Methylome and transcriptome profiles in three yak tissues revealed that DNA methylation and transcription factor ZGPAT co-regulate milk production

Jinwei Xin
State Key Laboratory of Hulless Barley and Yak Germplasm Resources and Genetic Improvement
Zhixin Chai
Southwest Minzu University
Chengfu Zhang
State Key Laboratory of Hulless Barley and Yak Germplasm Resources and Genetic Improvement
Qiang Zhang
State Key Laboratory of Hulless Barley and Yak Germplasm Resources and Genetic Improvement
Yong Zhu
State Key Laboratory of Hulless Barley and Yak Germplasm Resources and Genetic Improvement
Hanwen Cao
State Key Laboratory of Hulless Barley and Yak Germplasm Resources and Genetic Improvement
Yangji Cidan
State Key Laboratory of Hulless Barley and Yak Germplasm Resources and Genetic Improvement
Xiaoying Chen
State Key Laboratory of Hulless Barley and Yak Germplasm Resources and Genetic Improvement
Hui Jiang
State Key Laboratory of Hulless Barley and Yak Germplasm Resources and Genetic Improvement
Jincheng Zhong
Southwest minzu university
Qiumei Ji (jiqiumei07@163.com)
State Key Laboratory of Hulless Barley and Yak Germplasm Resources and Genetic Improvement

Research article

Keywords: milk production, DNA methylation, transcription factor, epigenetic regulation

Posted Date: August 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-20775/v2

Version of Record: A version of this preprint was published on October 20th, 2020. See the published version at https://doi.org/10.1186/s12864-020-07151-3.

Abstract

Background Domestic yaks play an indispensable role in sustaining the livelihood of Tibetans and other ethnic groups on the Qinghai-Tibetan Plateau (QTP), by providing milk and meat, and have evolved numerous physiological adaptabilities to high-altitude landscape, such as strong capacity of blood oxygen transportation and high metabolism. The role of DNA methylation and network of gene expression underlying milk production and adaptation to high altitudes of yak need further exploration.

Results We performed genome-wide DNA methylome and transcriptome analyses of breast, lungs, and gluteal muscle from yaks of different ages. We identified differentially methylated regions (DMRs) across age groups within the each tissue, and breast tissue had considerably more differentially methylated regions than that from the three younger age groups. Hypomethylated genes with high expression level might regulate milk production by influencing protein processing in the endoplasmic reticulum. Weighted gene correlation network analysis revealed that the “hub” gene ZGPAT was highly expressed in post-mature breast tissue and that it potentially regulated the transcription of 280 genes, which play roles in regulating protein synthesis, processing, and secretion. Besides, Tissue network analysis indicates that high expression of HIF1A regulates energy metabolism in the lung.

Conclusions The results of this comprehensive study provide a solid basis for understanding the epigenetic mechanisms underlying milk production in yaks, which could be helpful to breeding programs aimed at improving milk production.

Background

Domestic yaks play an indispensable role in sustaining the livelihood of Tibetans and other ethnic groups on the Qinghai-Tibetan Plateau (QTP) and in the Himalayas and connecting Central Asian highlands by providing milk, meat, hide, fiber, fuel, and transportation [1, 2]. Milk is an important source of high-quality protein because of its high content of essential amino acids, such as lysine, which is deficient in many human diets [3], and because of its well-known physiological effects, such as immunomodulatory and gastrointestinal activities [4]. The milk protein content and composition influence the technological properties of milk and are therefore important for the dairy industry, especially in Europe, where the majority of the milk produced is transformed into cheese. In recent decades, there have been extraordinary advances in our knowledge of the physiology and biochemistry of the lactating mammary gland. It has been clearly demonstrated that milk protein synthesis in the mammary gland depends on hormonal and developmental cues that modulate the transcriptional and translational regulation of genes through the activity of specific transcription factors, non-coding RNAs, and alterations of the chromatin structure in the mammary epithelial cells [5]. The interplay between all these aforementioned factors might play a key role in milk protein synthesis, which is crucial during the onset of and throughout lactation in high-producing dairy cattle. Despite such advances, little is currently known about the regulation of the physiological and cellular mechanisms required for milk protein synthesis and secretion in yak. We hypothesized that genes related to milk production might be regulated by DNA methylation

and distinct sub-modules of correlated expression variation might be indentified. In this study, we performed genome-wide DNA methylome and transcriptome analyses of lung, breast, and biceps brachii muscle tissues at four different month ages (MA) from yaks to identify the regulatory networks associated with milk protein synthesis, metabolism, and secretion in yak.

Results

Global DNA methylation and gene expression in the breast, lungs, and biceps brachii muscle at different ages

We generated the methylomes and transcriptomes of lung, breast, and biceps brachii muscle tissues at four different month ages (MA), representing four life stages with 3 replicates (MA= 6, 30, 54, and 90 months; young, per-mature, mature and post-mature), from a total of 12 female Riwoqe yaks. Among these yaks, only the individuals at 90 months were in lactation period with ~3.16 kg/day milk yield. After performing sequence quality control and filtering, we obtained a single-base resolution methylome covering 85.6% (27,471,373/32,092,725) of CpG sites across the genome with an average depth of 22.5×. We first calculated pairwise Pearson's correlations of CpG sites with at least 10× coverage depth across all samples, and the samples were well clustered by tissue (Figure 1a). CpG methylation levels of biological replicates highly correlated with each other (median Pearson's r = 0.74), and the correlation of CpG methylation level between ages (median Pearson's r = 0.72) was relatively weaker and showed the lowest coefficients among tissues (median Pearson's r = 0.66) (Figure 1c). The transcriptome sequencing data of all samples were aligned to our newly assembled yak genome reference (unpublished), and subsequently obtained the transcripts. In total, we obtained a total of 2,0504 transcripts, which were annotated to the Gene Ontology (GO) [6], InterPro [7], Kyoto Encyclopedia of Genes and Genomes (KEGG) [8], Swiss-Prot [9], and TrEMBL [10] databases (Table S1). We also calculated pairwise Pearson's correlations of all transcripts and obtained similar results as with DNA methylation (Figure 1b). Biological replicates showed the highest correlation coefficients, while different tissues showed the lowest correlation coefficients (Figure 1d).

Differential DNA methylation among the age groups was involved in protein processing in endoplasmic reticulum We looked for differentially methylated regions (DMRs) across age groups within the breast, lungs, and gluteal muscle (Table S2-4). Within the lung and gluteal-muscle tissues, age groups did not differ in age-related DMRs (A-DMRs), but post-mature breast tissue had considerably more differential methylation regions than the three younger age groups (Figure 2a). We investigated the correlations between DMRs and their associated differentially expressed genes, and the ratios of negatively to positively

correlated gene pairs were 1.02 for promoters with DMRs and 0.95 for genebody with DMRs. This might be explained that not every methylation was to be correlated with the expression of its associated gene due to the complexity in gene regulation [11].

At ~90 months, ~120 days after the third birth, yaks were in the lactation period (milk yield, ~3.16kg/day)(Li & Jiang, 2019), so it is possible that the observed methylation may at least partially control yak lactation. As methylation on promoter decrease the gene expression [11], we then selected 375 hypomethylated promoter (highly expressed) genes along with 207 hypermethylated promoter (lowly expressed) genes from post-mature yak breast tissue. The hypomethylated (highly expressed) genes were enriched in only “protein processing in endoplasmic reticulum (ER)” (nine genes, 2.964-fold enrichment, p = 0.0049). Specifically, the genes were involved in vesicle trafficking (SEC23B), oligosaccharide linking (MOGS, RPN2), folding and assembly (HSPA5), transportation (LMAN2, SEL1L), and ubiquitination and degradation (UBE2J1, UBE2J2, DERL2) [12, 13]. These genes were significantly uprelated at 90 months in breast tissues, but not in lung and biceps brachii muscle tissues (Figure 2e). Thus, methylation might be involved in regulation of milk production by influencing protein processing in the endoplasmic reticulum during the vigorous period of lactation.

