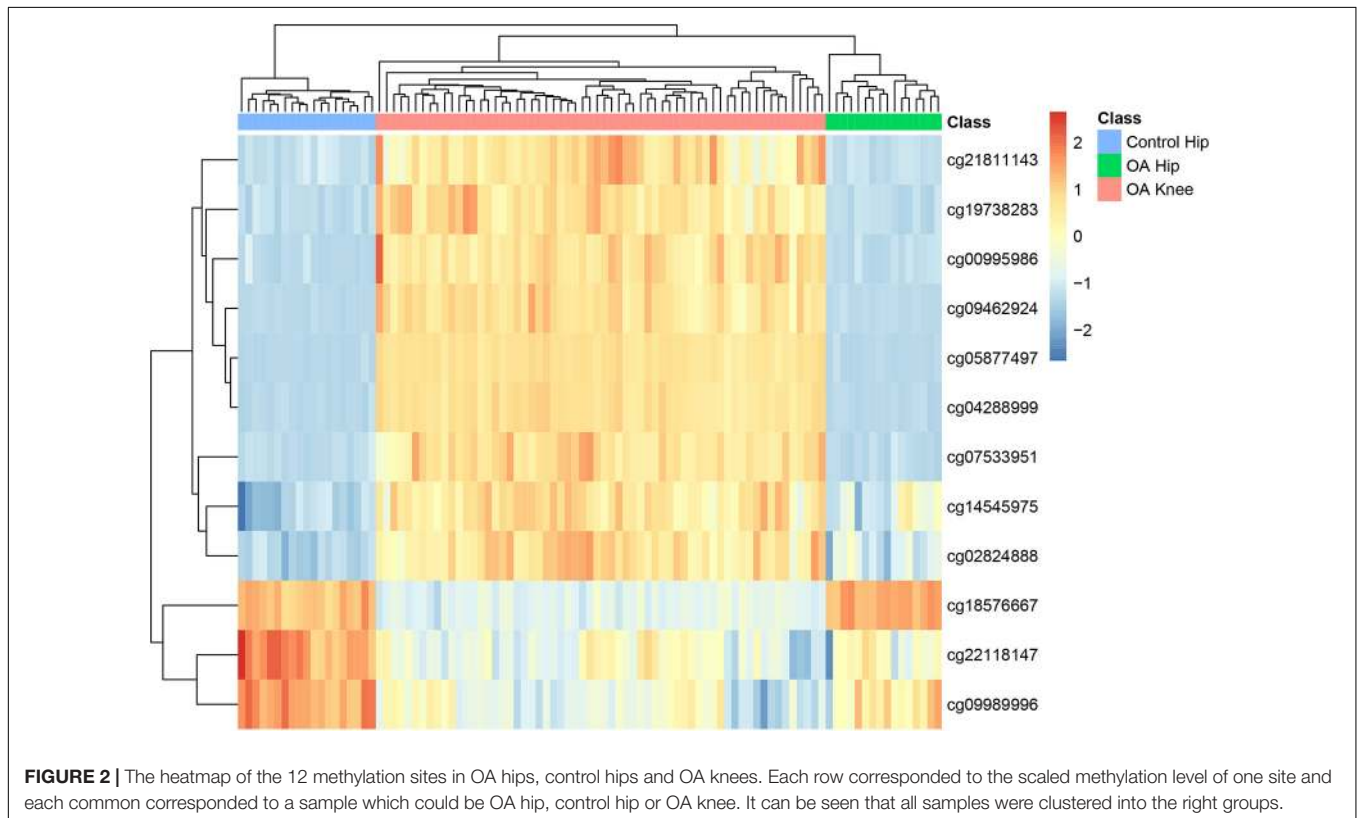
To find the best discriminative methylation sites for OA, we plotted the IFS results in **Figure 1** in which, the *x*-axis was the number of methylation sites and the *y*-axis was the LOOCV accuracy of the SVM classifier based these sites. It can be seen that when the top 12 mRMR sites was used, the ACC was the highest. All samples were correctly classified. The 12 chosen methylation sites were given in **Table 1**. The chromosome positions were from Genome Build 37 (hg19). We must be cautious that this accuracy

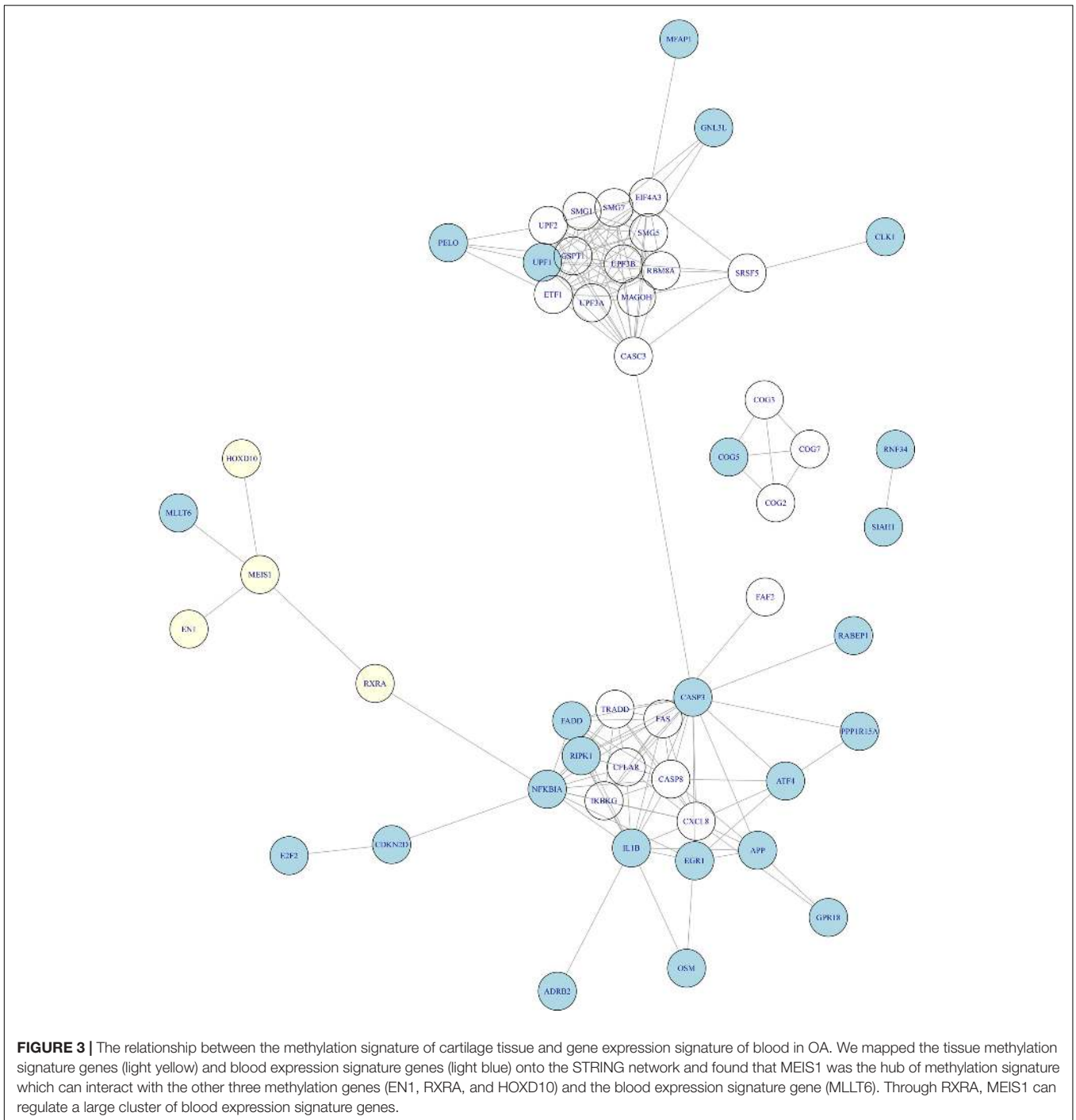
needs to be validated on an independent large OA cohort. But since there is no other similar dataset, we can only try our best and evaluate it with objective LOOCV method. The results provided clues about the difference between OA and control and the difference between hip OA and knee OA and worth to be further investigated.

To explore the methylation levels of the 12 sites in different disease status, we plotted the heatmap of them in all samples in **Figure 2**. It can be seen that OA hip were more similar with OA knee than with control hip. The disease status surpassed the tissue specificity. As for OA hip and OA knee, there were nine highly methylated sites in OA knee (cg21811143, cg19738283, cg00995986, cg09462924, cg05877497, cg04288999, cg07533951, cg14545975, and cg02824888) while there were three highly methylated sites in OA hip (cg18576667, cg22118147, and cg09989996).