We examined A-DMRs that overlapped across age groups. Young and post-mature tissues rarely shared A-DMRs when comparing the lung and muscle tissues (young, muscle: 586 A-DMRs, breast: 2,249, lung: 496; post-mature, muscle: 470, breast: 12,050, lung: 772) (Figure 2b, c). Pre-mature and mature stages also rarely shared A-DMRs across muscle and lung tissues (Figure 2d), suggesting that methylation patterns were already established at the young stage and that no extensively divergent epigenetic difference occurred through the different ages under natural high-altitude conditions.

Consensus network analysis for tissues and age groups

We first performed a multi-way ANOVA test for each gene among all samples (n=36), to test the null hypothesis that the gene expression level did not differ among age groups and tissues. At the threshold for significance ($p<0.05$), 417 age-related and 8,560 tissue-related genes were selected for further weighted gene correlation network analysis (WGCNA), which takes advantage of the correlations among genes and groups genes into modules using network topology [14]. Subsequently, we conducted WGCNA for tissue- and age-related gene expression separately to identify a “consensus network”–a common pattern of genes that are correlated in all conditions. We performed consensus network, module statistic, and eigengene network analyses to identify modules, assess relationships between modules and traits, and study the relationships between co-expression modules [15]. The consensus networks identified for tissues and age groups had clearly delineated modules

(Figure 3a, 3b), and the modules identified were significantly correlated with tissues and age groups (Figure 3c, 3d).

Age network analysis indicates that ZGPAT might regulate milk production

Within the age-related gene network, the largest module (“turquoise”, n=356) had opposite directions of correlation for breast tissue (r=-0.57, p=3e-04) and age (r= 0.37, p=0.03) and showed a positive correlation with biceps brachii muscle (Table S5). The breast tissue had a stronger signal than age and possibly overwhelmed the signal from age. Genes in this module were enriched in the GO categories of “protein polyubiquitination,” “RNA polymerase II core promoter proximal region sequence-specific DNA binding,” “ATP binding,” “transcription, DNA-templated,” and “negative regulation of transcription from RNA polymerase II promoter” (p-values of 0.000344, 0.000822, 0.000991, 0.00139, and 0.00244, respectively). The “blue” (n=48) and “grey” (n=13) modules showed only a negative correlation with age (r= -0.58, p= 2e-04; r= -0.76, p= 6e-08, respectively) and exhibited no enrichment of GO categories for genes. After applying the threshold of the absolute value of gene significance for age (|GS| >0.5) and module membership measures (|MM| >0.6) in each module, we defined 20 and 7 “hub” genes in the “turquoise” and “blue” modules (Table 1). The gene expression of the “hub” genes was well clustered by modules, which was consistent with the opposite directions of correlation with age (“turquoise” r=0.37, “blue” r=-0.58). The upregulated expression level of the “hubs” in breast tissue at 90 months of age indicated that the “turquoise” module had a stronger correlation with breast tissue than with age (Figure 4a).

Table 1. List “hub” genes in the consensus network for age

Yak ID	Gene symbol	Module color	Gene significance	p-value	Module membership	p-value
BmuPB009868	AP3D1	turquoise	0.513965075	0.001344116	0.868945999	6.37E-12
BmuPB004083	TMEM30A	turquoise	0.505722627	0.001652657	0.848054843	6.65E-11
BmuPB000726	DDRGK1	turquoise	0.53596122	0.000754414	0.842070133	1.22E-10
BmuPB019011	OTUD3	turquoise	0.543916749	0.000606215	0.828680521	4.37E-10
BmuPB000091	CNPPD1	turquoise	0.51195228	0.001414359	0.814537409	1.50E-09
BmuPB006815	TOR1AIP1	turquoise	0.544705745	0.000593032	0.809789221	2.21E-09
BmuPB007137	MGAT4A	turquoise	0.545969144	0.000572452	0.798170686	5.51E-09
BmuPB007385	HECA	turquoise	0.603273223	9.84E-05	0.770629907	3.85E-08
BmuPB019443	SYAP1	turquoise	0.500377628	0.001884511	0.733120646	3.68E-07
BmuPB009825	PHAX	turquoise	0.561938984	0.00036184	0.70497442	1.59E-06
BmuPB008032	UBE2G2	turquoise	0.52385538	0.001041698	0.70208134	1.83E-06
BmuPB012517	SYF2	turquoise	0.568328068	0.000299194	0.699367268	2.08E-06
BmuPB001610	FURIN	turquoise	0.567906296	0.000303008	0.697046314	2.32E-06
BmuPB000679	ZGPAT	turquoise	0.504624296	0.001698141	0.683958391	4.25E-06
BmuPB007049	SKP2	turquoise	-0.527608908	0.000943732	-0.669790419	7.91E-06
BmuPB000224	C2orf6	turquoise	0.531191165	0.000857925	0.652871802	1.59E-05
BmuPB007420	TAB2	turquoise	0.505905132	0.001645204	0.652844304	1.59E-05
BmuPB018569	ORAOV1	turquoise	0.51035386	0.001472421	0.637043276	2.95E-05
BmuPB010550	CCPG1	turquoise	0.644779605	2.19E-05	0.612874899	7.08E-05
BmuPB012521	TMEM57	turquoise	0.57317319	0.000258349	0.60967979	7.91E-05
BmuPB013324	MCM3	blue	-0.583585649	0.000186982	0.841505398	1.29E-10
BmuPB010064	SPC24	blue	-0.509962181	0.001486965	0.744497748	1.93E-07
BmuPB003102	C17orf49	blue	-0.526803776	0.000964031	0.741805787	2.26E-07
BmuPB015902	SERPINH1	blue	-0.607769763	8.44E-05	0.721017606	7.04E-07
BmuPB016582	SRPX2	blue	-0.51838468	0.001200558	0.698151166	2.20E-06
BmuPB012996	UCK2	blue	-0.588274454	0.000161067	0.677420455	5.68E-06
BmuPB015372	UMPS	blue	-0.503577413	0.001742514	0.648111658	1.92E-05

We then used the AnimalTFDB 3.0 database [16] to examine transcription factors in these 27 “hubs” and found that ZGPAT encodes a transcription regulator protein and was significantly upregulated in breast tissue at 90 months of age (Figure 4a). previous study reported that this protein specifically binds the 5'-GGAG[GA]A[GA]A-3' consensus sequence and represses transcription by recruiting the chromatin multi-complex NuRD to target promoters [17]. High expression of ZGPAT in breast tissue at 90 months of age was observed, and it potentially regulated the transcription of 280 genes (weight >0.15) in the network from the “turquoise” module. To identify the most important cellular activities controlled by this TF regulatory network, we analyzed over-represented GO biological process and molecular function terms and KEGG pathways. These potential target genes were enriched in the GO categories of “protein binding,” “ATP binding,” and “zinc ion binding,” among others, and the KEGG categories of “aminoacyl-tRNA biosynthesis,” “autophagy animal,” and “protein processing in endoplasmic reticulum” (Figure 4b). These enriched GO terms and KEGG pathways are likely to play roles in regulating protein synthesis, processing, and secretion in breast tissue. For example, 6 of 7 genes from “protein processing in endoplasmic reticulum” were also upregulated at 90 months of age in breast tissue (Figure 4c) and involved in multiple processes in the endoplasmic reticulum, including vesicle trafficking (SEC24C), folding and assembly (SELENOS), transportation (BCAP31), and ubiquitination and degradation (BAG1, UBE2G2, and MARCH6) [12, 13]. Only DNAJC10 was downregulated at 90 months of age in breast tissue, and this gene encodes an endoplasmic reticulum co-chaperone that is part of the endoplasmic reticulum-associated degradation complex involved in recognizing and degrading misfolded proteins [13].

Tissue network analysis indicates that high expression of HIF1A regulates genes involved in energy metabolism in the lung

Within the tissue-related gene module network, four modules showed a positive correlation and two modules showed a negative correlation with the lung, and all significant module-trait relationships were negative in the muscle but positive in the breast (Figure 3d). Moreover, 99.54% of the total 8,560 tissue-related genes were related to the top 4 modules (“turquoise,” n=3833, “blue,” n=2795, “brown,” n=1052, “yellow,” n=339) (Figure 5a, Table S6), and these modules were also highly correlated with other modules; for example, “brown,” “yellow,” and “black” showed a high eigengene adjacency with each other (Figure 5b).