### The Relationship Between the Methylation Signature of Cartilage Tissue and Gene Expression Signature of Blood in OA

There have been studies of blood expression profiles of OA (Ramos et al., 2014; Li et al., 2018). Ramos et al. identified 27 blood expression signature genes of OA (Ramos et al., 2014) and Li et al. (2018) identified 23 blood expression signature genes. By combining them, we obtain 46 blood expression signature genes. There was no overlap between the 46 blood





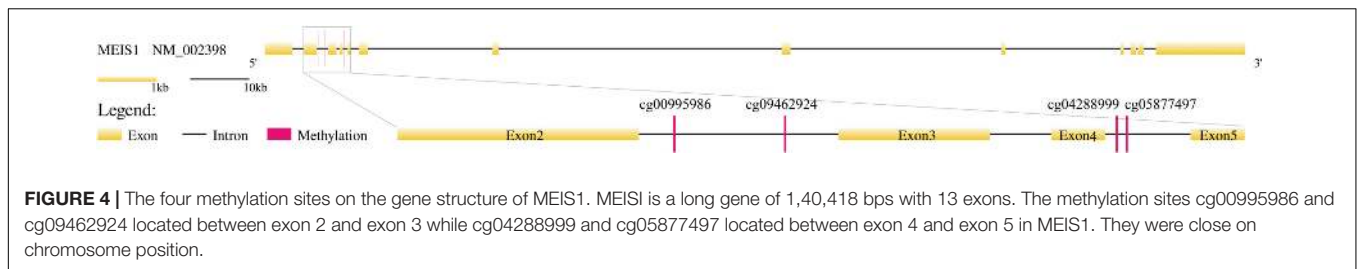
expression signature genes and our cartilage tissue methylation signature genes. But when we mapped them onto STRING network<sup>2</sup> (Szklarczyk et al., 2018), the network relationship between tissue methylation genes and blood expression genes (Figure 3) was clear. It can be seen that MEIS1 was the hub of methylation signature which can interact with the other three methylation genes (EN1, RXRA, and HOXD10) and the

blood expression signature gene (MLLT6). Through RXRA and MEIS1 can regulate a large cluster of blood expression signature genes.

### The Biological Functions of the Discriminative Methylation Sites for OA

In Table 1, we annotated the methylation sites to genes based on their chromosome coordinates. Four of them were corresponding

<sup>2</sup><http://string-db.org>, version: 11.0



to the same gene, MEIS1 (Meis Homeobox 1). We checked the consistency of these four MEIS1 methylation sites in OA hips, control hips and OA knees in **Figure 2** and we found they shared similar pattern although they were not identical. MEIS1 is a long gene of 1,40,418 bps with 13 exons. We mapped the four methylation sites onto the gene structure of MEIS1 (**Figure 4**). The methylation sites cg00995986 and cg09462924 located between exon 2 and exon 3 while cg04288999 and cg05877497 located between exon 4 and exon 5. They were close on chromosome position. There have been reports that MEIS1 was associated with blood and epithelial cancers, such as acute leukemia and skin cancer (Okumura et al., 2014; Wang et al., 2014). As a transcription factor, MEIS1 can bind chromatin DNA and regulate pluripotency of stem cells (Okumura et al., 2014). It is generally considered as an oncogene which accelerate development (Wang et al., 2014). There are no reported associations between MEIS1 and osteoarthritis. MEIS1 may serve as a novel OA biomarker, especially for distinguishing the difference between hip OA and knee OA.

Another interesting methylation site was cg21811143 which locate in EN1 (Engrailed Homeobox 1). It plays a role in controlling development of the central nervous system (CNS) (Webb et al., 2008) and segmentation course, where it is required for the formation of posterior compartments (Lawrence and Johnston, 1984). Recent studies have identified EN1 gene as a decisive factor of bone mineral density (BMD) via whole-genome sequencing (Zheng et al., 2015), while higher BMD was considered to be involved with OA in many cross-sectional and longitudinal epidemiological studies (Belmonte-Serrano et al., 1993). SNP rs4144782 in EN1 was reported to be significantly associated with increased risk of knee OA (OR = 1.26; 95% CI: 1.05–1.50,  $p$ -value = 0.012) (Li et al., 2017). What's more, the EN1 mRNA levels were differentially expressed between Knee and hip Intra-articular adipose tissues (IAATs) and Subcutaneous adipose tissue (SCAT) (Eymard et al., 2017). Its expression was strongly decreased in all IAATs compared with SCAT (Infrapatellar fat pad, IFP: 0.3-fold,  $p$  = 0.006; Suprapatellar fat pad, SPFP: 0.2-fold,  $p$  = 0.006; and Acetabular fat pad, AFP: 0.3-fold,  $p$  = 0.046) (Eymard et al., 2017). The genomic, methylation and transcriptomic evidences of the association between EN1 and OA make it a strong candidate OA gene worth further experimental validation.

GABRG3 and RXRA, were gamma-aminobutyric acid (GABA) receptor and retinoid X receptor, respectively. Beside autistic disorder (Menold et al., 2001) and Alzheimer's disease (Iwakiri et al., 2009), GABRG3 is associated with alcohol

dependence by affecting disinhibition and hyperexcitability of CNS (Dick et al., 2004; Edenberg and Foroud, 2006). Maybe it can also affect how OA patients feel about pain. For RXRA, its methylation status was reported to be associated with childhood bone mineral content (BMC) (Harvey et al., 2014).

The methylation sites that can't be annotated to genes may functions through trans-regulation (van Eijk et al., 2012). Their roles can be investigated with integrative analysis of multi-omics OA data in the future.

Another issue of this work was the small sample size. It will overestimate the prediction performance. The confounding factors and disease heterogeneity will be difficult to detect in such small dataset. The signature needs to be validated on large cohorts from different medical centers.

## CONCLUSION

As a chronic joint disease, osteoarthritis is very common in elder people. Even young people are more and more likely to have OA symptoms due to the quickening pace of life. OA happens in knees and hips mostly, but knee OA is more common than hip OA. The disease progression and treatment of hip and knee OA are different. Since the epigenetic factor plays key roles in OA, we systematically analyzed the DNA methylation profiles of cartilage from 16 OA hip samples, 19 control hip samples and 62 OA knee samples. With advanced feature selection methods, 12 OA discriminative methylation sites were selected from a total of 4,82,421 sites. These sites corresponded to genes like MEIS1, EN1, GABRG3, and RXRA. These results provided not only novel OA biomarkers, but also possible mechanisms that worth further investigation in an independent large cohort.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63695>.

## AUTHOR CONTRIBUTIONS

TH and XX designed the experiment. ZW performed the experiment. LS and JW participant in the data analysis. ZW, TH, and XX wrote the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China (31701151), National Key R&D Program of China (2018YFC0910403),

Shanghai Municipal Science and Technology Major Project (Grant No. 2017SHZDZX01), Shanghai Sailing Program (16YF1413800), and the Youth Innovation Promotion Association of Chinese Academy of Sciences (CAS) (2016245).