We applied the more stringent threshold of the absolute value of gene significance for the age and module membership measures in the top four modules to identify “hub” genes in

the “turquoise,” “blue,” “brown,” and “yellow” modules. With the threshold values of |GS|>0.7 and |MM| >0.8, 34 “hub” genes were identified in the “turquoise” module. Then, 24 “hubs” were filtered from the gene significance of module-lung relationships, and 10 “hubs” were filtered from the gene significance of module-breast relationships; these were further divided into 3 clusters by hierarchical clustering, and they showed high expression levels in the breast (cluster 1), lung (cluster 2), and biceps brachii muscle (cluster 3) tissues, respectively, with distinct clustering patterns by tissue (Figure 5c).

Table 2. List “hub” genes in the consensus network for tissue

	Gene symbol	Module color	Tissue	Gene significance	p-value	Module membership	p-value
3014336	EEF1G	turquoise	lung	-0.73907039	2.64E-07	0.826748126	5.20E-10
3017352	PMS1	turquoise	lung	-0.71599797	9.13E-07	0.850552958	5.12E-11
3000878	MTPAP	turquoise	lung	-0.715583465	9.33E-07	0.912877613	8.73E-15
3018762	CXHXorf58	turquoise	lung	-0.709433358	1.27E-06	0.82262518	7.50E-10
3014450	RWDD4	turquoise	lung	-0.708526806	1.33E-06	0.923713618	9.94E-16
3011005	HUS1	turquoise	lung	-0.707137304	1.43E-06	0.875052199	2.97E-12
3005540	MTUS1	turquoise	lung	-0.704482832	1.62E-06	0.871626441	4.58E-12
3011453	EBF3	turquoise	lung	-0.703373458	1.71E-06	0.870717052	5.13E-12
3010608	LAMB3	turquoise	lung	0.70783756	1.38E-06	-0.85067459	5.05E-11
3004299	C3H3orf58	turquoise	lung	0.708872549	1.31E-06	-0.88522675	7.61E-13
3020871	G6PD	turquoise	lung	0.708930411	1.30E-06	-0.90727231	2.41E-14
3007438	ZCCHC6	turquoise	lung	0.709740967	1.25E-06	-0.84738754	7.13E-11
3007592	CTDSPL	turquoise	lung	0.711634321	1.14E-06	-0.8813581	1.30E-12
3004894	HIF1A	turquoise	lung	0.713237417	1.05E-06	-0.87302984	3.84E-12
3020508	FAM122B	turquoise	lung	0.720139618	7.37E-07	-0.82428698	6.48E-10
3018102	CCDC82	turquoise	lung	0.720592854	7.20E-07	-0.90665968	2.68E-14
3014539	CNTRL	turquoise	lung	0.722130497	6.64E-07	-0.87149974	4.65E-12
3003882	VAV3	turquoise	lung	0.725601285	5.53E-07	-0.89496378	1.82E-13
3020153	EPS8L1	turquoise	lung	0.727495147	4.99E-07	-0.82286844	7.34E-10
3010308	PGM2	turquoise	lung	0.746811608	1.69E-07	-0.83795918	1.83E-10
3004008	GDAP2	turquoise	lung	0.754806874	1.05E-07	-0.88661303	6.26E-13
3012568	WASF2	turquoise	lung	0.757428944	8.92E-08	-0.83874427	1.69E-10
3011966	ACTG1	turquoise	lung	0.772114072	3.49E-08	-0.84373776	1.03E-10
3014173	STAT6	turquoise	lung	0.80896084	2.37E-09	-0.84981355	5.53E-11
3015106	MAN2A1	turquoise	breast	0.705124693	1.57E-06	-0.84782348	6.81E-11
3008614	VPS26A	turquoise	breast	0.731553581	4.01E-07	-0.88669366	6.18E-13
3018496	FAM92A	turquoise	breast	0.718223061	8.14E-07	-0.85598454	2.85E-11
3005078	RASA1	turquoise	breast	0.705902582	1.51E-06	-0.85869193	2.11E-11

3011476	FAM53B	turquoise	breast	-0.728187435	4.81E-07	0.848680696	6.23E-11
3008348	EPDR1	turquoise	breast	-0.710771064	1.19E-06	0.832481139	3.08E-10
3003453	TROVE2	turquoise	breast	0.72654448	5.26E-07	-0.80669368	2.84E-09
3021200	VWA7	turquoise	breast	-0.722814498	6.41E-07	0.814033145	1.56E-09
3010528	RNF111	turquoise	breast	0.737234077	2.92E-07	-0.80146106	4.28E-09
3015811	FRMD3	turquoise	breast	-0.739223421	2.61E-07	0.809147869	2.33E-09

According to the AmalTFDB 3.0 database [16], EBF3, HIF1A, and STAT6 were annotated as transcription factors. EBF3 encodes a member of the early B-cell factor (EBF) family of DNA binding transcription factors. EBF proteins are involved in B-cell differentiation, bone development, and neurogenesis and may also function as tumor suppressors [18]. STAT6 is a member of the STAT family of transcription factors. In response to cytokines and growth factors, STAT family members are phosphorylated by the receptor associated kinases and then form homo- or heterodimers that translocate to the cell nucleus where they act as transcription activators [19]. HIF1A, hypoxia-inducible factor-1, functions as a master regulator of the cellular and systemic homeostatic response to hypoxia by activating the transcription of many genes, including those involved in energy metabolism, angiogenesis, and apoptosis and other genes whose protein products increase oxygen delivery or facilitate metabolic adaptation to hypoxia [20]. HIF1A was found to be upregulated in the lung and to potentially regulate the transcription of 2008 genes (weight >0.15), which were enriched in multiple GO biological process and molecular function categories and KEGG pathways. Notably, most of the enriched GO terms and KEGG pathways were related to energy metabolism, such as “mitochondrial respiratory chain complex I assembly,” “NADH dehydrogenase (ubiquinone) activity,” “ATP binding,” “mitochondrial translation,” “tricarboxylic acid cycle,” and “GTP binding” in GO terms and “thermogenesis,” “carbon metabolism,” and “citrate cycle TCA cycle” in KEGG pathways (Figure 5d). Mitochondria function as the primary energy producers of the cell and serve as the center of biosynthesis, the oxidative stress response, and cellular signaling, placing them at the hub of a variety of immune pathways [21]. NADH dehydrogenase is a core subunit of the mitochondrial membrane respiratory chain and believed to belong to the minimal assembly required for catalysis [22]. Protein which binds ATP or GTP, carries three phosphate groups esterified to a sugar moiety and represents energy and phosphate sources for the cell [23, 24]. The tricarboxylic acid cycle, a series of metabolic reactions in aerobic cellular respiration, occurs in the mitochondria of animals and plants, and during this cycle, acetyl-CoA, formed from pyruvate produced during glycolysis, is completely oxidized to CO_2 via the interconversion of various carboxylic acids. It results in the

reduction of NAD and FAD to NADH and FADH2, whose reducing power is then used indirectly in the synthesis of ATP by oxidative phosphorylation [25]. The “thermogenesis” pathway is essential for warm blooded animals, ensuring normal cellular and physiological functioning under conditions of environmental challenge [26].

Discussion

Numerous studies reported transcription profiling of mammary gland in different livestock animals such as cattle [27], sheep [28], and goat [29], and DNA methylation profiling of mammary gland in cattle [30]. These studies have indicated temporal and spatial specificity in the methylome and transcriptome profiles of the mammary gland in different species. However, these work only reported differential gene expression profile in mammary gland, and the regulatory network is still unknown. In the present study, we generated the methylomes and transcriptomes of lung, breast, and biceps brachii muscle tissues at four different month ages (MA) from yak for the first time, representing four life stages (MA= 6, 30, 54, and 90 months; young, pre-mature, mature, and post-mature). We found that breast tissue at 90 months showed considerably differential methylation levels compared with other month ages, but not lung and biceps brachii muscle tissues. Enrichment analysis for upregulated genes with hypomethylated DMRs from breast tissues of yaks at 90 months showed that the activation of “protein processing in endoplasmic reticulum (ER)” pathway might be regulated by DNA methylation. As only the individuals at 90 months were in lactation period, this result support that DNA methylation is involved in milk production by influencing protein processing in the endoplasmic reticulum during the vigorous period of lactation.

One of the contributions of this study is that hub genes were identified by WGCNA. The data show that the hub genes with the highest MM and GS in modules of interest should be considered as the natural candidates for further research. This study identified turquoise module genes associated with milk yield, and 20 genes were considered as the hub genes and showed the highest mRNA expression level in breast tissue at 90 months, when yaks enter the lactation period. In these hub genes, ZGPAT was annotated as transcription factor and it potentially regulated the transcription of 280 genes in the network from the “turquoise” module, which were enriched in the KEGG categories of “aminoacyl-tRNA biosynthesis,” “autophagy animal,” and “protein processing in endoplasmic reticulum”. This result revealed that ZGPAT is likely to play driving role in regulating protein synthesis, processing, and secretion in breast tissue. Moreover, the 7 genes potentially regulated by ZGPAT in “protein processing in endoplasmic reticulum” were totally different from aforementioned 9 genes regulated by hypomethylation, illustrating that DNA methylation and transcription factor possibly co-regulate milk production. In addition, the tissue network analysis indicates the central role of HIF1A in regulating energy metabolism, which is important in adaptation to low temperature and hypoxia in high altitude environment.

Conclusions

The results of this comprehensive study provide a solid basis for understanding the roles of DNA methylation and transcriptional network underlying milk protein synthesis and high-altitude adaptation in yaks. This information advances our understanding of regulatory network in mammary gland at different development stages and could be helpful to breeding programs aimed at improving milk production.

Methods

Animals and samples

In total, twelve female yaks that were 6, 30, 54 or 90 months old with 3 replicates for each month age (sampled from private farms in Riwoqe at altitudes of 3800-4000 meters above sea level; an indigenous yak breed distributed in Riwoqe, Tibet, China) were collected in June-December of 2016. At the time of slaughter, their mean live weights were 44.93 (6 months old), 153.06 (30 months old), 188.3 (54 months old) and 243.56 kg (90 months old). Among these yaks, only the individuals at 90 months were in lactation period with ~3.16 kg/day milk yield (~120 days after the third birth), and the individuals at 54 months were in dry period (~540 days after the first birth) [31]. The blood relationship within the last three generations among individuals was unknown, which were housed simultaneously and fed the same diets. The yaks were not fed the night before they were slaughtered as necessary to ameliorate suffering, and were humanely sacrificed by performing the following procedures: (1) taking showers for the yaks with clean water close to body temperature (35-38°C), (2) yaks were electrically stunned (120V dc, 12s) prior to exsanguination, (3) during the coma, yaks were sacrificed by bloodletting from carotid artery and jugular vein, (4) after further dissection, each tissue including breast, lung, and biceps brachii muscle samples were rapidly obtained from each individual, immediately frozen in liquid nitrogen, and stored at −80°C until RNA and DNA extraction.

Whole genome bisulfite sequencing

The QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) was used to isolate the high-quality DNA from each sample. 1 μg of genomic DNA was fragmented by sonication to a mean size of approximately 250 bp and subsequently used for whole genome bisulfite sequencing (WGBS) library construction using an Acegen Bisulfite-Seq Library Prep Kit (Acegen, Shenzhen, GD, China) following the manufacturer's instructions. Briefly, fragmented DNA was end-repaired, 5'-phosphorylated, 3'-dA-tailed, and then ligated to methylated adapters. The methylated adapter-ligated DNAs were purified using 1× Agencourt AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA) and subjected to bisulfite conversion with a ZYMO EZ DNA Methylation-Gold Kit (Zymo research, Irvine, CA, USA). The converted DNAs were then amplified using 25 μl HiFi HotStart U+ RM and 8-bp index primers with a final concentration of 1 μM each. The constructed WGBS libraries were then analyzed with an Agilent 2100 Bioanalyzer (Agilent Technologies, SantaClara, CA, USA), quantified with a Qubit fluorometer with Quant-iT dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA,USA), and finally sequenced on an Illumina Hiseq X ten sequencer (PE150 mode) (Illumina, San Diego, CA, USA)

Methylation calculation and identification of DMRs

We filtered low quality reads that contained more than 5 ‘N’s or over 50% of the sequence with low quality value (Phred score < 5). The sequencing reads of the samples were aligned to the yak reference genome [32] using BSMAP (Version 2.74) [33]. The methylated CpG (mCG) sites were identified following a previously described algorithm [34]. The methylation levels for each sample were calculated using in-house perl scripts. Differentially methylated regions (DMRs) were identified using metilene (Version 0.2-6) within a 500 bp sliding window at 250 bp steps with at least 10 CpGs covered by over 10× sequence reads, applying the thresholds of differential methylation β >=15%, FDR for two-dimensional Kolmogorov-Smirnov-Test p-value <0.05. [35]. The enrichment analysis were conducted using WebGestalt (WEB-based Gene SeT AnaLysis Toolkit) [36].

Total RNA extraction, library preparation, and sequencing

The TRIzol reagent (Invitrogen, Carlsbad, CA, USA) was used to isolate the total RNA of each sample. The purity, concentration, and integrity of RNA were checked using the NanoPhotometer spectrophotometer (IMPLEN, Westlake Village, CA, USA), the Qubit RNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA), and the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 System (Agilent Technologies, SantaClara, CA, USA), respectively. We utilized 3 μg high-quality RNA per sample as input material for RNA-seq library preparation. First, we removed ribosomal RNA by the Epicentre Ribo-Zero rRNA Removal Kit (Epicentre, Madison, WI, USA). Second, the rRNA-depleted RNA was used to create sequencing libraries by the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, Ipswich, MA, USA). Finally, the library products were purified using 1× Agencourt AMPure XP magnetic beads (Beckman Coulter, Brea, CA, USA) and the Agilent Bioanalyzer 2100 System (Agilent Technologies, SantaClara, CA, USA) was employed to assess the library quality. After completing the clustering of the index-coded samples on a cBot Cluster Generation System using the TruSeq PE Cluster Kit v3-cBot-HS (Illumina, San Diego, CA, USA), the libraries were sequenced on the Illumina HiSeq X Ten Platform to generate 150 bp paired-end reads.

Quality analysis, transcriptome assembly, and abundance estimation

Clean reads were obtained by removing reads containing adapter or poly-N and low-quality reads (over 10% of the sequence with quality value < 30) from the raw data using in-house perl scripts.. All the downstream analyses were based on the good-quality clean reads. Paired-end clean reads were mapped to the yak reference genome (unpublished data) with STAR (available at https://github.com/alexdobin/STAR/releases). The mapped reads of each sample were assembled using StringTie [37]. The mapped reads of each sample were assembled using StringTie. Then, all transcriptomes from the samples were merged to reconstruct a comprehensive transcriptome using perl scripts. After the final transcriptome was generated, StringTie and edgeR were used to estimate the expression levels of all transcripts [38]. StringTie was used to assess the expression level of mRNAs by calculating fragments per kilobase of transcript per million fragments mapped (FPKM). Differentially expressed mRNAs were identified using the DESeq2 package, with the criteria of fold-change log2 > 1 or log2 < -1 and with the statistical significance set to FDR < 0.05.

Weighted gene correlation network analysis

A WGCNA network [14] was generated for age-related genes and tissue-related genes. Consensus networks and module statistics followed the overall approach described by Langfelder et al. (2008). Briefly, the network was derived based on a signed Spearman correlation using a b of 10 as a weight function. The topological overlap metric (TOM) [15] was derived from the resulting adjacency matrix and used to cluster the modules using the blockwiseModules function (blockwise Consensus Modules, for the consensus modules) and the dynamic tree cut algorithm [15] with a height of 0.25 and a deep split level of 2, a reassign threshold of 0.2, and a minimum module size of 30 (100 for the consensus network). The eigenmodules–essentially the first principal component of the modules, which can be used as a "signature" of a module's gene expression–were then correlated with the dose, and each module that was correlated with the dose-response curve with a p-value < 0.01 (p-value < 0.05 for the consensus network) was considered statistically significant.