## REFERENCES

- Aref-Eshghi, E., Zhang, Y., Liu, M., Harper, P. E., Martin G, Furey, A (2015). Genome-wide DNA methylation study of hip and knee cartilage reveals embryonic organ and skeletal system morphogenesis as major pathways involved in osteoarthritis. *BMC Musculoskelet. Disord.* 16:287. doi: 10.1186/s12891-015-0745-5
- Barter, M. J., and Young, D. A. (2013). Epigenetic mechanisms and non-coding RNAs in osteoarthritis. *Curr. Rheumatol. Rep.* 15:353. doi: 10.1007/s11926-013-0353-z
- Belmonte-Serrano, M. A., Bloch, D. A., Lane, N. E., Michel, B. E., and Fries, J. F. (1993). The relationship between spinal and peripheral osteoarthritis and bone density measurements. *J. Rheumatol.* 20, 1005–1013.
- Cai, L., Huang, T., Su, J., Zhang, X., Chen, W., Zhang, F., et al. (2018). Implications of newly identified brain eQTL genes and their interactors in schizophrenia. *Mol. Ther. Nucleic Acids* 12, 433–442. doi: 10.1016/j.omtn.2018.05.026
- Cai, Y. D., Huang, T., Feng, K. Y., Hu, L., and Xie, L. (2010). A unified 35-gene signature for both subtype classification and survival prediction in diffuse large B-cell lymphomas. *PLoS One* 5:e12726. doi: 10.1371/journal.pone.0012726
- Chen, L., Li, J., Zhang, Y. H., Feng, K., Wang, S., Zhang, Y., et al. (2017a). Identification of gene expression signatures across different types of neural stem cells with the Monte-Carlo feature selection method. *J. Cell. Biochem.* 119, 3394–3403.
- Chen, L., Wang, S., Zhang, Y. H., Li, J., Xing, Z. H., Yang, J., et al. (2017b). Identify key sequence features to improve CRISPR sgRNA efficacy. *IEEE Access* 5, 26582–26590. doi: 10.1109/ACCESS.2017.2775703
- Chen, L., Zhang, Y.-H., Wang, S., Zhang, Y., Huang, T., and Cai, Y.-D. (2017c). Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways. *PLoS One* 12:e0184129. doi: 10.1371/journal.pone.0184129
- Chen, L., Li, J., Zhang, Y. H., Feng, K., Wang, S., Zhang, Y., et al. (2018a). Identification of gene expression signatures across different types of neural stem cells with the Monte-Carlo feature selection method. *J. Cell. Biochem.* 119, 3394–3403. doi: 10.1002/jcb.26507
- Chen, L., Zhang, Y. H., Huang, G., Pan, X., Wang, S., Huang, T., et al. (2018b). Discriminating cirRNAs from other lncRNAs using a hierarchical extreme learning machine (H-ELM) algorithm with feature selection. *Mol. Genet. Genomics* 293, 137–149. doi: 10.1007/s00438-017-1372-7
- Chen, L., Lu, J., Zhang, N., Huang, T., and Cai, Y. D. (2014). A hybrid method for prediction and repositioning of drug anatomical therapeutic chemical classes. *Mol. Biosyst.* 10, 868–877. doi: 10.1039/c3mb70490d
- Chen, L., Pan, X., Zeng, T., Zhang, Y., Huang, T., and Cai, Y. (2019a). Identifying essential signature genes and expression rules associated with distinctive development stages of early embryonic cells. *IEEE Access* 7, 128570–128578. doi: 10.1109/ACCESS.2019.2939556
- Chen, L., Pan, X., Zhang, Y.-H., Kong, X., Huang, T., and Cai, Y.-D. (2019b). Tissue differences revealed by gene expression profiles of various cell lines. *J. Cell. Biochem.* 120, 7068–7081. doi: 10.1002/jcb.27977
- Chen, L., Pan, X., Zhang, Y.-H., Liu, M., Huang, T., and Cai, Y.-D. (2019c). Classification of widely and rarely expressed genes with recurrent neural network. *Comput. Struct. Biotechnol. J.* 17, 49–60. doi: 10.1016/j.csbj.2018.12.002
- Chen, L., Pan, X., Zhang, Y. H., Hu, X., Feng, K., Huang, T., et al. (2019d). Primary tumor site specificity is preserved in patient-derived tumor xenograft models. *Front. Genet.* 10:738. doi: 10.3389/fgene.2019.00738
- Cheng, S., Zhu, C., Chu, C., Huang, T., Kong, X., and Zhu, L. C. (2016). Prediction of bioactive compound pathways using chemical interaction and structural information. *Comb. Chem. High Throughput Screen.* 19, 161–169. doi: 10.2174/1386207319666151110123611
- Chou, K. C. (2011). Some remarks on protein attribute prediction and pseudo amino acid composition. *J. Theor. Biol.* 273, 236–247. doi: 10.1016/j.jtbi.2010.12.024
- den Hollander, W., and Meulenbelt, I. (2015). DNA methylation in osteoarthritis. *Curr. Genomics* 16, 419–426. doi: 10.2174/1389202916666150817212711
- Dick, D. M., Edenberg, H. J., Xuei, X., Goate, A., Kuperman, S., Schuckit, M., et al. (2004). Association of GABRG3 with alcohol dependence. *Alcohol. Clin. Exp. Res.* 28, 4–9. doi: 10.1097/01.alc.0000108645.54345.98
- Edenberg, H. J., and Foroud, T. (2006). The genetics of alcoholism: identifying specific genes through family studies. *Addict. Biol.* 11, 386–396. doi: 10.1111/j.1369-1600.2006.00035.x
- Eymard, F., Pigenet, A., Citadelle, D., Tordjman, J., Foucher, L., Rose, C., et al. (2017). Knee and hip intra-articular adipose tissues (IAATs) compared with autologous subcutaneous adipose tissue: a specific phenotype for a central player in osteoarthritis. *Ann. Rheum. Dis.* 76, 1142–1148. doi: 10.1136/annrheumdis-2016-210478
- Felson, D. T. (2004). An update on the pathogenesis and epidemiology of osteoarthritis. *Radiol. Clin. North Am.* 42, 1–9. doi: 10.1016/s0033-8389(03)00161-1
- Harvey, N. C., Sheppard, A., Godfrey, K. M., McLean, C., Garratt, E., Ntani, G., et al. (2014). Childhood bone mineral content is associated with methylation status of the RXRA promoter at birth. *J. Bone Miner. Res.* 29, 600–607. doi: 10.1002/jbmr.2056
- Hashimoto, K., Oreffo, R. O., Gibson, M. B., Goldring, M. B., and Roach, H. I. (2009). DNA demethylation at specific CpG sites in the IL1B promoter in response to inflammatory cytokines in human articular chondrocytes. *Arthritis Rheum* 60, 3303–3313. doi: 10.1002/art.24882
- Huang, T., and Cai, Y.-D. (2013). An information-theoretic machine learning approach to expression QTL analysis. *PLoS One* 8:e67899. doi: 10.1371/journal.pone.0067899
- Huang, T., Shu, Y., and Cai, Y.-D. (2015). Genetic differences among ethnic groups. *BMC Genomics* 16:1093. doi: 10.1186/s12864-015-2328-0
- Huang, T., Tu, K., Shyr, Y., Wei, C. C., Xie, L., and Li, Y. X. (2008). The prediction of interferon treatment effects based on time series microarray gene expression profiles. *J. Transl. Med.* 6:44. doi: 10.1186/1479-5876-6-44
- Hunter, D. J., March, L., and Sambrook, P. N. (2002). Knee osteoarthritis: the influence of environmental factors. *Clin. Exp. Rheumatol.* 20, 93–100.
- Im, G. I., and Choi, Y. J. (2013). Epigenetics in osteoarthritis and its implication for future therapeutics. *Expert Opin. Biol. Ther.* 13, 713–721. doi: 10.1517/14712598.2013.764410
- Iwakiri, M., Mizukami, K., Ikonovic, M. D., Ishikawa, M., Abrahamson, E. E., DeKosky, S. T., et al. (2009). An immunohistochemical study of GABA A receptor gamma subunits in Alzheimer's disease hippocampus: relationship to neurofibrillary tangle progression. *Neuropathology* 29, 263–269. doi: 10.1111/j.1440-1789.2008.00978.x
- Jiang, Y., Huang, T., Chen, L., Gao, Y. F., Cai, Y., and Chou, K. C. (2013). Signal propagation in protein interaction network during colorectal cancer progression. *Biomed Res. Int.* 2013:287019. doi: 10.1155/2013/287019
- Kevorkian, L., Young, D. A., Darrach, C., Donell, S. T., Shepstone, L., Porter, S., et al. (2004). Expression profiling of metalloproteinases and their inhibitors in cartilage. *Arthritis Rheum* 50, 131–141. doi: 10.1002/art.11433
- Lawrence, P. A., and Johnston, P. (1984). On the role of the engrailed+ gene in the internal organs of *Drosophila*. *EMBO J.* 3, 2839–2844. doi: 10.1002/j.1460-2075.1984.tb02217.x