Abbreviations

QTP: Qinghai-Tibetan Plateau

DMRs: differentially methylated regions

WGBS: whole genome bisulfite sequencing

FPKM: fragments per kilobase of transcript per million fragments

WGCNA: Weighted gene correlation network analysis

TOM: topological overlap metric

MA: month ages

GO: Gene Ontology

KEGG: Kyoto Encyclopedia of Genes and Genomes

A-DMRs: age-related DMRs

ER: endoplasmic reticulum

GS: gene significance

MM: module membership measures

Declarations

Ethics approval and consent to participate

All protocols for collection of the semen samples of yaks were reviewed and approved by the Ethics Committee at Institute of Animal Science and Veterinary, Tibet Academy of Agricultural and Animal Husbandry Sciences (Permit Number: 2015-216).

Consent for publication

Not applicable.

Availability of data and material

The DNA methylation data and RNA transcriptome data in this study are available in SRA under the accession numbers PRJNA530286 and PRJNA512958, respectively.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by a program of Provincial Department of Finance of the Tibet Autonomous Region (No: XZNKY-2019-C-052), Program National Beef Cattle and Yak Industrial Technology System (No: CARS-37), Basic Research Programs of Sichuan Province (No: 2019YJ0256) and the Open Project Program of State Key Laboratory of Hulless Barley and Yak Germplasm Resources and Genetic Improvement (NO:XZNKY-2019-C-007K10). The funding bodies had no role in design of study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

JX, QJ and JZ planned and coordinated the study and wrote the manuscript. CY, XC and HJ collected the samples. ZC and CZ performed the library constraction and sequencing and the quality control analysis. QZ, YZ and HC performed downstream analysis of the data and assisted in the generation of additional files for the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the animal husbandry station of Chang Du and the agricultural bureau of Riwoqe for all support.

References

1. Wiener G, Han J-L, Long R-J: **The Yak**, 2nd edn. Bangkok, Thailand: Regional Office for Asia and the Pacific, Food and Agriculture Organization of the United Nations; 2003.
2. Qiu Q, Zhang G, Ma T, Qian W, Wang J, Ye Z, Cao C, Hu Q, Kim J, Larkin DM: **The yak genome and adaptation to life at high altitude**. *Nature Genetics* 2012, **44**(8):946-949.

3. Nichols K, Doelman J, Kim JJM, Carson M, Metcalf JA, Cant JP: **Exogenous essential amino acidsstimulate an adaptive unfolded protein response in the mammary glands of lactating cows**. *J Dairy Sci* 2017, **100**(7):5909-5921.
4. Kanwar JR, Kanwar RK, Sun X, Punj V, Matta H, Morley SM, Parratt A, Puri M, Sehgal R: **Molecular and biotechnological advances inmilk proteins in relation to human health.** *Curr Protein Pept Sci* 2009, **10**(4):308-38.
5. Osorio JS, Lohakare J, Bionaz M: **Biosynthesis ofmilk fat, protein, and lactose: roles of transcriptional and posttranscriptional regulation**. *Physiol Genomics* 2016, **48**(4):231-56.
6. Gene Ontology Consortium: **Expansion of theGene Ontology knowledgebase and resources**. *Nucleic Acids Res* 2017, **45**(D1):D331-D338.
7. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto SC, Wu CH, Xenarios I, Yeh LS, Young SY, Mitchell AL: **InterPro in 2017-beyond protein family and domain annotations**. *Nucleic Acids Res* 2017, **45**(D1):D190-D199.
8. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K: **KEGG: new perspectives on genomes, pathways, diseases and drugs**. *Nucleic Acids Res* 2017, **45**(D1):D353-D361.
9. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L, Xenarios I: **UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View**. *Methods Mol Biol* 2016, **1374**:23-54.
10. Junker V, Contrino S, Fleischmann W, Hermjakob H, Lang F, Magrane M, Martin MJ, Mitaritonna N, O'Donovan C, Apweiler R: **The role SWISS-PROT and TrEMBL play in the genome research environment.** *J Biotechnol* 2000, **78**(3):221-34.
11. Burgess, D.J: **Gene expression: Principles of gene regulation across tissues.** *Nature Reviews Genetics,* 2017, **18**(12):701-701.
12. Preston GM, Brodsky JL: **The evolving role of ubiquitin modification in endoplasmic reticulum-associated degradation**. *Biochem J* 2017, **474**(4):445-469.
13. McCaffrey K, Braakman I: **Protein quality control at the endoplasmic reticulum**. *Essays Biochem* 2016, **60**(2):227-235
14. Langfelder P, Horvath S: WGCNA: **an R package for weighted correlation network analysis**. *BMC Bioinformatics* 2008, **9**:559.
15. Liu X, Hu AX, Zhao JL, Chen FL: **Identification of Key Gene Modules in Human Osteosarcoma by Co-Expression Analysis Weighted Gene Co-Expression Network Analysis (WGCNA)**. *J Cell Biochem* 2017, **118**(11):3953-3959.
16. Hu H, Miao YR, Jia LH, Yu QY, Zhang Q, Guo AY: **AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors**. *Nucleic Acids Res* 2019, **7**(D1):D33-D38.

17. Sun LD, Xiao FL, Li Y, Zhou WM, Tang HY, Tang XF, Zhang H, Schaarschmidt H, Zuo XB, Foelster-Holst R, He SM, Shi M, Liu Q, Lv YM, Chen XL, Zhu KJ, Guo YF, Hu DY, Li M, Li M, Zhang YH, Zhang X, Tang JP, Guo BR, Wang H, Liu Y, Zou XY, Zhou FS, Liu XY, Chen G, Ma L, Zhang SM, Jiang AP, Zheng XD, Gao XH, Li P, Tu CX, Yin XY, Han XP, Ren YQ, Song SP, Lu ZY, Zhang XL, Cui Y, Chang J, Gao M, Luo XY, Wang PG, Dai X, Su W, Li H, Shen CP, Liu SX, Feng XB, Yang CJ, Lin GS, Wang ZX, Huang JQ, Fan X, Wang Y, Bao YX, Yang S, Liu JJ, Franke A, Weidinger S, Yao ZR, Zhang XJ: **Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population**. *Nat Genet* 2011, **43**(7):690-4.
18. Liao D: **Emerging roles of the EBF family of transcription factors in tumor suppression**. *Mol Cancer Res* 2009, **7**(12):1893-901.
19. Jargosch M, Kröger S, Gralinska E, Klotz U, Fang Z, Chen W, Leser U, Selbig J, Groth D, Baumgrass R: **Data integration for identification of important transcription factors of STAT6-mediated cell fate decisions**. *Genet Mol Res* 2016, **15**(2).
20. Moniz S, Biddlestone J, Rocha S: **Grow2: the HIF system, energy homeostasis and the cell cycle**. *Histol Histopathol* 2014, **29**(5):589-600.
21. van der Bliek AM, Sedensky MM, Morgan PG: **Cell Biology of theMitochondrion**. *Genetics* 2017, **207**(3):843-871.
22. Ganesan V, Sivanesan D, Yoon S: **Correlation between the Structure and Catalytic Activity of [Cp*Rh(Substituted Bipyridine)] Complexes for NADH Regeneration**. *Inorg Chem* 2017, **56**(3):1366-1374.
23. Li T, Liu J, Smith WW: **Synphilin-1 binds ATP and regulates intracellular energy status**. *PLoS One* 2014, **9**(12):e115233.
24. Goody RS: **The significance of the free energy of hydrolysis of GTP for signal-transducing and regulatory GTPases**. *Biophys Chem* 2003, **100**(1-3):535-44.
25. Nunes-Nesi A, Araújo WL, Obata T, Fernie AR: **Regulation of the mitochondrial tricarboxylic acid cycle**. *Curr Opin Plant Biol* 2013, **16**(3):335-43.
26. Bal NC, Singh S, Reis FCG, Maurya SK, Pani S, Rowland LA, Periasamy M: **Both brown adipose tissue and skeletal musclethermogenesis processes are activated during mild to severe cold adaptation in mice**. *J Biol Chem* 2017, **292**(40):16616-16625.
27. Xiaogang Cui, Yali Hou, Shaohua Yang, Yan Xie, Shengli Zhang, Yuan Zhang, Qin Zhang, Xuemei Lu, George E Liu, Dongxiao Sun: **Transcriptional profiling of mammary gland in Holstein cows with extremely different milk protein and fat percentage using RNA sequencing**.*BMC Genomics*. 2014, **15**:226.
28. Suárez-Vega A, Gutiérrez-Gil B, Klopp C, Tosser-Klopp G, Arranz JJ: **Comprehensive RNA-Seq profiling to evaluate lactating sheep mammary gland transcriptome**.*Sci Data*. 2016, **3**:160051.
29. Mobuchon L, Marthey S, Boussaha M, Le Guillou S, Leroux C, Le Provost F: **Annotation of the goat genome using next generation sequencing of microRNA expressed by the lactating mammary gland: comparison of three approaches**. *BMC Genomics*. 2015, **16**(1):285.

30. Yang Zhou, Erin E Connor, Derek M Bickhart, Congjun Li, Ransom L Baldwin, Steven G Schroeder, Benjamin D Rosen, Liguo Yang, Curtis P Van Tassell, George E Liu: **Comparative whole genome DNA methylation profiling of cattle sperm and somatic tissues reveals striking hypomethylated patterns in sperm**. *Gigascience.* 2018, **7**(5):giy039.
31. Zhixiong Li, Mingfeng Jiang: **Metabolomic profiles in yak mammary gland tissue during the lactation cycle**. *PLoS One* 2019, **14**(7):e0219220.
32. Qiumei Ji, Jinwei Xin, Zhixin Chai, Chengfu Zhang, Dawa Yangla, Luo Sang, Qiang Zhang, Pingcuo Zhandui, Minsheng Peng, Yong Zhu, Hanwen Cao, Hui Wang, Jianlin Han, Jincheng Zhong: **A chromosome-scale reference genome and genome-wide genetic variations elucidate adaptation in yak**. *Molecular Ecology Resources*, DOI:10.1111/1755-0998.13236. In press.
33. Xi Y, Li W: **BSMAP: whole genome bisulfite sequence MAPping program**. *BMC Bioinformatics* 2009, **10**(1):232-232.
34. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon GC, Tontifilippini J, Nery JR, Lee LK, Ye Z, Ngo Q: **Human DNA methylomes at base resolution show widespread epigenomic differences**. *Nature* 2009, **462**(7271):315-322.
35. Juhling F, Kretzmer H, Bernhart SH, Otto C, Stadler PF, Hoffmann S: **metilene: Fast and sensitive calling of differentially methylated regions from bisulfite sequencing data**. *Genome Research* 2016, **26**(2):256-262.
36. Wang, J., Duncan, D., Shi, Z., Zhang, B: **WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update.** *Nucleic Acids Res* **2013**, 41: W77-83.
37. Pertea M, Pertea G, Antonescu C, Chang TC, Mendell JT, Salzberg SL: **StringTie enables improved reconstruction of a transcriptome from RNA-seq reads**. *Nature Biotechnology* 2015, **33**(3):290-295.
38. Robinson MD, Mccarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinformatics* 2010, 26(1):139-140.

Figures

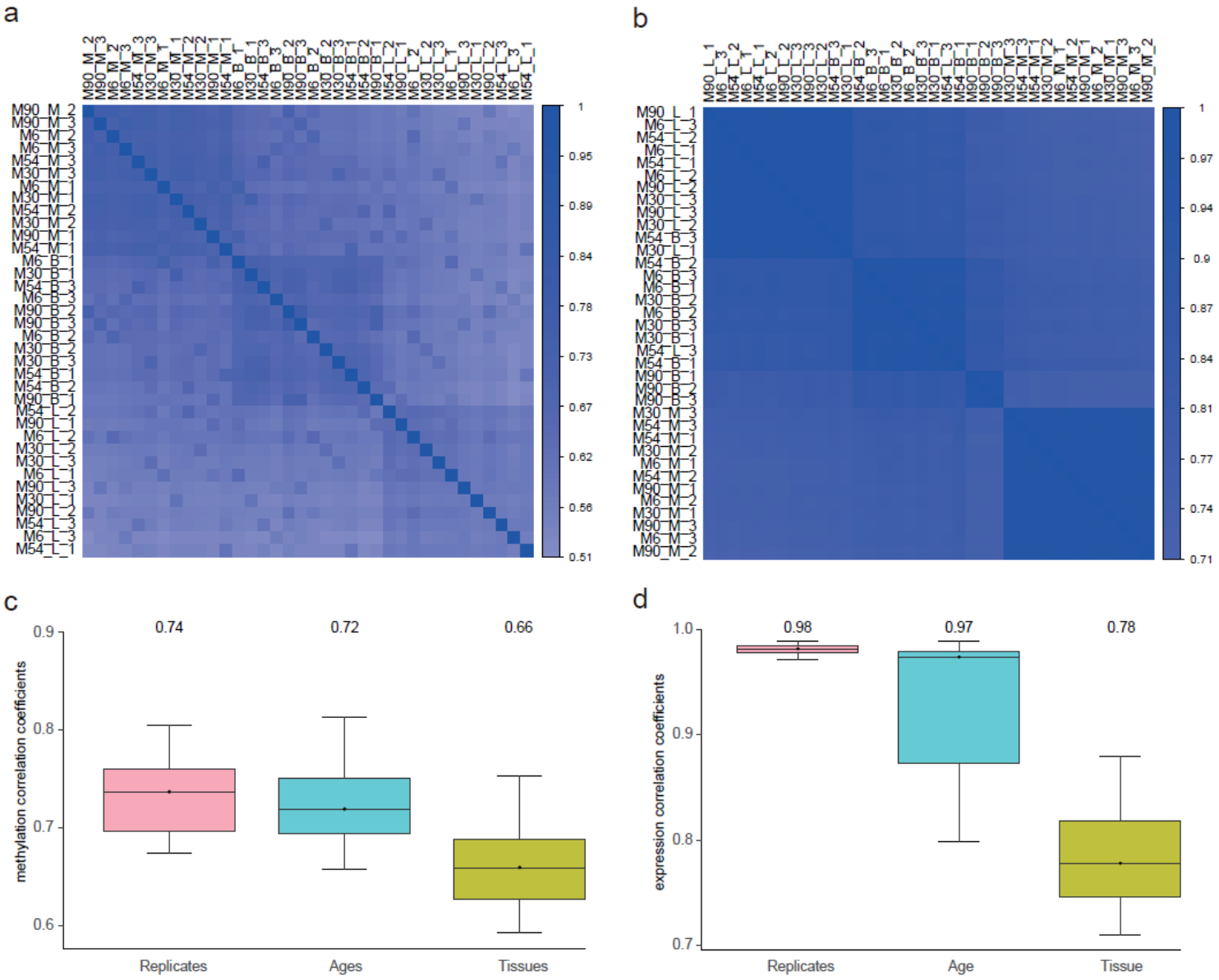

Figure 1

Global DNA methylation and gene expression among samples. Pearson's correlation analysis based on the methylation of CpG sites (a) and gene expression (b) among samples. Boxplot of Pearson's correlation coefficients between replicates, ages, or tissues for methylation (c) and gene expression (d). M6, M30, M54 and M90 represent different month ages. B, L and M represent breast, lung and gluteal muscle, respectively. 1, 2 and 3 represent different replicates.