- Li, B. Q., You, J., Huang, T., and Cai, Y. D. (2014). Classification of non-small cell lung cancer based on copy number alterations. *PLoS One* 9:e88300. doi: 10.1371/journal.pone.0088300
- Li, H., Zhang, X., Cao, Y., Hu, S., Peng, F., Zhou, J., et al. (2017). Association between EN1 rs4144782 and susceptibility of knee osteoarthritis: a case-control study. *Oncotarget* 8, 36650–36657. doi: 10.18632/oncotarget.16842
- Li, J., and Huang, T. (2017). Predicting and analyzing early wake-up associated gene expressions by integrating GWAS and eQTL studies. *Biochim. Biophys. Acta Mol. Basis Dis.* 1864(6 Pt B), 2241–2246. doi: 10.1016/j.bbadis.2017.10.036
- Li, J., and Huang, T. (2018). Predicting and analyzing early wake-up associated gene expressions by integrating GWAS and eQTL studies. *Biochim. Biophys. Acta* 1864(6 Pt B), 2241–2246. doi: 10.1016/j.bbadis.2017.10.036
- Li, J., Lan, C.-N., Kong, Y., Feng, S.-S., and Huang, T. (2018). Identification and analysis of blood gene expression signature for osteoarthritis with advanced feature selection methods. *Front. Genet.* 9:246. doi: 10.3389/fgene.2018.00246
- Li, J., Lu, L., Zhang, Y. H., Liu, M., Chen, L., Huang, T., et al. (2019a). Identification of synthetic lethality based on a functional network by using machine learning algorithms. *J. Cell. Biochem.* 120, 405–416. doi: 10.1002/jcb.27395
- Li, J., Lu, L., Zhang, Y.-H., Xu, Y., Liu, M., Feng, K., et al. (2019b). Identification of leukemia stem cell expression signatures through Monte Carlo feature selection strategy and support vector machine. *Cancer Gene Ther.* 27, 56–69. doi: 10.1038/s41417-019-0105-y
- Liu, L., Chen, L., Zhang, Y. H., Wei, L., Cheng, S., Kong, X., et al. (2017). Analysis and prediction of drug-drug interaction by minimum redundancy maximum relevance and incremental feature selection. *J. Biomol. Struct. Dyn.* 35, 312–329. doi: 10.1080/07391102.2016.1138142
- Loeser, R. F. (2008). Molecular mechanisms of cartilage destruction in osteoarthritis. *J. Musculoskelet. Neuronal Interact.* 8, 303–306.
- Loughlin, J., and Reynard, L. N. (2015). Osteoarthritis: epigenetics of articular cartilage in knee and hip OA. *Nat. Rev. Rheumatol.* 11, 6–7. doi: 10.1038/nrrheum.2014.189
- Martel-Pelletier, J., Boileau, C., Pelletier, J. P., and Roughley, P. J. (2008). Cartilage in normal and osteoarthritis conditions. *Best Pract. Res. Clin. Rheumatol.* 22, 351–384. doi: 10.1016/j.berh.2008.02.001
- Menold, M. M., Shao, Y., Wolpert, C. M., Donnelly, S. L., Raiford, K. L., Martin, E. R., et al. (2001). Association analysis of chromosome 15 GABAA receptor subunit genes in autistic disorder. *J. Neurogenet.* 15, 245–259. doi: 10.3109/01677060109167380
- Miranda-Duarte, A. (2018). DNA methylation in osteoarthritis: current status and therapeutic implications. *Open Rheumatol. J.* 12, 37–49. doi: 10.2174/1874312901812010037
- Niu, B., Huang, G., Zhang, L., Wang, X., Chen, F., Zhang, Y., et al. (2013). Prediction of substrate-enzyme-product interaction based on molecular descriptors and physicochemical properties. *Biomed Res. Int.* 2013:674215. doi: 10.1155/2013/674215
- Okumura, K., Saito, M., Isogai, E., Aoto, Y., Hachiya, T., Sakakibara, Y., et al. (2014). Meis1 regulates epidermal stem cells and is required for skin tumorigenesis. *PLoS One* 9:e102111. doi: 10.1371/journal.pone.0102111
- Pan, X., Chen, L., Feng, K. Y., Hu, X. H., Zhang, Y. H., Kong, X. Y., et al. (2019). Analysis of expression pattern of snoRNAs in different cancer types with machine learning algorithms. *Int. J. Mol. Sci.* 20:2185. doi: 10.3390/ijms20092185
- Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1226–1238. doi: 10.1109/tpami.2005.159
- Pereira, D., Severo, M., Santos, R. A., Barros, H., Branco, J., Lucas, R., et al. (2016). Knee and hip radiographic osteoarthritis features: differences on pain, function and quality of life. *Clin. Rheumatol.* 35, 1555–1564. doi: 10.1007/s10067-015-3087-7
- Ramos, Y. F., Bos, S. D., Lakenberg, N., Böhringer, S., den Hollander, W. J., Kloppenburg, M., et al. (2014). Genes expressed in blood link osteoarthritis with apoptotic pathways. *Ann. Rheum. Dis.* 73, 1844–1853. doi: 10.1136/annrheumdis-2013-203405
- Reynard, L. N., and Loughlin, J. (2012). Genetics and epigenetics of osteoarthritis. *Maturitas* 71, 200–204. doi: 10.1016/j.maturitas.2011.12.001
- Rice, S. J., Beier, F., Young, D. A., and Loughlin, J. (2020). Interplay between genetics and epigenetics in osteoarthritis. *Nat. Rev. Rheumatol.* 16, 268–281. doi: 10.1038/s41584-020-0407-3
- Rushton, M. D., Reynard, L. N., Barter, M. J., Refaie, R., Rankin, K. S., Young, D. A., et al. (2014). Characterization of the cartilage DNA methylome in knee and hip osteoarthritis. *Arthritis Rheumatol.* 66, 2450–2460. doi: 10.1002/art.38713