A

Breast	M06	M30	M54	M90
M06	0	20,952	23,868	62,576
M30	20,952	0	12,347	46,898
M54	23,868	12,347	0	46,483
M90	62,576	46,898	46,483	0

Lung	M06	M30	M54	M90
M06	0	22,108	22,580	27,140
M30	22,108	0	20,208	22,231
M54	22,580	20,208	0	22,060
M90	27,140	22,231	22,060	0

Muscle	M06	M30	M54	M90
M06	0	15,025	13,538	14,427
M30	15,025	0	11,688	12,339
M54	13,538	11,688	0	15,882
M90	14,427	12,339	15,882	0

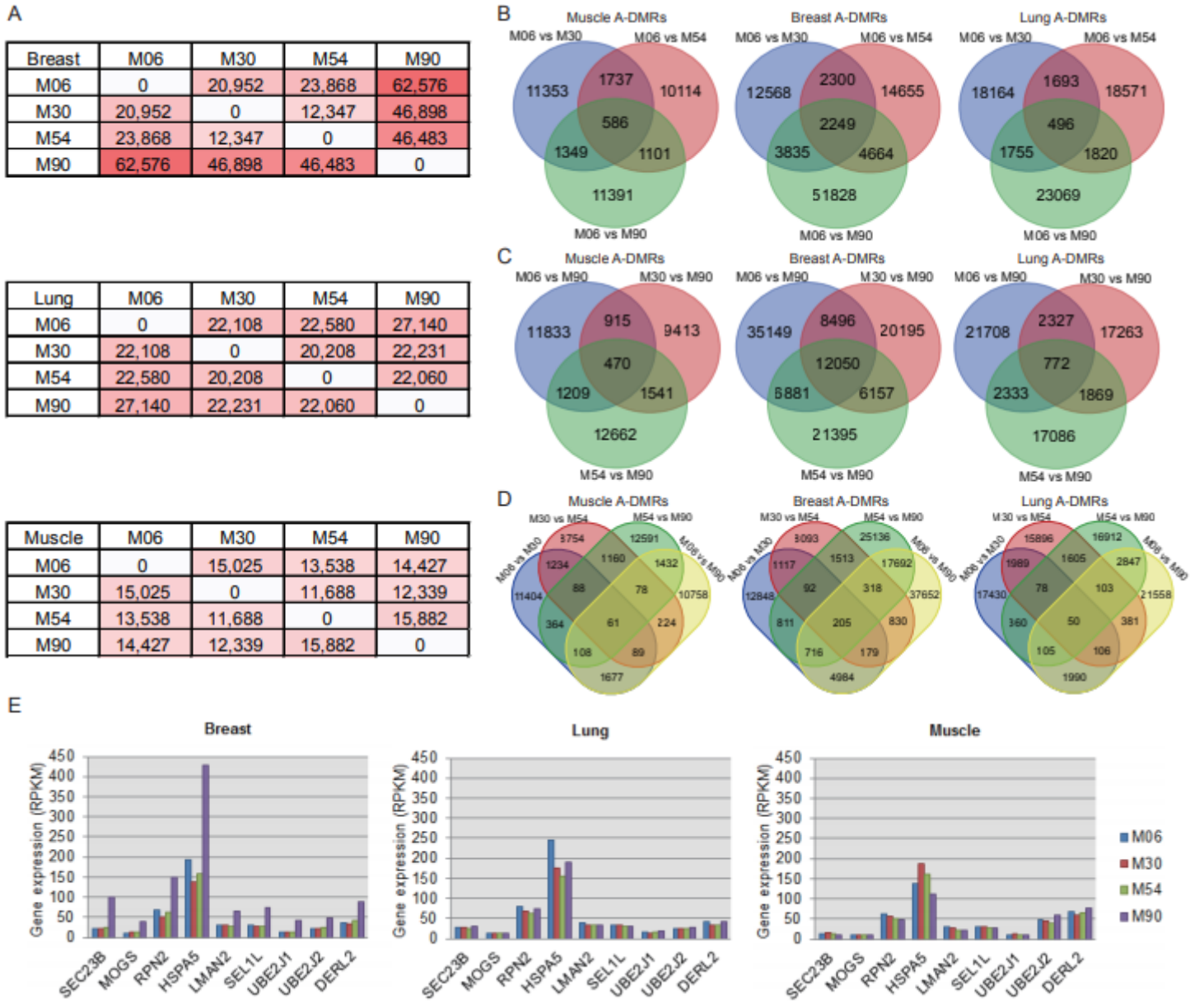

Figure 2

Overview of month age-associated DMRs. (a) Basic statistics for A-DMRs within each tissue. Overlap of A-DMRs associated with the 6 months age group (b), 90 months age group (c), and 30 and 54 months age groups (d) in the muscle, breast, and lung respectively.

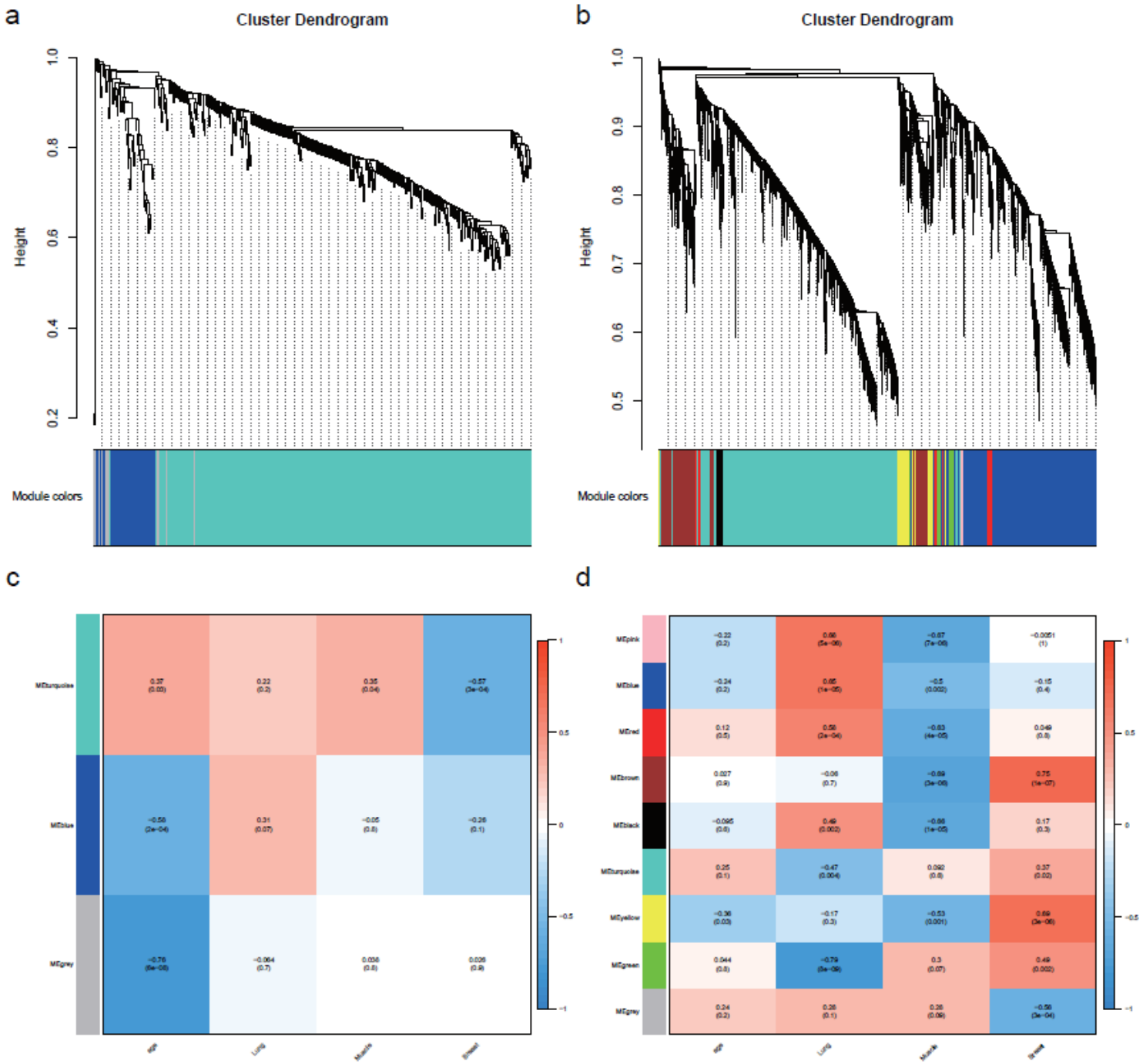

Figure 3

Modules of consensus networks and correlation with traits. Consensus networks from the age (a) or tissue (b) curve. Gene expression similarity was determined using a pair-wise weighted correlation metric and clustered according to a topological overlap metric into modules; assigned modules are colored on the bottom, and gray genes were not assigned to any module. Consensus network modules for age (c) and tissue (d) correlated with traits using the eigenmodule (the first principal component of the module). The correlation coefficients along with the p-value in parenthesis are provided underneath; color-coding refers to the correlation coefficient (legend at right).