- Shu, Y., Zhang, N., Kong, X., Huang, T., and Cai, Y. D. (2014). Predicting A-to-I RNA editing by feature selection and random forest. *PLoS One* 9:e110607. doi: 10.1371/journal.pone.0110607
- Sun, L., Yu, Y., Huang, T., An, P., Yu, D., Yu, Z., et al. (2012). Associations between ionomic profile and metabolic abnormalities in human population. *PLoS One* 7:e38845. doi: 10.1371/journal.pone.0038845
- Sun, X., Li, J., Gu, L., Wang, S., Zhang, Y., Huang, T., et al. (2018). Identifying the characteristics of the hypusination sites using SMOTE and SVM algorithm with feature selection. *Curr. Proteomics* 15, 111–118. doi: 10.2174/1570164614666171109120615
- Sun, X., Xiao, L., Chen, J., Chen, X., Chen, X., Yao, S., et al. (2020). DNA methylation is involved in the pathogenesis of osteoarthritis by regulating CtBP expression and CtBP-mediated signaling. *Int. J. Biol. Sci.* 16, 994–1009. doi: 10.7150/ijbs.39945
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2018). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. doi: 10.1093/nar/gky1131
- van Eijk, K. R., de Jong, S., Boks, M. P., Langeveld, T., Colas, F., Veldink, J. H., et al. (2012). Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. *BMC Genomics* 13:636. doi: 10.1186/1471-2164-13-636
- Wang, D., Li, J.-R., Zhang, Y.-H., Chen, L., Huang, T., and Cai, Y.-D. (2018). Identification of differentially expressed genes between original breast cancer and xenograft using machine learning algorithms. *Genes* 9:155. doi: 10.3390/genes9030155
- Wang, Q. F., Li, Y. J., Dong, J. F., Li, B., Kaberlein, J. J., Zhang, L., et al. (2014). Regulation of MEIS1 by distal enhancer elements in acute leukemia. *Leukemia* 28, 138–146. doi: 10.1038/leu.2013.260
- Wang, S. B., and Huang, T. (2019). The early detection of asthma based on blood gene expression. *Mol. Biol. Rep.* 46, 217–223. doi: 10.1007/s11033-018-4463-6
- Webb, B. T., Sullivan, P. F., Skelly, T., and van den Oord, E. J. (2008). Model-based gene selection shows engrailed 1 is associated with antipsychotic response. *Pharmacogenet. Genomics* 18, 751–759. doi: 10.1097/FPC.0b013e32830162bc
- Zhang, N., Huang, T., and Cai, Y. D. (2014). Discriminating between deleterious and neutral non-frameshifting indels based on protein interaction networks and hybrid properties. *Mol. Genet. Genomics* 290, 343–352.
- Zhang, N., Huang, T., and Cai, Y. D. (2015). Discriminating between deleterious and neutral non-frameshifting indels based on protein interaction networks and hybrid properties. *Mol. Genet. Genomics* 290, 343–352. doi: 10.1007/s00438-014-0922-5
- Zhang, N., Wang, M., Zhang, P., and Huang, T. (2016). Classification of cancers based on copy number variation landscapes. *Biochim. Biophys. Acta Gen. Subj.* 1860(11 Part B), 2750–2755. doi: 10.1016/j.bbagen.2016.06.003
- Zhang, P. W., Chen, L., Huang, T., Zhang, N., Kong, X. Y., and Cai, Y. D. (2015). Classifying ten types of major cancers based on reverse phase protein array profiles. *PLoS One* 10:e0123147. doi: 10.1371/journal.pone.0123147
- Zhang, X., Chen, C., Wu, M., Chen, L., Zhang, J., Zhang, X., et al. (2012). Plasma microRNA profile as a predictor of early virological response to interferon treatment in chronic hepatitis B patients. *Antivir. Ther.* 17, 1243–1253. doi: 10.3851/imp2401
- Zhang, Y. H., Huang, T., Chen, L., Xu, Y., Hu, Y., Hu, L. D., et al. (2017). Identifying and analyzing different cancer subtypes using RNA-seq data of blood platelets. *Oncotarget* 8, 87494–87511. doi: 10.18632/oncotarget.20903
- Zhao, T. H., Jiang, M., Huang, T., Li, B. Q., Zhang, N., Li, H. P., et al. (2013). A novel method of predicting protein disordered regions based on sequence features. *Biomed Res. Int.* 2013:414327. doi: 10.1155/2013/414327
- Zheng, H. F., Forgetta, V., Hsu, Y. H., Estrada, K., Rosello-Diez, A., Leo, P. J., et al. (2015). Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature* 526, 112–117. doi: 10.1038/nature14878



- Zhou, Y., Yang, L., Wang, H., Chen, X., Jiang, W., Wang, Z., et al. (2020). Alterations in DNA methylation profiles in cancellous bone of postmenopausal women with osteoporosis. *FEBS Open Bio* 10, 1516–1531. doi: 10.1002/2211-5463.12907
- Zhou, Y., Zhang, N., Li, B. Q., Huang, T., Cai, Y. D., and Kong, X. Y. (2015). A method to distinguish between lysine acetylation and lysine ubiquitination with feature selection and analysis. *J. Biomol. Struct. Dyn.* 33, 2479–2490. doi: 10.1080/07391102.2014.1001793

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Wu, Shou, Wang, Huang and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*