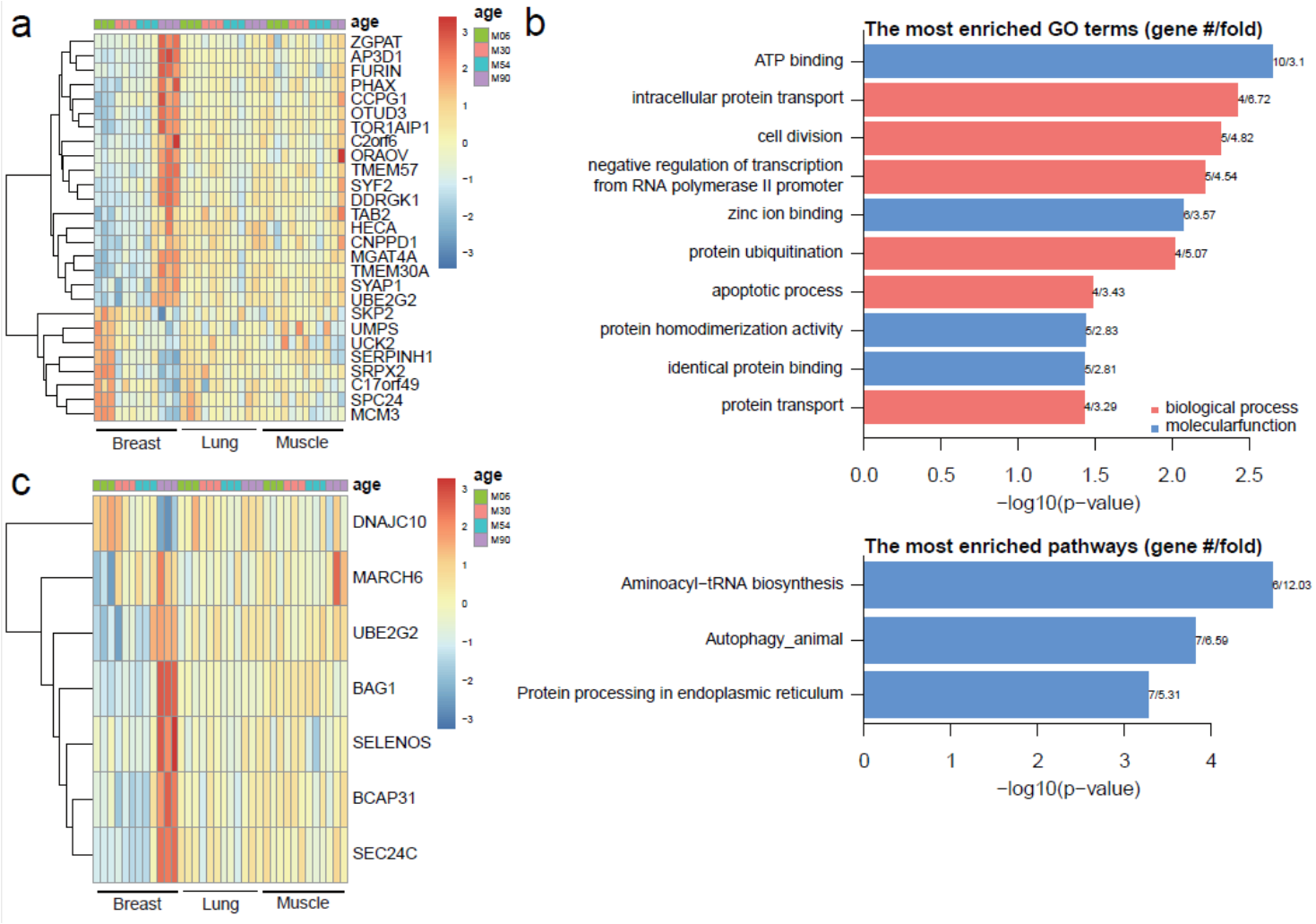

Figure 4

“Hub” genes and potential target genes of ZGPAT in the age network. (a) Expression level of 27 “hub” genes. (b) Enrichment analysis of ZGPAT’s potential target genes. (c) Expression level of 7 genes in “protein processing in endoplasmic reticulum”, which was enriched from potential target genes of ZGPAT.

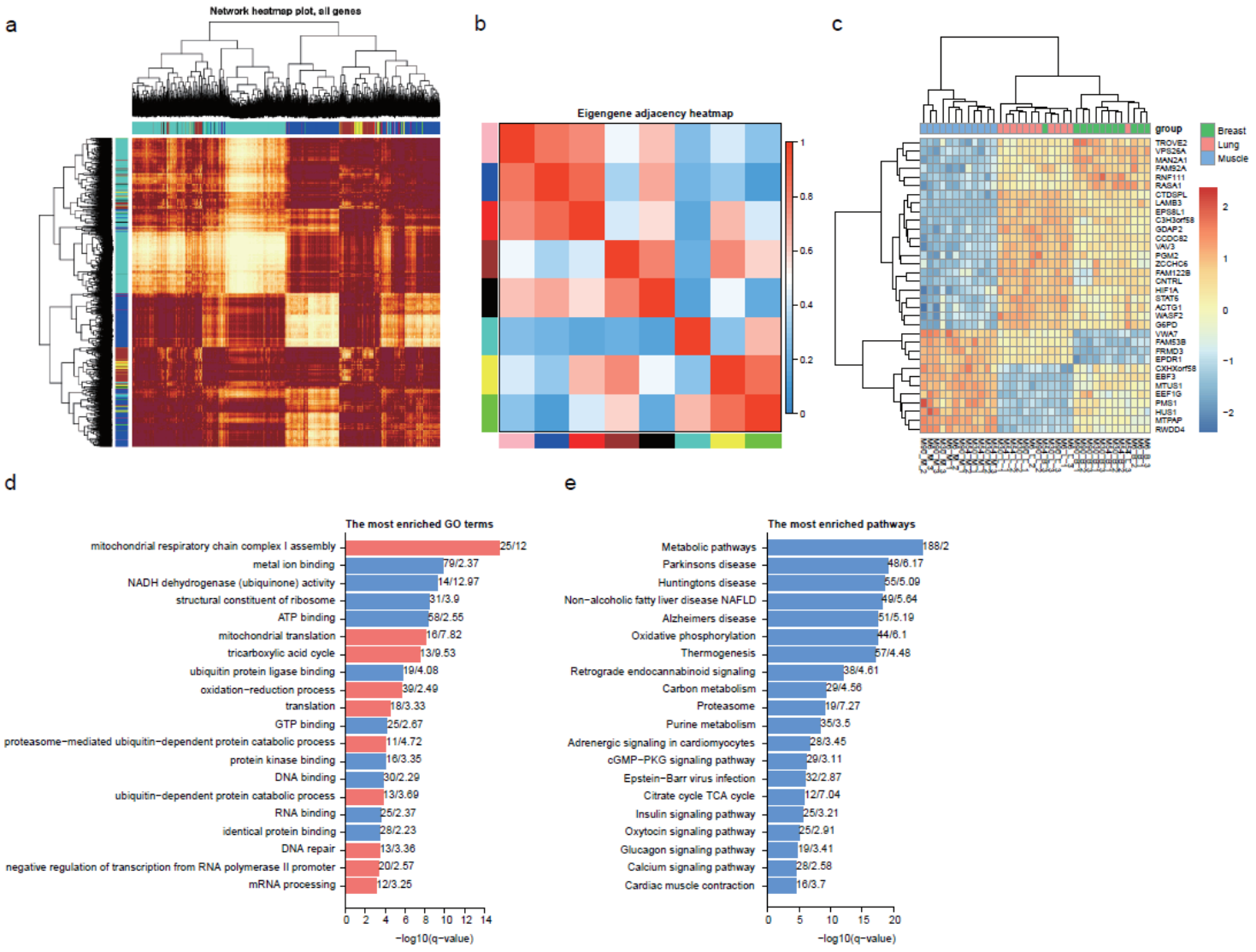

Figure 5

Modules and “hub” genes in the tissue network. (a) WGCNA modules of the tissue-related genes, (b) correlations between modules showed by the eigenmodule adjacency heatmap, (c) expression level of “hub” genes in the tissue network, (d) enrichment analysis of potential target genes of HIF1A, and the number of enriched genes and enrichment fold were indicated on the right.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ARRIVEchecklist.docx
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx

- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx