KLF4, a Key Regulator of a Transitive Triplet, Acts on the TGF-β Signaling Pathway and Contributes to High-Altitude Adaptation of Tibetan Pigs

Tao Wang
Northeast Agricultural University
Yuanyuan Guo
Northeast Agricultural University
Shengwei Liu
Northeast Agricultural University
Chaoxin Zhang
Northeast Agricultural University
Tongyan Cui
Northeast Agricultural University
Kun Ding
Inner Mongolia Normal University
Peng Wang
Heilongjiang Provincial Husbandry Department
Xibiao Wang
Northeast Agricultural University
Zhipeng Wang (wangzhipeng@neau.edu.cn)
School of animal science Northeast Agricultural University https://orcid.org/0000-0002-6041-8906

Research

Keywords: Tibetan pig, Multi-tissue, Transcriptome, Hypoxia adaptation, Gene network

Posted Date: November 5th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-101172/v1

Version of Record: A version of this preprint was published at Frontiers in Genetics on April 15th, 2021.
See the published version at https://doi.org/10.3389/fgene.2021.628192.

KLF4, a Key Regulator of a Transitive Triplet, Acts on the TGF-β Signaling Pathway and Contributes to High-Altitude Adaptation of Tibetan Pigs

Tao Wang[1, 2, a], Yuanyuan Guo[1, 2, a], Shengwei Liu[1, 2], Chaoxin Zhang[1, 2], Tongyan Cui[1,2], Kun Ding[3], Peng Wang[4], Xibiao Wang[1, *], Zhipeng Wang[1, 2, *]

[1]College of Animal Science and Technology, Northeast Agricultural University, Harbin 150000, China;

[2]Bioinformatics Center, Northeast Agricultural University, Harbin 150000, China;

[3]College of Computer Science and Technology, Inner Mongolia Normal University, Hohhot 010000, China;

[4]HeiLongJiang provincial Husbandry Dapartment, Harbin 150000, China.

[a]Equal contribution:

Tao Wang: 2407159664@qq.com

Yuanyuan Guo: 303992052@qq.com

[*]Corresponding Author:

Xibiao Wang: wangxibiao@neau.edu.cn

Zhipeng Wang: wangzhipeng@neau.edu.cn

Abstract

Background: Tibetan pigs are native mammalian species on the Tibetan Plateau that have evolved distinct physiological traits that allow them to tolerate high-altitude hypoxic environments. They can be used as a suitable animal model for exploring the molecular mechanism of hypoxia adaptation in high-altitude organisms.

Results: Here, based on multi-tissue transcriptional data from high-altitude Tibetan pigs and low-altitude Rongchang pigs, we performed a weighted correlation network analysis (WGCNA) and identified key modules related to these tissues. Complex network analysis and bioinformatics analysis were integrated to identify key genes and size-3 network motifs. The results showed that compared to other tissues, the lungs of Tibetan pigs and Rongchang pigs are more significantly different, showing more adaptive transcriptional changes. In the lung tissues of Tibetan pigs, we identified *KLF4, BCL6B, EGR1, EPAS1, SMAD6, SMAD7, KDR, ATOH8* and *CCN1* genes as potential regulators of hypoxia adaption. We found that *KLF4* and *EGR1* genes simultaneously regulate the *BCL6B* gene, forming a *KLF4-EGR1-BCL6B* transitive triplet. This transitive triplet, dominated by *KLF4*, may enhance the hypoxia adaptability of Tibetan pigs by mediating the TGF-β signaling pathway. This triplet also regulates the *KDR* gene, which is involved in the PI3K-Akt signaling pathway and plays an important role in hypoxia adaptation.

Conclusions: We postulate that the *KLF4-EGR1-BCL6B* transitive triplet may contribute to the adaptation of Tibetan pigs to hypoxic environments. These findings provide new details of the regulatory architecture of hypoxia-adaptive genes and are

valuable for understanding the genetic mechanism of hypoxic adaptation in mammals.

Keywords: Tibetan pig; Multi-tissue; Transcriptome; Hypoxia adaptation; Gene network

Background

Hypoxia is a significant environmental characteristic of high altitude, which exerts a marked impact on biological organisms and imposes extreme physiological challenges in mammals. The Tibetan pig was originally distributed at altitudes of 2900-4300 m in the Tibetan Plateau [1]. Physiological studies show that Tibetan pigs have evolved physiological adaptations to high-altitude hypoxia, such as a thicker alveolar septum with more highly developed capillaries [2] and a larger and strong heart [3]. Therefore, they represent a suitable animal model for exploring the molecular mechanism of hypoxia adaptation in high-altitude organisms.

With the development of sequencing technology, at present, the majority of studies have explored the genetic basis of hypoxic adaptation in Tibetan pigs from the perspective of selection signals [1-6] or by using differential expression analysis between differential conditional gene expression in one tissue based on the transcriptome [7,8]. Although previous studies have identified the *EPAS1, HIF1A, EGLN1, RGCC KLF6, TGFB2, EGLN3,* and *ACE* genes as related to hypoxia, these genes may only explain a minority of genetic variance due to the case of the missing heritability. Therefore, the most detailed solution to the missing heritability problem would involve identifying all causal genetic variants [9] and exploring related gene networks that have facilitated high altitude adaptation of Tibetan pigs.

The adaptation of Tibetan pigs to hypoxia is a very complex biological process that may involve multiple genes and transcriptional regulation among genes. The gene network provides a systemic view of gene regulation by the coordinated activity of

multiple genes and regulatory factors and serves as a medium for understanding the mechanism of gene regulation [10]. Based on the gene expression profile, a gene network was constructed by quantitative modeling, which can be used for rational design of molecular approaches to target specific biological processes [11] and infer new biological functions [12,13]. Moreover, the gene network can not only intuitively elucidate the regulatory relationship between genes but also identify important hub genes. These hub genes represent candidates for further experimental investigation and potential biomarkers for complex traits [14-16].

Transcription factors (TFs) and microRNAs (miRNAs) regulate gene expression at the transcriptional and post-transcriptional levels, respectively. They coordinately control the dynamics and output of gene transcription and tightly control spatial and temporal patterns of gene expression. Therefore, constructing a gene regulatory network involving TFs and miRNAs is helpful for understanding the regulatory mechanism of genes in adaptation to hypoxia.

Moreover, most cellular tasks are performed not by individual genes but by groups of functionally associated genes, often referred to as modules. In a gene regulatory network, modules appear as groups of densely interconnected nodes, also called communities or clusters [17]. Among these clusters of gene regulatory networks, size-3 network motifs, such as the feed-forward loop (FFL), have a significant biological function. The FFL motif governs many aspects of normal cell functions, such as creating bistable switches of gene expression in developing tissues for spatial avoidance, controlling the time sequence of gene expression to create temporal

avoidance, and minimizing expression fluctuation against noise [18].

In this study, based on transcriptional data from six tissues in Tibetan pigs and Rongchang pigs, key module of lung tissues were identified by constructing a gene network. Integrating complex network analysis and bioinformatics analysis, we identified key genes and size-3 network motifs and found that *KLF4*, a key regulator of the transitive triplet, may enhance hypoxia adaptability in Tibetan pigs by mediating the TGF-β signaling pathway. This study provides a theoretical basis for further understanding the molecular mechanism of adaptability to high-altitude hypoxia.

Materials and methods

Gene expression data collection

In this study, Tibetan pigs living in a high-altitude hypoxic environment were obtained from Songpan County, Tibet, China (altitude up to 3000 m). As a control, Rongchang pigs were obtained from Chongqing, China (the altitude is approximately 400 m), where the oxygen content in the air is normal. The protein-encoding genes and miRNAs expression profile data of 6 tissues (muscle, liver, heart, spleen, kidney and lung) from three Tibetan pigs and three Rongchang pigs were obtained from the Gene Expression Omnibus (GEO) database at the National Center for Biotechnology Information (NCBI) under accession numbers GSE93855 and GSE124418, respectively. Genes in the dataset were re-annotated based on the *Sus scrofa* 11.1 genome assembly. Taking into account that genes with very low expression are less

reliable and indistinguishable from sampling noise, we selected the top 50% of protein-encoding genes of the median absolute deviation (MAD) of expression level.

Co-expression network analysis

Network analysis was performed according to the protocol of the WGCNA R package (version 1.68) [19]. First, cluster analysis of Tibetan pig and Rongchang pig samples based on the "hclust" function in WGCNA was used to verify the clustering of samples and to detect outliers. Then, the soft threshold power β was obtained to meet the scale-free topology criterion [20]. Based on β, the Pearson correlation matrix between genes was transformed into an adjacency matrix:

$$\alpha_{ij} = \left|r_{ij}\right|^{\beta} \quad [1]$$

The topological overlap measure (TOM) representing the overlap in shared neighbors was calculated using the adjacency matrix:

$$TOM_{i,j} = \frac{\sum_{K=1}^{N} \alpha_{i,k} \cdot \alpha_{k,j} + \alpha_{i,j}}{\min(K_i, K_j) + 1 - \alpha_{i,j}} \quad [2]$$

where α is the adjacency matrix given by formula [1]. Based on the TOM matrix, genes with similar expression profiles were classified into the same modules using hierarchical clustering and dynamic branch cutting procedures. Relationships between modules can be studied using the correlation between module eigengenes. Here, we merged modules with a correlation higher than 0.9. The number of genes in the merged module should be more than 50.

We used the following criteria to identify the key module of each tissue: (1) the p-value of the correlation between the module and the tissue was less than $3.97 \times$

10^{-4} (0.05/126); and (2) the median of the key module gene significance (GS) value was greater than 0.8. GS is a parameter to characterize the correlation between genes and phenotypic traits. The higher the absolute value of GS_i is, the more biologically significant the i-th gene [19].

In addition, we calculated the fundamental topology concepts of each key module, including density, mean cluster coefficient, centralization and heterogeneity.

Analysis of gene expression patterns in multiple tissues

In this study, we used the Mfuzz package in R [21] to identify multi-tissue expression patterns of each gene in each key module. Based on the fuzzy c-means algorithm, this software implements soft clustering methods for microarray data analysis, which makes the clustering process less sensitive to noise and effectively reflects the strength of a gene's association with a given cluster.

Gene tissue-specific analysis

We used the tissue-specificity index (TSI, τ) [22] to grade the scalar measure of the specificity of an expression profile, which ranged from 0 for housekeeping genes to 1 for strictly TS genes. The index τ was defined as follows:

$$\tau = \frac{\sum_{i=1}^{N}(1-x_i)}{N-1} \quad [3]$$

where N is the number of tissues and x_i is the expression normalized by the maximal component value. According to Yania et al. (2005) [22], genes with TSI> 0.9 were considered TS genes.

157

158 **Functional enrichment analysis of genes in key modules**

159 We used the online software DAVID (v6.8) [23] to perform functional enrichment
160 analysis of genes in each key module, including gene ontology (GO) and KEGG
161 pathway analysis.

162

163 **Identification of hub genes in key modules**

164 For each module, Langfelder et al. (2008) [19] define a quantitative measure of
165 module membership (MM) as the correlation of the module eigengene and the gene
166 expression profile. The MM of gene i in module q can be defined as follows:

167 $$\mathrm{MM}^{\mathrm{q}} = \mathrm{K}_{\mathrm{cor,i}}^{(\mathrm{q})} = \mathrm{cor}\left(\mathrm{x_i}, \mathrm{E}^{(\mathrm{q})}\right) \quad [4]$$

168 where $\mathrm{x_i}$ is the profile of gene i and $\mathrm{E}^{(\mathrm{q})}$ is the module eigengene of module q.
169 If MM_i^q is close to 0, the i-th gene is not part of the q module. On the other hand,
170 if MM_i^q is close to 1 or -1, it is highly connected to the q module genes.

171 In co-expression networks, the connectivity ($\mathrm{k_i}$) is defined as the sum of
172 connection strengths with the other genes:

173 $$\mathrm{k_i} = \sum_{\mu \neq \mathrm{i}} \mathrm{a_{\mu i}} \quad [5]$$

174 The Kwithin of a gene is the sum of the connectivity of this gene in the module. We
175 identified the hub genes in each key module according to the following criteria: (1)
176 GS value of the gene ≥ 0.8; (2) MM value of the gene ≥ 0.95; and (3) in each module,
177 Kwithin ranked in the top 20% of genes.

178

Gene regulatory network construction

First, we removed the co-expression relationship with a weight value of less than 0.2 in the network of the six key modules of Tibetan pigs. Using the AnimalTFDB database [24], we obtained the TFs in each key module. The biomaRt package of R [25] was used to obtain the sequence of the 1 kb region upstream of the transcription start site of all protein-encoding genes in the pig genome. The TF position weight matrix (PWM) of pigs was obtained from the CIS-BP database [26]. Using the TFBSTools package in R [27] to predict the target genes of TFs, the relScore value was set to 0.85, and other parameters were defaulted. Next, the biomaRt package was used to obtain the sequence of the 3'UTR region of pig protein coding protein genes. We obtained all mature miRNA sequences from the miRBase database [28]. Based on the miRanda tool [29], we predicted target genes of the miRNAs, and the Tot Score and Tot Energy were set to 140 and -20, respectively. Finally, the gene regulatory network in each Tibetan pig tissue was constructed by combining TFs, miRNAs, target genes, co-expressed genes, hub genes and their interactions.

Motif analysis of the gene regulatory network

Gene networks may contain various subgraphs, and the detection of motifs contributes to identifying the typical local connection pattern [30-33]. The 3-node motifs in the gene regulatory network of each tissue were obtained using mfinder1.2 [34]. Mfinder1.2 implements a switching method to generate random network, which can switch between edges while maintaining the number of incoming edges, outgoing

edges and mutual edges of each node of the input network. In this study, the number of random networks was set to 10000. Moreover, mfinder1.2 describes the significance of the difference between the frequency of motifs in the real network and that in the corresponding randomized network using the Z-test in statistics. The Z-test is defined as follows:

$$Z_j \triangleq Z(j) = \frac{N(j) - \overline{N_r\ (j)}}{\sigma_r(j)} \quad [6]$$

where $N(j)$ is the number of times the subgraph appears in the real network, and $\overline{N_r(j)}$ and $\sigma_r(j)$ are the mean and standard deviation of its appearances in the randomized network ensemble. The larger the absolute value of Z is, the more significant the difference. The significance profile (SP) is the vector of Z scores normalized to length 1, describing the statistical significance of each motif in the network [35]:

$$SP_i = \frac{Z_i}{(\sum Z_i^2)^{1/2}} \quad [7]$$

We constructed the triad significance profile (TSP) of the six tissues from Tibetan pigs, which display certain relations between subgraph types.

Identification of important genes and size-3 subgraphs in the lung-specific gene regulation network

Each node was scored according to the connectivity, differential expression between different conditions, tissue-specific expression and TF characteristics using the following formula [8]:

$$S_{node_i} = \omega_i K_i q_i TSI_i \quad \omega = \begin{cases} 0.5 & the\ number\ of\ TG\ of\ TF \geq \overline{TGs} \\ 0.3 & the\ number\ of\ TG\ of\ TF < \overline{TGs} \\ 0.2 & Non\ TFs \end{cases}$$

[8]

Where K_i is the scaled connectivity of each gene in the regulatory network, and K_i of the i-th node is defined as follows:

$$K_i = \frac{Connectivity_i}{\max(Connectivity)} \quad [9]$$

q_i is the estimated probability of differentially expressed genes in lung tissues between Tibetan pigs and Rongchang pigs, calculated by NOISeq [36]; TSI_i is the tissue-specificity index of the gene; and ω_i is the weighting coefficient. $\overline{TGs}$ is the average number of target genes regulated by TFs. If the TF regulated more than $\overline{TGs}$, ω is set to 0.5, the target gene is less than $\overline{TGs}$, ω is set to 0.3. The ω of non-TF genes is set to 0.2.

The score of each candidate size-3 subgraph was calculated by combining the node score and the edge score as follows:

$$S_{motif_i} = \frac{\sum_{node \in motif} S_{node}}{\sqrt{n_{node}}} + \frac{\sum_{edge \in motif} S_{edge}}{\sqrt{n_{edge}}} \quad [10]$$

where S_{edge} denotes the score of each edge, which was the weight value of the edge from WGCNA, and n_{node} and n_{edge} are the number of nodes and edges in the motif, respectively.

Verification of important genes in lung tissue

The lung tissue expression profile of Tibetan sheep and yak was obtained from the GEO database (accession: GSE93855), the expression profile of Diqing Tibetan pig lung tissue from another dataset (accession: GSE84409), and WGCNA was performed.

The Hmisc package in R was used to statistically test the correlation between genes.

Results

WGCNA and identification of key modules in tissues

We calculated the MAD of protein-encoding gene expression in this study. A total of 5723 protein-encoding genes in the top 50% of MAD values were selected for subsequent analysis. Cluster analysis revealed that different samples from the same tissue of Tibetan pigs and Rongchang pigs clustered together, and no outlier samples were observed, as shown in Fig. 1.

We constructed a co-expression network for Tibetan pigs and Rongchang pigs. To fulfill the criteria of approximate scale-free topology, the soft threshold power β was set to 20 (the scale-free topological index $R^2 = 0.85$ for Tibetan pigs (Fig. 2a) and $R^2 = 0.80$ for Rongchang pigs (Additional file 1: Figure S1 a)). Through hierarchic clustering and dynamic branch cutting procedures, 36 modules were identified in the Tibetan pig co-expression network. According to the similarity between modules, the modules with a correlation higher than 0.9 were merged, and 21 merged modules were ultimately obtained. Clustering of the modules is shown in Fig. 2b.

Next, the GS values of genes contained in these modules and the correlation between each module and different tissues in Tibetan pigs (Fig. 2c) were calculated. According to the screening criteria, key modules from six tissues in Tibetan pigs were determined. Among them, there was only one key module in muscle, liver, heart, spleen and kidney, which are designated M1, M5, M2, M9 and M20, respectively.

These modules contained 267, 215, 157, 201 and 420 genes, respectively. There were three key modules (M12, M13 and M14) in the lung tissue of Tibetan pigs. Since the correlation between these three modules was close to 0.9, we merged them into a single module representing the key module of the lung, named module 22 (M22), which contained 350 genes. The co-expression network of six key tissue modules of Tibetan pigs was visualized by using Cytoscape v3.8.0 software, as shown in additional file 2: Figure S2.

For the co-expression network of Rongchang pigs, 20 merged modules were ultimately obtained. According to key module screening criteria, key modules of muscle, liver, heart, spleen, kidney and lung were designated M14, M8, M21, M3, M13 and M1, respectively. These modules contained 335, 201, 269, 940, 367 and 126 genes, respectively, as shown in additional file 1: Figure S1.

Network topology analysis

We calculated the network topology of each tissue key module from the Tibetan pigs and Rongchang pigs, including density, mean cluster coefficient, centralization and heterogeneity. Results are shown in Table 1. Among them, the density and mean cluster coefficient describe the cohesive characteristics of the network. We observed that the network density of Tibetan pig lung and heart tissue was the lowest (0.03), and the clustering coefficient was the lowest (0.13-0.14), while the network density (0.12) and clustering coefficient (0.28) of the spleen were the highest. These network concepts indicate that the key modules of the lung and heart were a sparse network,

while the densification of the spleen tissue network was higher than that of the other five tissues.

Heterogeneity and centralization describe the distribution of connectivity (degree) in the network. Generally, if the network is highly heterogeneous, its centralization will be low. Moreover, the higher the heterogeneity of the network, the more uneven the distribution of degree in the network, that is, only a few nodes in the network have high connectivity, while most other nodes have low connectivity. We found that the network heterogeneity of key modules in six tissues from Tibetan pigs was ≥ 0.8, and the centralization was < 0.2, indicating that the degree of distribution of the key module network determined by us in each tissue was approximately scale-free.

The network density, centralization, heterogeneity, and mean cluster coefficient of the six tissues from Rongchang pigs were similar to those of Tibetan pigs.

Multi-tissues gene expression patterns

According to the analysis of gene expression patterns, we found that compared to other tissues, the expression patterns of key module genes in lung tissues were the most significantly different between Tibetan pigs and Rongchang pigs. Genes in the Tibetan pig lung key module were highly expressed in lung tissue compared to other tissues and divided into 8 clusters based on their multi-tissues expression pattern. Genes in cluster 2 and cluster 8 had the second-highest expression in heart tissue (Fig. 3a). For genes in the Rongchang pig lung key module, some genes were expressed the highest in lung tissue, but other genes were expressed the highest in spleen tissue.

These genes were also divided into 8 clusters based on their multi-tissue expression pattern for Rongchang pigs (Fig. 3b).

Tissue-specific gene analysis

A total of 266 genes (4.65%) were identified as tissue-specific ($\tau > 0.9$) in Tibetan pigs. Among them, there were 32, 50, 23, 36, 47, and 22 TS genes in the key modules of muscle, liver, heart, spleen, kidney and lung, respectively, accounting for 80% of the total specific genes (210/266). TS genes exhibited the highest expression levels in one tissue compared to other tissues.

In Rongchang pigs, a total of 206 TS genes were detected. There were 39, 51, 21, 29, 45 and 8 TS genes in muscle, liver, heart, spleen, kidney and lung, respectively, and 31, 41, 19, 22, 35 and 4 overlapped with Tibetan pigs corresponding to the same tissues, respectively. Among them, the lung tissue presented the greatest difference. There were more TS genes in Tibetan lung tissue than in Rongchang pig lung tissue.

Functional enrichment analysis of genes in key modules

To further understand the biological functions of genes in each key module in Tibetan pigs and Rongchang pigs, we conducted gene function enrichment analysis. After the Benjamini correction, we identified significant pathway enrichment in five tissues, except for kidney tissue in Tibetan pigs. Compared to Rongchang pigs, there were 10, 4, 1, and 13 pathways in muscle, lung, heart, and spleen that were only significantly enriched in Tibetan pigs, as shown in Table 2.

Pathways enriched only in Tibetan pig muscles, including some signaling pathways related to the hypoxic stress response and energy metabolism homeostasis, such as the AMPK signaling pathway (ssc04152), proteasome (ssc03050) and adrenergic signaling in cardiomyocytes (ssc04261). Pathways enriched in lung tissue regulated cell growth, proliferation, migration and apoptosis include focal adhesion (ssc04510), ECM-receptor interaction (ssc04512), PI3K-Akt signaling pathway (ssc04151) and TGF-β signaling pathway (ssc04350). In the heart, Tibetan pigs were enriched in arrhythmogenic right ventricular cardiomyopathy (ssc05412). In the spleen, most significantly enriched pathways were related to the protein translation process and ribosomes.

Identification of hub genes in key modules

According to the screening criteria of hub genes, 23, 41, 20, 40, 81 and 31 hub genes were identified in the muscle, liver, heart, spleen, kidney and lung, respectively, in Tibetan pigs. In addition, 61, 39, 26, 123, 68, and 14 hub genes were identified in the same six tissues, respectively, of Rongchang pigs. Compared to Rongchang pigs, Tibetan pigs had more hub genes in liver, kidney and lung tissues. There was no hub gene overlap between the lung tissues of Tibetan pigs and Rongchang pigs. There were 2 and 6 overlapping hub genes in the heart and spleen, respectively, while 22, 20 and 45 overlapping hub genes were found in muscle, liver and kidney, respectively. In addition, 11, 30, 8, 20, 26, and 8 hub genes were TS in six tissues of Tibetan pigs. However, the number of TS genes in the heart, spleen and lung of Rongchang pigs

was only 3, 0 and 1, respectively. Table 3 summarizes the hub gene information in Tibetan pigs and Rongchang pigs.

Gene regulatory network construction

The gene regulatory network of Tibetan pig tissues was constructed by combining TFs, miRNAs, target genes, co-expression genes, and hub genes. There were 115, 80, 35, 117, 160, and 157 nodes (genes) and 986, 1786, 298, 1976, 5315 and 1075 edges (regulatory relationship) in the gene regulatory network of muscle, liver, heart, spleen, kidney and lung, respectively, as shown in Fig. 4. There were 9, 3, 1, 3, 3, and 16 TFs, respectively, in the gene regulatory network of each tissue. In total, 35 TFs belonged to 10 TF families, among which 10 TFs were also hub genes. According to the PWM provided by the CIS-BP database, 20 TFs target genes were predicted. We found that these 20 TFs regulate 237 genes (94 genes are hub genes) in each tissue key modules, predicting a total of 408 regulatory relationships.

Through the prediction of miRNA target genes, we found that genes in the key modules of muscle, liver, heart, spleen, kidney and lung were regulated by 8, 3, 3, 2, 4, and 6 miRNAs, respectively. Table 4 summarizes the information on TFs, miRNAs, target genes and hub genes in the gene regulatory network of six tissues in Tibetan pigs.

Identification of gene regulatory network motifs

In gene networks, some motifs displayed much higher frequencies than expected in

randomized networks [30,37], and these motifs were suggested to be recurring circuit elements that perform key information-processing tasks [37-40]. Among them, the motif composed of three nodes contains 13 kinds, including V-out, 3-Chain, FFL, 3-Loop, Clique and so on. Using mfinder1.2 software, we identified 8894, 13067, 993, 19899, 78959 and 14692 motifs in muscle, liver, heart, spleen, kidney and lung tissue gene regulatory networks in Tibetan pigs, respectively. There were significant differences in the distribution of motifs among different gene regulatory networks ($p < 2.2e-16$).

In the lung and muscle gene regulatory network, 12 types of 3-node motifs were found, excluding the 3-loop motif. Especially, in the lung tissue gene regulatory network, there were 5160, 18, 133, 380, 3152, 3098, 135, 810, 88, 28, 382, and 1308 motifs of V-out, V-in, 3-Chian, Mutual in, Mutual out, Mutual V, FFL, Regulated mutual, Regulating mutual, Mutual and 3-Chain, Semi clique and Clique, respectively. In liver, heart, spleen and kidney gene regulatory networks, there were 7, 5, 2, and 7 kinds of 3-node motifs, respectively. The Clique motif is the most frequent motif in liver and spleen gene regulatory networks, and Mutual V and Clique motifs are primarily found in the heart and kidney. The motif information in the six tissue gene regulatory networks is shown in Table 5.

To analyze the statistical significance of each motif type, we generated 10,000 random networks representing a conservation rule. The value of the constant in each of the randomized networks that conserves the degree sequence is equal to its value in the real network. TSP analysis was performed on each motif in the six tissue gene

regulatory networks. The distribution of TSP in the lung tissue gene regulatory network is shown in Fig. 5. Any network with significant deviations from randomness in its local structure will have a Z-score vector with a standard deviation larger than one. We found that the frequency of FFL, Regulated mutual, Regulating mutual and Clique motif in the lung tissue gene regulatory network was significantly different from that of random networks ($p<1E-04$). In the muscle and heart tissue gene regulatory network, Regulated mutual and Clique motif were significant motif types, while V-out, Semi clique and Clique motif were significant in the kidney gene regulatory network.

Motif analysis of the gene regulatory network in lung tissue

An ordered triplet of nodes (x, y, z) is transitive if x→y, y→z and x→z. An ordered triplet of nodes is intransitive, as defined for example by Harary and Kommel [41], if x→y, y→z but no edge is directed from x to z. There were different numbers of transitive and intransitive triplets for all 13 triad subgraphs. Among them, the FFL motif is a classical type of transitive triplet, and Regulating mutual, Mutual and 3-Chain, Regulated mutual, and Clique have 2, 1, 2, and 6 transitive triplets, respectively.

All FFL motifs in the lung tissue gene regulatory network were $TF_1 \rightarrow TF_2$, including *KLF4→EPAS1, KLF4→BCL6B, KLF4→FOS, EGR1→BCL6B, EGR1→EPAS1, BCL6B→EPAS1, TBX3→EPAS1,* and *TBX3→BCL6B*. Then, the two TFs shared a target gene. As a result, 51 target genes were regulated, including 4 TFs,

forming 13 transitive triplets, and 21 hub genes, forming 71 transitive triplets. In addition, three of these target genes were both TF and hub gene, forming 8 transitive triplets.

There are two main types of Regulating mutual motifs. One includes two TFs regulating each other, including *EGR1-KLF4, EGR1-TBX3*, and *KLF4-TBX3*, and jointly regulating the same target gene. A total of 47 target genes were regulated, including 6 TFs, forming 24 transitive triplets, and 27 hub genes, forming 98 transitive triplets. Among these target genes, 4 target genes were both TF and Hub genes, forming 20 transitive triplets. The other type of Regulating mutual motifs includes two TFs that are co-expressed and share a target gene. We found that *FOS* and *JUNB* co-expressed and co-regulated the *DUSP1* gene.

In the Regulated mutual motif, one TF regulated two genes, and there was a co-expression relationship between the two target genes. It is composed of TFs, including *EGR1, KLF4, EPAS1, BCL6B,* and *TBX3*, and their regulated target genes, forming a total of 1620 transitive triplets. Of these triplets, there are 8 in which both target genes are TFs and 593 in which both are hub genes. In the Clique motif, only *EGR1-KLF4-TBX3* motif is the mutual regulation of these three TFs and the remaining motifs were co-expressed relationships among genes.

Identification of important genes and regulatory relationships related to hypoxia in the lung gene regulatory network

Formulas [8] and [10] were used to evaluate the importance of each gene and the

3-node motif, including FFL, Regulating mutual, and Regulated mutual type motif, in the lung tissue gene regulatory network. We found that the top several important genes were *KLF4, BCL6B, EGR1, SMAD6* and *EPAS1* transcription factor genes, which are also hub genes. The top 25% of the node importance scores in the Tibetan pig lung gene regulatory network are shown in Table 6.

And the Regulating mutual motif formed by *KLF4-EGR1-BCL6B* was the most important motif based on motif score. We call it the "*KLF4-EGR1-BCL6B*" triplet. In this triplet, the *KLF4* and *EGR1* genes regulate the same target gene, *BCL6B*. This triplet preferred to synergistically regulate the *EPAS1, KDR, SMAD6, SMAD7, CCN1,* and *ATOH8* genes (Fig. 6), which comprised 18 motifs (Table 7). The "*KLF4-EGR1-BCL6B*" triplet coordinately regulated the *SMAD6* and *SMAD7* genes, which play an important role in the TGF-β signaling pathway. *EPAS1* is an important hypoxia-inducible factor. This triplet may also indirectly regulate *SMAD6* and *SMAD7* genes by regulating the *EPAS*1 gene. This triplet regulated the *KDR* gene, which is involved in the PI3K-Akt signaling pathway. TGF-β and PI3K-Akt signaling pathways both play an important role in hypoxia response and hypoxia adaptation [7,42-44].

Validation of important genes in lung tissue

To confirm the relationship between *KLF4, EGR1, BCL6B, SMAD6*, *EPAS1, KDR, SMAD7, CCN1, ATOH8,* and *MMP23B* genes, we used lung tissue transcriptome data from the Tibetan sheep, yak and Diqing Tibetan pig population for validation via

co-expression analysis. In Diqing Tibetan pig lung tissue, six genes were highly expressed, including *KLF4, EGR1, EPAS1, SMAD6, SMAD7* and *KDR*. The *KLF4, EGR1* and *SMAD7* genes clustered into one module, while *EPAS1* and *SMAD6* clustered into the other module. Overall, 81.82% of the co-expression relationships among the above genes were confirmed.

The *KLF4, BCL6B, EPAS1, EGR1, SMAD6, KDR, CCN1* and *ATOH8* genes had the highest expression in lung tissues compared with the other five tissues of Tibetan sheep (muscle, liver, heart, spleen and kidney). Except for *EGR1* and *ATOH8*, the other genes were all clustered into the same module. In total, 73.68% of the relationships between genes were validated.

With the exception of *ATOH8* and *MMP23B*, the other genes were most highly expressed in the lung tissues of yak compared to the other five tissues. *KLF4, BCL6B, EPAS1, SMAD6*, and *SMAD7* clustered into key modules related to the lungs. We successfully verified 60% of the relationships between genes.

Discussion

Many previous studies primarily focused on identifying differentially expressed genes through gene expression profile analysis, but interactions between genes in different cell states may not be fully considered [45]. Compared to expression level analysis, network-based analysis not only captures local patterns but also identifies global patterns in a biological context, revealing molecular regulation details of hub genes at the network level. Therefore, through gene network analysis, not only are hub genes

related to biological processes identified but also their important regulatory relationships.

In this study, we detected the gene regulatory network related to Tibetan pig lung tissue. Combining topological characteristics, differential expression, and tissue-specific expression, we identified a list of genes related to hypoxia adaptation in Tibetan pig lung tissue, such as *EPAS1, LOXL1, KLF4, EGR1, BCL6B, KDR, CCN1, ATOH8,* and *MRC2*.

Some studies have shown that the *EPAS1* and *LOXL1* genes might be associated with adaptation to hypoxic conditions. The *EPAS1* gene encodes one subunit of hypoxia-inducible factor (HIF), which shows multifarious effects involved in complex oxygen sensing [46] and regulation of angiogenesis, hemoglobin concentration and erythrocytosis [47]. Based on selection signature analysis, many studies have identified the *EPAS1* gene as a key evolutionary molecular adaptation to high-altitude hypoxic environments in humans [48-50], the Tibetan horse [51], and the Tibetan pig [2]. Lysyl oxidase-like-1 (*LOXL1*) is essential for the stability and strength of elastic vessels and tissues [52] and may play important roles in the enhanced angiogenesis promoted by hypoxia [53].

In this study, some key genes were involved in lung tissue development, such as *MRC2, KLF4,* and *EGR1.* The *MRC2* gene is a member of the mannose receptor family, which plays an important role in the development and remodeling of the lung [54]. Angiogenesis also plays an important role in lung growth and development. A majority of identified hub genes participate in the angiogenesis process, such as the

508 *EGR1* gene, which plays an important role in the process of angiogenesis [55,56]. The
509 *KLF4* gene tended to be pleiotropic. It is abundantly expressed in pulmonary vascular
510 endothelial cells [57]. Not only does it promote pulmonary angiogenesis and blood
511 transport [58] and accelerate the acquisition and transport of oxygen, but it also
512 protect the lungs from oxygen deficiency, facilitating adaptation to a hypoxic
513 environment [57]. These genes might influence growth and development in the
514 Tibetan pig lung, which contributes to obtaining and transporting oxygen better in
515 hypoxic environments.

516 Based on the Tibetan pig lung tissue-specific gene network, we found that the
517 *KLF4* and *EGR1* simultaneously regulate the *BCL6B* gene, forming
518 *KLF4-EGR1-BCL6B* transitive triplets, which are dominated by the *KLF4* gene and
519 affect the expression of *EPAS1*, *SMAD6*, *SMAD7*, *CCN1*, *KDR*, and *ATOH8*. These
520 key genes and regulatory relationships were validated in the lung tissue of Tibetan
521 pigs from Jia et al. (2016) [7] and Tibetan sheep and yak from Tang et al. (2017) [59].
522 After a large literature review and verification of gene function annotation, we
523 postulate that *KLF4-EGR1-BCL6B* transitive triplets may contribute to the adaptation
524 of Tibetan pigs to hypoxic environments.

525 The *KLF4, EGR1,* and *BCL6B* genes jointly regulate the *SMAD6* and *SMAD7*
526 genes, which are important regulators of the TGF-β signaling pathway. In the TGF-β
527 signaling pathway, SMAD family genes are very important signal transduction and
528 regulatory molecules. *SMAD6* and *SMAD7* are antagonists of the TGF-β gene family.
529 High expression of *SMAD7* inhibits the transcription of *SMAD2* and *SMAD3* genes

induced by the *TGF-β* gene and antagonizes tissue fibrosis [60]. Therefore, the *KLF4-EGR1-BCL6B* transitive triplet in Tibetan pig lungs may mediate the TGF-β signaling pathway by regulating expression of *SMAD6* and *SMAD7*, thereby enhancing the anti-fibrotic effect of the lungs and improving adaptation to the hypoxic environment.

Moreover, the *KLF4-EGR1-BCL6B* transitive triplet regulates the *KDR* gene, which is primarily expressed in pulmonary vascular endothelial cells and has important proangiogenic activity [61]. The *KDR* gene is an important regulator of the PI3K-Akt signaling pathway. Jia et al. (2016) [9] and Qi et al. (2018) [44] found that the PI3K-Akt signaling pathway was involved in hypoxia adaptation in both Tibetan pigs and yaks. Under hypoxic conditions, the combination of *KDR* and *VEGF* activates the downstream *PI3K* gene, thereby regulating proliferation and differentiation of neovascular endothelial cells and playing an important role in the development of angiogenesis [62]. Therefore, the *KLF4-EGR1-BCL6B* transitive triplet may act on the PI3K-Akt pathway by mediating the *KDR* gene and accelerating the acquisition and transportation of oxygen under hypoxic conditions.

In addition, the *KLF4-EGR1-BCL6B* transitional triplet also regulated the *ATOH8, CCN1* and *EPAS1* genes. High expression of *CCN1* suppresses pulmonary vascular smooth muscle contraction in response to hypoxia [63]. The *ATOH8* gene participates in the *ALK-1/SMAD/ATOH8* axis, which attenuates the hypoxic response in endothelial cells in the pulmonary circulation and may help prevent the development of pulmonary arterial hypertension [64]. The *MMP23B* gene is a member of the MMP

gene family, and MMP matrix metalloproteinases play an important role in tissue remodeling and angiogenesis [65]. Moreover, *MMP23B* is regulated by *EPAS1* and ssc-miR-296-3p. Studies have shown that miR-296 can regulate angiogenesis [66,67].

Conclusions

In summary, through gene network analysis, we found that lung tissue may play an important role in hypoxia adaptation in Tibetan pigs. We comprehensively profiled the gene regulatory network of Tibetan pig lung tissue, identifying a series of genes related to hypoxia adaptation and discovering that *KLF4* is the core regulator of the *KLF4-EGR1-BCL6B* transitive triplet, which may mediate the TGF-β signaling pathway and improve the ability of Tibetan pigs to adapt to hypoxia. These findings contribute to a better understanding of the molecular mechanisms potentially underlying hypoxia adaptation.

List of abbreviations

WGCNA: weighted correlation network analysis;

TF: transcription factors;

miRNA: microRNA;

FFL: feed-forward loop;

MAD: median absolute deviation;

TOM: topological overlap measure;

GS: gene significance;

TSI(τ)：tissue-specificity index;

MM：module membership;

GO: gene ontology;

KEGG: Kyoto Encyclopedia of Genes and Genomes;

PWM: position weight matrix;

3'UTR: 3'-untranslated region;

SP: significance profile;

TSP: triad significance profile.

Figure legends

Fig. 1 Clustering dendrogram of 36 tissue samples of Tibetan pigs and Rongchang pigs

The figure shows the clustering of a total of 36 tissue samples of Tibetan pigs and Rongchang pigs, where "T" represents Tibetan pigs and "R" represents Rongchang pigs. For example, "T_muscle1" represents the muscle sample of the first individual Tibetan pig.

Fig. 2 Weighted gene co-expression network analysis of Tibetan pigs

(a) Network topology of different soft-thresholding power of Tibetan pig Co-expression Network. The left panel displays the influence of soft-thresholding power (x-axis) on scale-free fit index (y-axis). The right panel shows the influence of soft-thresholding power (x-axis) on the mean connectivity (degree, y-axis). **(b)** Gene

clustering module of Tibetan pig co-expression network. The dissimilarity was based on topological overlap. The “Merged dynamic” is the result of merging modules with a correlation higher than 0.9. The y-axis is the distance determined by the extent of topological overlap. **(c)** Heatmap of the correlation between module eigengenes and the six tissues of Tibetan Pig. The x-axis is the six tissues of Tibetan pigs, and the y-axis is the module eigengene (ME). In the heatmap, red represents high adjacency (positive correlation) and blue represents low adjacency (negative correlation). In brackets is the p-value of the correlation test.

Fig. 3 Multi-tissue expression patterns of genes in key modules of lung tissue of two pig breeds

(a) Multi-tissue expression patterns of key module genes in Tibetan pigs lung tissue. **(b)** Multi-tissue expression patterns of key module genes in Rongchang pigs lung tissue. The 1, 2, 3, 4, 5, and 6 in the figure represent muscle, liver, heart, spleen, kidney, and lung tissues, respectively. And Yellow or green colored lines correspond to genes with low membership value; red and purple colored lines correspond to genes with high membership value.

Fig. 4 Gene regulatory network of six tissues of Tibetan pigs

In each network in the figure, the yellow dots represent TFs, the green dots represent miRNAs, and the hub genes are represented by triangles. The red edges with arrows represent the regulatory relationship between TFs and miRNAs and target genes. The

gray edge indicates that there is only a co-expression relationship between the two genes.

Fig. 5 The triad significance profile (TSP) of Tibetan pig lung gene regulatory network

The ordinate in the figure is the normalized Z value, and the abscissa is 13 motifs types. And the point marked with "*" is that the frequency of the corresponding motif in lung tissues gene regulatory network is significantly different from that of random networks ($p<1E-04$). The motifs are FFL (7), Regulated mutual (9), Regulating mutual (10) and Clique (13) in order.

Fig. 6 *"KLF4-EGR1-BCL6B"* transitive triplet and their regulated genes in Tibetan pig lung tissue

The transitive triplet formed by *KLF4-EGR1-BCL6B* regulates *EPAS1, SMAD6, SMAD7, KDR, ATOH8, CCN1* genes, and mediates the TGF-β and PI3K-Akt signaling pathways by regulating *SMAD6, SMAD7* and *KDR* genes, respectively. The green edge in the figure represents regulation, and the red edge represents inhibition.

Table 1 The fundamental network topology concepts of key modules in Tibetan pig and Rongchang pig tissues

Table 2 Pathways that are only significantly enriched in Tibetan pig tissue

modules

Table 3 Hub gene information of key modules in Tibetan pigs and Rongchang pigs

Table 4 Detailed information of gene regulatory networks in six tissues of Tibetan pigs

Table 5 Motif information in regulatory networks of six tissues in Tibetan pigs

Table 6 The top 25% of S_{node} genes in the Tibetan pig lung gene regulatory network

Table 7 The motifs formed between the *"KLF4-EGR1-BCL6B"* triplet and its regulatory genes in the lung

Additional file

Figure S1 Weighted gene co-expression network analysis of Rongchang pigs

(a) Analysis of network topology of Rongchang pig showed that it meet the scale-free topology threshold of 0.8 when $\beta = 20$. The left panel shows the scale-free fit index as a function of the soft-threshold power. The right panel displays the mean connectivity as a function of the soft-threshold power. **(b)** The dissimilarity was based on

topological overlap. The “Merged dynamic” is the result of merging modules with a correlation higher than 0.9. The y-axis is the distance determined by the extent of topological overlap. **(c)** Heatmap displaying the correlations and significant differences between gene modules and six tissues of Rongchang pigs. Red represents high adjacency (positive correlation) and blue represents low adjacency (negative correlation). In brackets is the p-value of the correlation test.

Figure S2 The co-expression network of six tissues key modules of Tibetan pig

The co-expression network of muscle, liver, heart, spleen, kidney and lung in the figure shows the co-expression relationship of weight above 0.35, 0.35, 0.25, 0.35, 0.35 and 0.25, respectively. The dark dots in the figure represent the hub genes of each network.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was financially supported by Natural Science Foundation of China (No. 32070571), the Academic Backbone Project of Northeast Agricultural University (No.15XG14), NEAU Research Founding for Excellent Young Teachers (2010RCB29).

Authors' contributions

ZW, XW and TW conceived the project. TW, YG, SL and CZ performed the bioinformatics and data analysis. TW and ZW wrote the manuscript. TC, KD and PW collected the samples and data. All authors read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Acknowledgments

Not applicable.

References

[1]Ai H, Yang B, Li J, Xie X, Chen H, Ren J. Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. BMC Genomics. 2014; 15(1):834. doi: 10.1186/1471-2164-15-834.

[2]Ma Y, Han X, Huang C, Zhong L, Adeola AC, Irwin DM, et al. Population Genomics Analysis Revealed Origin and High-altitude Adaptation of Tibetan Pigs. Sci Rep. 2019; 9(1):11463. doi: 10.1038/s41598-019-47711-6.

[3]Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. Nat Genet. 2013; 45(12):1431-8. doi: 10.1038/ng.2811.

[4]Shang P, Li W, Tan Z, Zhang J, Dong S, Wang K, et al. Population Genetic Analysis of Ten Geographically Isolated Tibetan Pig Populations. Animals (Basel). 2020; 10(8):1297. doi: 10.3390/ani10081297.

[5]Li M, Jin L, Ma J, Tian S, Li R, Li X. Detecting mitochondrial signatures of selection in wild Tibetan pigs and domesticated pigs. Mitochondrial DNA A DNA Mapp Seq Anal. 2016; 27(1):747-52. doi: 10.3109/19401736.2014.913169.

[6]Huang M, Yang B, Chen H, Zhang H, Wu Z, Ai H, et al. The fine-scale genetic structure and selection signals of Chinese indigenous pigs. Evol Appl. 2019; 13(2):458-475. doi: 10.1111/eva.12887.

[7]Jia C, Kong X, Koltes JE, Gou X, Yang S, Yan D, et al. Gene Co-Expression

727 Network Analysis Unraveling Transcriptional Regulation of High-Altitude
728 Adaptation of Tibetan Pig. PLoS One. 2016; 11(12):e0168161. doi:
729 10.1371/journal.pone.0168161.
730 [8]Zhang B, Chamba Y, Shang P, Wang Z, Ma J, Wang L, et al. Comparative
731 transcriptomic and proteomic analyses provide insights into the key genes
732 involved in high-altitude adaptation in the Tibetan pig. Sci Rep. 2017; 7(1):3654.
733 doi: 10.1038/s41598-017-03976-3.
734 [9]Young AI. Solving the missing heritability problem. PLoS Genet. 2019;
735 15(6):e1008222. doi: 10.1371/journal.pgen.1008222.
736 [10]Narang V, Ramli MA, Singhal A, Kumar P, de Libero G, Poidinger M, et al.
737 Automated Identification of Core Regulatory Genes in Human Gene Regulatory
738 Networks. PLoS Comput Biol. 2015; 11(9):e1004504. doi:
739 10.1371/journal.pcbi.1004504.
740 [11]Nishio Y, Usuda Y, Matsui K, Kurata H. Computer-aided rational design of the
741 phosphotransferase system for enhanced glucose uptake in Escherichia coli. Mol
742 Syst Biol. 2008; 4:160. doi: 10.1038/msb4100201.
743 [12]McLeay RC, Lesluyes T, Cuellar Partida G, Bailey TL. Genome-wide in silico
744 prediction of gene expression. Bioinformatics. 2012; 28:2789–2796. doi:
745 10.1093/bioinformatics/bts529.
746 [13]Cheng C, Gerstein M. Modeling the relative relationship of transcription factor
747 binding and histone modifications to gene expression levels in mouse embryonic
748 stem cells. Nucleic Acids Res. 2012; 40:553–568. doi: 10.1093/nar/gkr752.

[14]Döhr S, Klingenhoff A, Maier H, Hrabé de Angelis M, Werner T, Schneider R. Linking disease-associated genes to regulatory networks via promoter organization. Nucleic Acids Res. 2005; 33(3):864-72. doi: 10.1093/nar/gki230.

[15]Buckingham M, Rigby PW. Gene regulatory networks and transcriptional mechanisms that control myogenesis. Dev Cell. 2014; 28:225–238. doi: 10.1016/j.devcel.2013.12.020.

[16]Chen J, Alvarez MJ, Talos F, Dhruv H, Rieckhof GE, Iyer A,et al. Identification of Causal Genetic Drivers of Human Disease through Systems-Level Analysis of Regulatory Networks. Cell. 2016; 166(4):1055. doi: 10.1016/j.cell.2016.07.036.

[17]Adamcsek B, Palla G, Farkas IJ, Derényi I, Vicsek T. CFinder: locating cliques and overlapping modules in biological networks. Bioinformatics. 2006; 22(8):1021-3. doi: 10.1093/bioinformatics/btl039.

[18]Shalgi R, Brosh R, Oren M, Pilpel Y, Rotter V. Coupling transcriptional and post-transcriptional miRNA regulation in the control of cell fate. Aging. 2009; 1:762–770. doi: 10.18632/aging.100085.

[19]Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559. doi: 10.1186/1471-2105-9-559.

[20]Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005; 4:Article17. doi: 10.2202/1544-6115.1128.

[21]Kumar L, E Futschik M. Mfuzz: a software package for soft clustering of

771 microarray data. Bioinformation. 2007; 2(1):5-7. doi: 10.6026/97320630002005.
772 [22]Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al.
773 Genome-wide midrange transcription profiles reveal expression level relationships
774 in human tissue specification. Bioinformatics. 2005; 21(5):650-9. doi:
775 10.1093/bioinformatics/bti042.
776 [23]Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of
777 large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;
778 4(1):44-57. doi: 10.1038/nprot.2008.211.
779 [24]Hu H, Miao YR, Jia LH, Yu QY, Zhang Q, Guo AY. AnimalTFDB 3.0: a
780 comprehensive resource for annotation and prediction of animal transcription
781 factors. Nucleic Acids Res. 2019; 47(D1):D33-D38. doi: 10.1093/nar/gky822.
782 [25]Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the
783 integration of genomic datasets with the R/Bioconductor package biomaRt. Nat
784 Protoc. 2009; 4(8):1184-91. doi: 10.1038/nprot.2009.97.
785 [26]Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al.
786 Determination and inference of eukaryotic transcription factor sequence
787 specificity. Cell. 2014; 158(6):1431-1443. doi: 10.1016/j.cell.2014.08.009.
788 [27]Tan G, Lenhard B. TFBSTools: an R/bioconductor package for transcription
789 factor binding site analysis. Bioinformatics. 2016; 32(10):1555-6. doi:
790 10.1093/bioinformatics/btw024.
791 [28]Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA
792 sequences to function. Nucleic Acids Res. 2019; 47(D1):D155-D162. doi:

10.1093/nar/gky1141.

[29]Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. MicroRNA targets in Drosophila. Genome Biol. 2003; 5(1):R1. doi: 10.1186/gb-2003-5-1-r1.

[30]Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science. 2002; 298(5594):824-7. doi: 10.1126/science.298.5594.824.

[31]Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. Hierarchical organization of modularity in metabolic networks. Science. 2002; 297(5586):1551-5. doi: 10.1126/science.1073374.

[32]Bascompte J. Disentangling the web of life. Science. 2009; 325(5939):416-9. doi: 10.1126/science.1170749.

[33]Alon U. Network motifs: theory and experimental approaches. Nat Rev Genet. 2007; 8(6):450-61. doi: 10.1038/nrg2102.

[34]Kashtan N, Itzkovitz S, Milo R, Alon U. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. Bioinformatics. 2004; 20(11):1746-58. doi: 10.1093/bioinformatics/bth163.

[35]Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, et al. Superfamilies of evolved and designed networks. Science. 2004; 303(5663):1538-42. doi: 10.1126/science.1089167.

[36]Tarazona S, Furió-Tarí P, Turrà D, Pietro AD, Nueda MJ, Ferrer A, et al. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res. 2015; 43(21):e140. doi: 10.1093/nar/gkv711.

[37]Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet. 2002; 31(1):64-8. doi: 10.1038/ng881.

[38]Rosenfeld N, Elowitz MB, Alon U. Negative autoregulation speeds the response times of transcription networks. J Mol Biol. 2002; 323(5):785-93. doi: 10.1016/s0022-2836(02)00994-4.

[39]Mangan S, Alon U. Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci USA. 2003; 100(21):11980-5. doi: 10.1073/pnas.2133841100.

[40]Mangan S, Zaslaver A, Alon U. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. J Mol Biol. 2003; 334(2):197-204. doi: 10.1016/j.jmb.2003.09.049.

[41]Harary F , Kommel H J . Matrix measures for transitivity and balance. Journal of Mathematical Sociology. 1979, 6(2):199-210.

[42]Ambalavanan N, Nicola T, Hagood J, Bulger A, Serra R, Murphy-Ullrich J, et al. Transforming growth factor-beta signaling mediates hypoxia-induced pulmonary arterial remodeling and inhibition of alveolar development in newborn mouse lung. Am J Physiol Lung Cell Mol Physiol. 2008; 295(1):L86-95. doi: 10.1152/ajplung.00534.2007.

[43]Chen Y, Feng J, Li P, Xing D, Zhang Y, Serra R, et al. Dominant negative mutation of the TGF-beta receptor blocks hypoxia-induced pulmonary vascular remodeling. J Appl Physiol (1985). 2006; 100(2):564-71. doi: 10.1152/japplphysiol.00595.2005.

837 [44]Qi X, Zhang Q, He Y, Yang L, Zhang X, Shi P, et al. The Transcriptomic
838 Landscape of Yaks Reveals Molecular Pathways for High Altitude Adaptation.
839 Genome Biol Evol. 2019; 11(1):72-85. doi: 10.1093/gbe/evy264.
840 [45]Kostka D., Spang R. Finding disease specific alterations in the co-expression of
841 genes. Bioinformatics. 2004; 20(1):i194–i199. doi:
842 10.1093/bioinformatics/bth909.
843 [46]Henderson J, Withford-Cave JM, Duffy DL, Cole SJ, Sawyer NA, Gulbin JP, et al.
844 The EPAS1 gene influences the aerobic-anaerobic contribution in elite endurance
845 athletes. Hum Genet. 2005; 118(3-4):416-23. doi: 10.1007/s00439-005-0066-0.
846 [47]Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, et al. Natural
847 selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in
848 Tibetan highlanders. Proc Natl Acad Sci USA. 2010; 107(25):11459-64. doi:
849 10.1073/pnas.1002443107.
850 [48]Peng Y, Yang Z, Zhang H, Cui C, Qi X, Luo X, et al. Genetic variations in Tibetan
851 populations and high-altitude adaptation at the Himalayas. Mol Biol Evol. 2011;
852 28(2):1075-81. doi: 10.1093/molbev/msq290.
853 [49]Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, et al. Genetic
854 evidence for high-altitude adaptation in Tibet. Science. 2010; 329(5987):72-5. doi:
855 10.1126/science.1189406.
856 [50]Xu S, Li S, Yang Y, Tan J, Lou H, Jin W, et al. A genome-wide search for signals
857 of high-altitude adaptation in Tibetans. Mol Biol Evol. 2011; 28(2):1003-11. doi:
858 10.1093/molbev/msq277.

[51]Liu X, Zhang Y, Li Y, Pan J, Wang D, Chen W, et al. EPAS1 gain-of-function mutation contributes to high-altitude adaptation in Tibetan horses. Mol Biol Evol. 2019; 36(11):2591–603. doi: 10.1093/molbev/msz158.

[52]Li G, Schmitt H, Johnson WM, Lee C, Navarro I, Cui J, et al. Integral role for lysyl oxidase-like-1 in conventional outflow tissue function and behavior. FASEB J. 2020. doi: 10.1096/fj.202000702RR.

[53]Xie Q, Xie J, Tian T, Ma Q, Zhang Q, Zhu B, et al. Hypoxia triggers angiogenesis by increasing expression of LOX genes in 3-D culture of ASCs and ECs. Exp Cell Res. 2017; 352(1):157-163. doi: 10.1016/j.yexcr.2017.02.011.

[54]Smith L, Wagner TE, Huizar I, Schnapp LM. uPARAP expression during murine lung development. Gene Expr Patterns. 2008; 8(7-8):486-93. doi: 10.1016/j.gep.2008.06.006.

[55]Adamson ED, Mercola D. Egr1 transcription factor: multiple roles in prostate tumor cell growth and survival. Tumour Biol. 2002; 23(2):93-102. doi: 10.1159/000059711.

[56]Sheng J, Liu D, Kang X, Chen Y, Jiang K, Zheng W. Egr-1 increases angiogenesis in cartilage via binding Netrin-1 receptor DCC promoter. J Orthop Surg Res. 2018; 13(1):125. doi: 10.1186/s13018-018-0826-x.

[57]Shatat MA, Tian H, Zhang R, Tandon G, Hale A, Fritz JS, et al. Endothelial Krüppel-like factor 4 modulates pulmonary arterial hypertension. Am J Respir Cell Mol Biol. 2014; 50(3):647-53. doi: 10.1165/rcmb.2013-0135OC.

[58]Ghaleb AM, Yang V. Krüppel-like factor 4 (KLF4): What we currently know.

Gene. 2017; 611:27-37. doi: 10.1016/j.gene.2017.02.025.

[59]Tang Q, Gu Y, Zhou X, Jin L, Guan J, Liu R, et al. Comparative transcriptomics of 5 high-altitude vertebrates and their low-altitude relatives. Gigascience. 2017; 6(12):1-9. doi: 10.1093/gigascience/gix105.

[60]Yan X, Liu Z, Chen Y. Regulation of TGF-beta signaling by Smad7. Acta Biochim Biophys Sin (Shanghai). 2009; 41(4):263-72. doi: 10.1093/abbs/gmp018.

[61]Melincovici CS, Boşca AB, Şuşman S, Mărginean M, Mihu C, Istrate M, et al. Vascular endothelial growth factor (VEGF) - key factor in normal and pathological angiogenesis. Rom J Morphol Embryol. 2018; 59(2):455-467.

[62]Graupera M, Potente M. Regulation of angiogenesis by PI3K signaling networks. Exp Cell Res. 2013; 319(9):1348-55. doi: 10.1016/j.yexcr.2013.02.021.

[63]Lee SJ, Zhang M, Hu K, Lin L, Zhang D, Jin Y. CCN1 suppresses pulmonary vascular smooth muscle contraction in response to hypoxia. Pulm Circ. 2015; 5(4):716-22. doi: 10.1086/683812.

[64]Morikawa M, Mitani Y, Holmborn K, Kato T, Koinuma D, Maruyama J, et al. The ALK-1/SMAD/ATOH8 axis attenuates hypoxic responses and protects against the development of pulmonary arterial hypertension. Sci Signal. 2019; 12(607):eaay4430. doi: 10.1126/scisignal.aay4430.

[65]Białkowska K, Marciniak W, Muszyńska M, Baszuk P, Gupta S, Jaworska-Bieniek K, et al. Polymorphisms in MMP-1, MMP-2, MMP-7, MMP-13 and MT2A do not contribute to breast, lung and colon cancer risk in

903 polish population. Hered Cancer Clin Pract. 2020; 18:16. doi:
904 10.1186/s13053-020-00147-w.
905 [66]Anand S, Cheresh DA. MicroRNA-mediated regulation of the angiogenic switch.
906 Curr Opin Hematol. 2011; 18(3):171-6. doi: 10.1097/MOH.0b013e328345a180.
907 [67]Li H, Ouyang XP, Jiang T, Zheng X, He P, Zhao G. MicroRNA-296: a promising
908 target in the pathogenesis of atherosclerosis? Mol Med. 2018; 24(1):12. doi:
909 10.1186/s10020-018-0012-y.
910

Figures

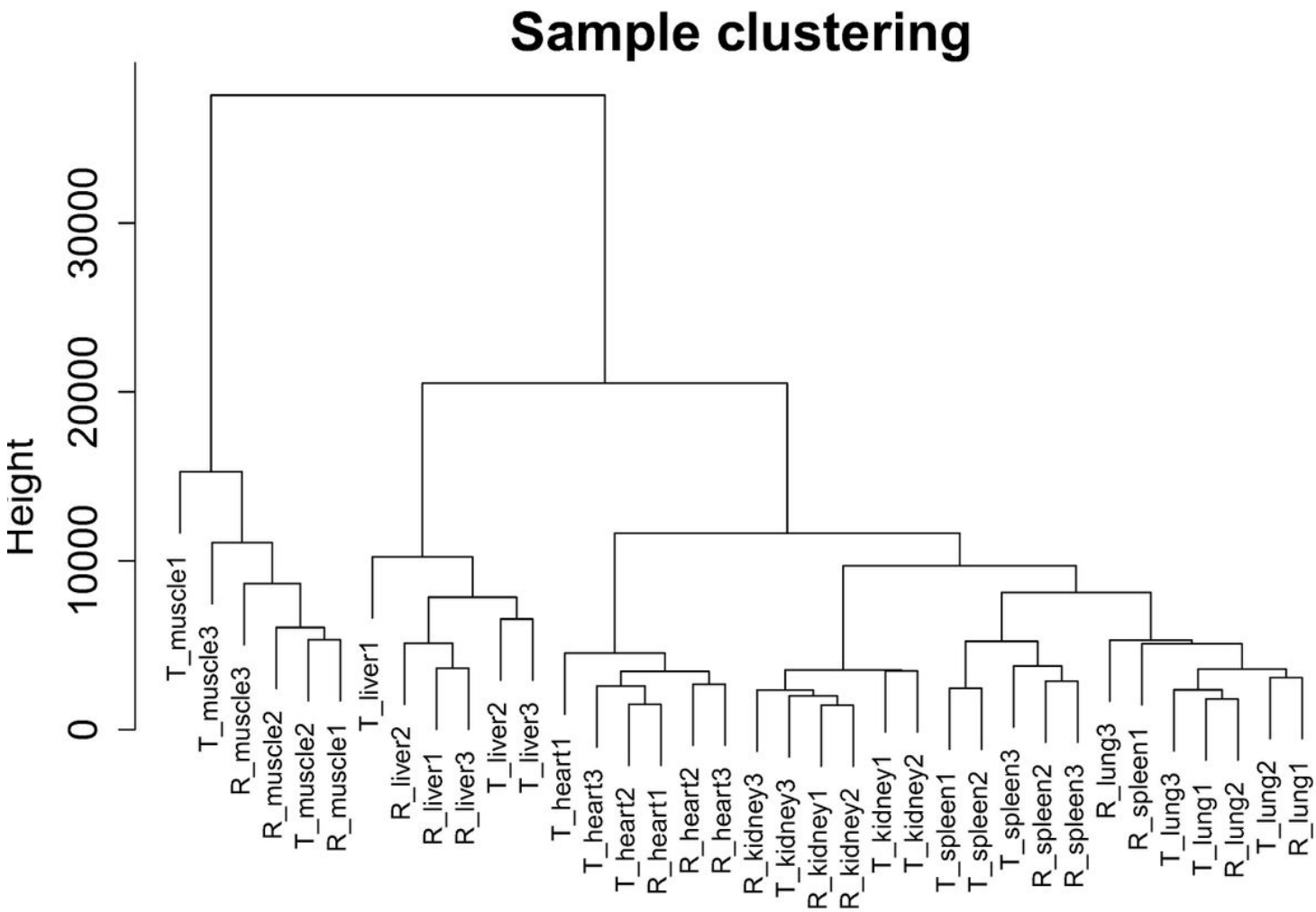

Figure 1

Clustering dendrogram of 36 tissue samples of Tibetan pigs and Rongchang pigs The figure shows the clustering of a total of 36 tissue samples of Tibetan pigs and Rongchang pigs, where “T” represents Tibetan pigs and “R” represents Rongchang pigs. For example, “T_muscle1” represents the muscle sample of the first individual Tibetan pig.

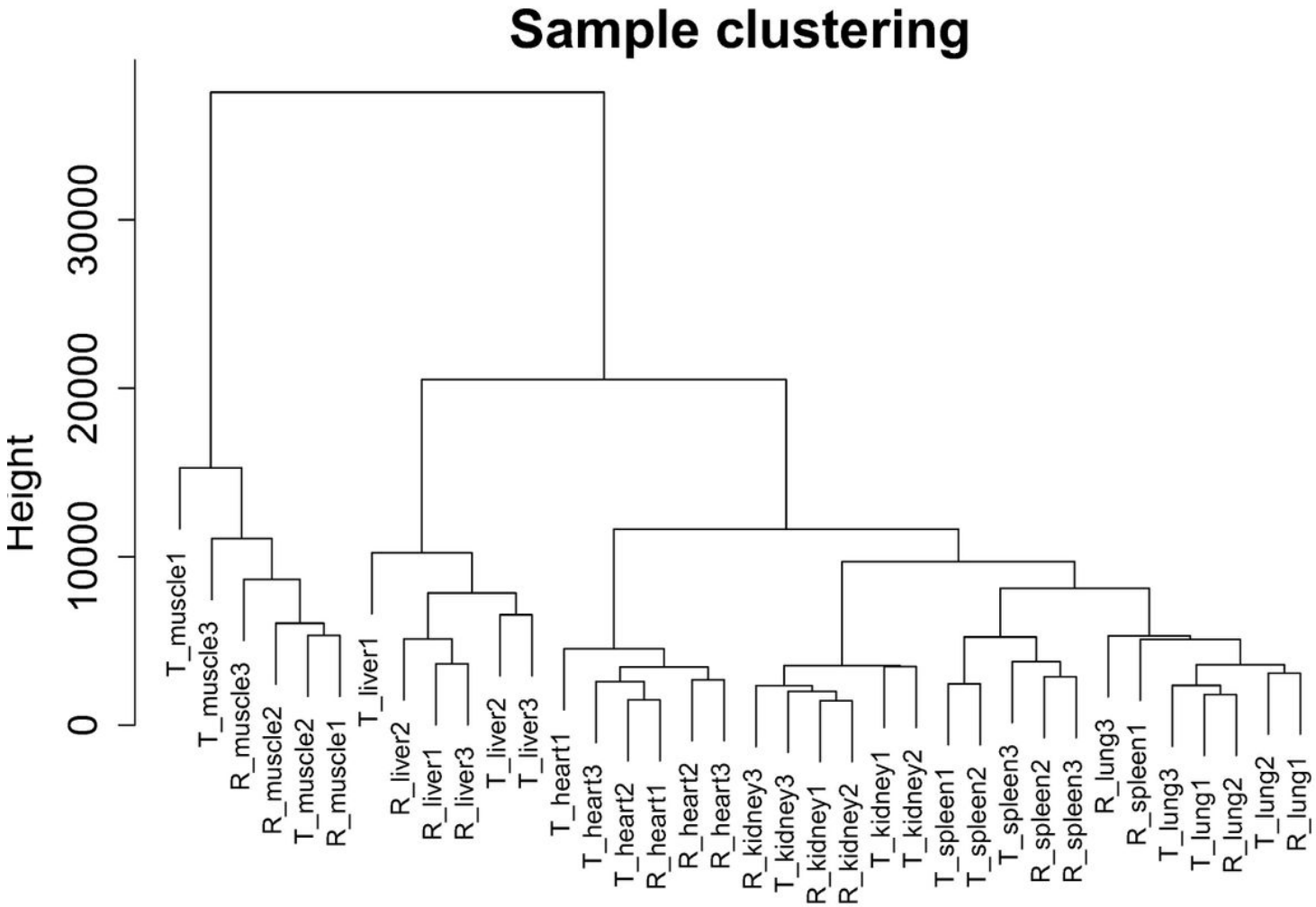

Figure 1

Clustering dendrogram of 36 tissue samples of Tibetan pigs and Rongchang pigs The figure shows the clustering of a total of 36 tissue samples of Tibetan pigs and Rongchang pigs, where “T” represents Tibetan pigs and “R” represents Rongchang pigs. For example, “T_muscle1” represents the muscle sample of the first individual Tibetan pig.

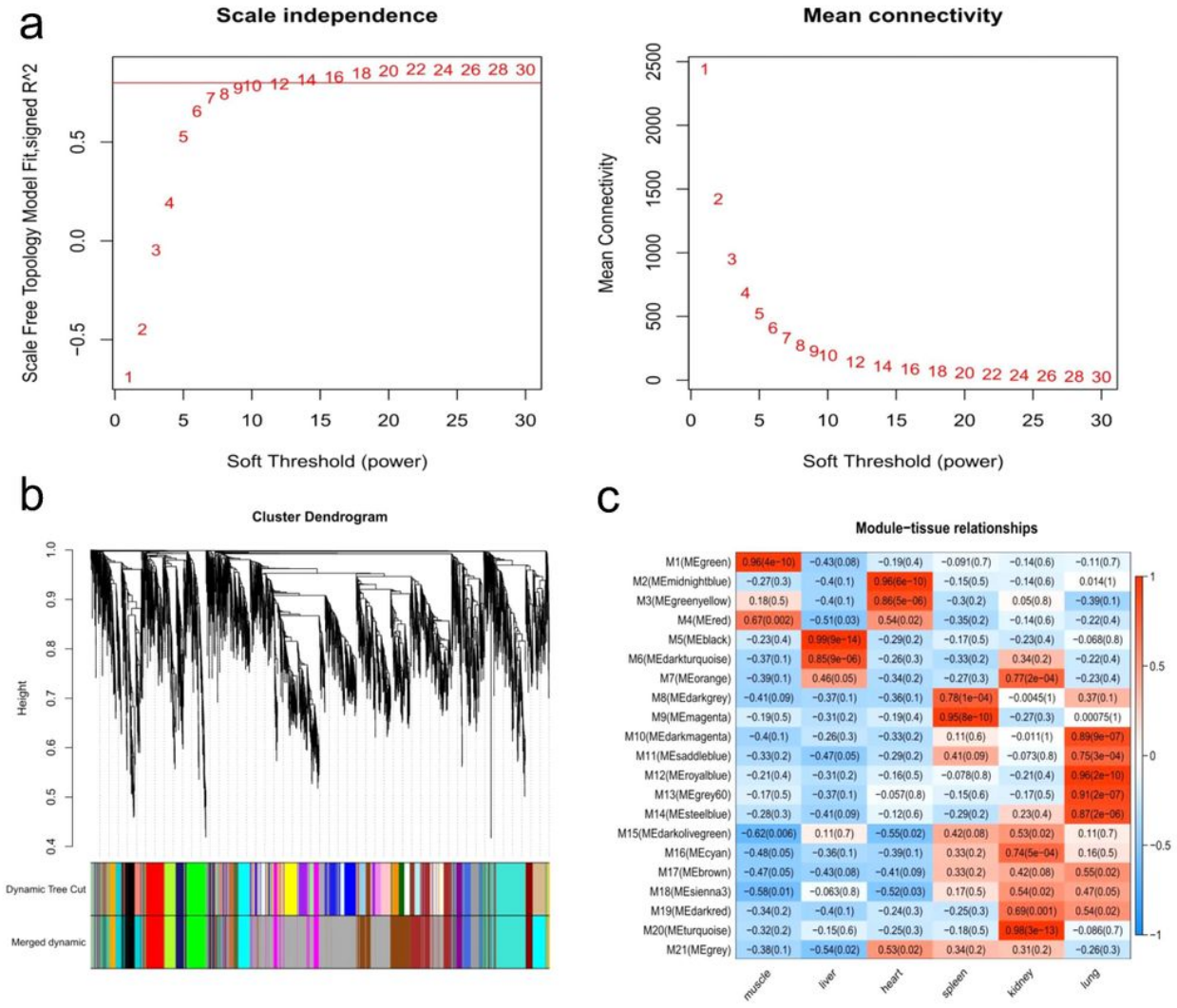

Figure 2

Weighted gene co-expression network analysis of Tibetan pigs (a) Network topology of different soft-thresholding power of Tibetan pig Co-expression Network. The left panel displays the influence of soft-thresholding power (x-axis) on scale-free fit index (y-axis). The right panel shows the influence of soft-thresholding power (x-axis) on the mean connectivity (degree, y-axis). (b) Gene clustering module of Tibetan pig co-expression network. The dissimilarity was based on topological overlap. The “Merged dynamic” is the result of merging modules with a correlation higher than 0.9. The y-axis is the distance determined by the extent of topological overlap. (c) Heatmap of the correlation between module eigengenes and the six tissues of Tibetan Pig. The x-axis is the six tissues of Tibetan pigs, and the y-axis is the module eigengene (ME). In the heatmap, red represents high adjacency (positive correlation) and blue represents low adjacency (negative correlation). In brackets is the p-value of the correlation test.

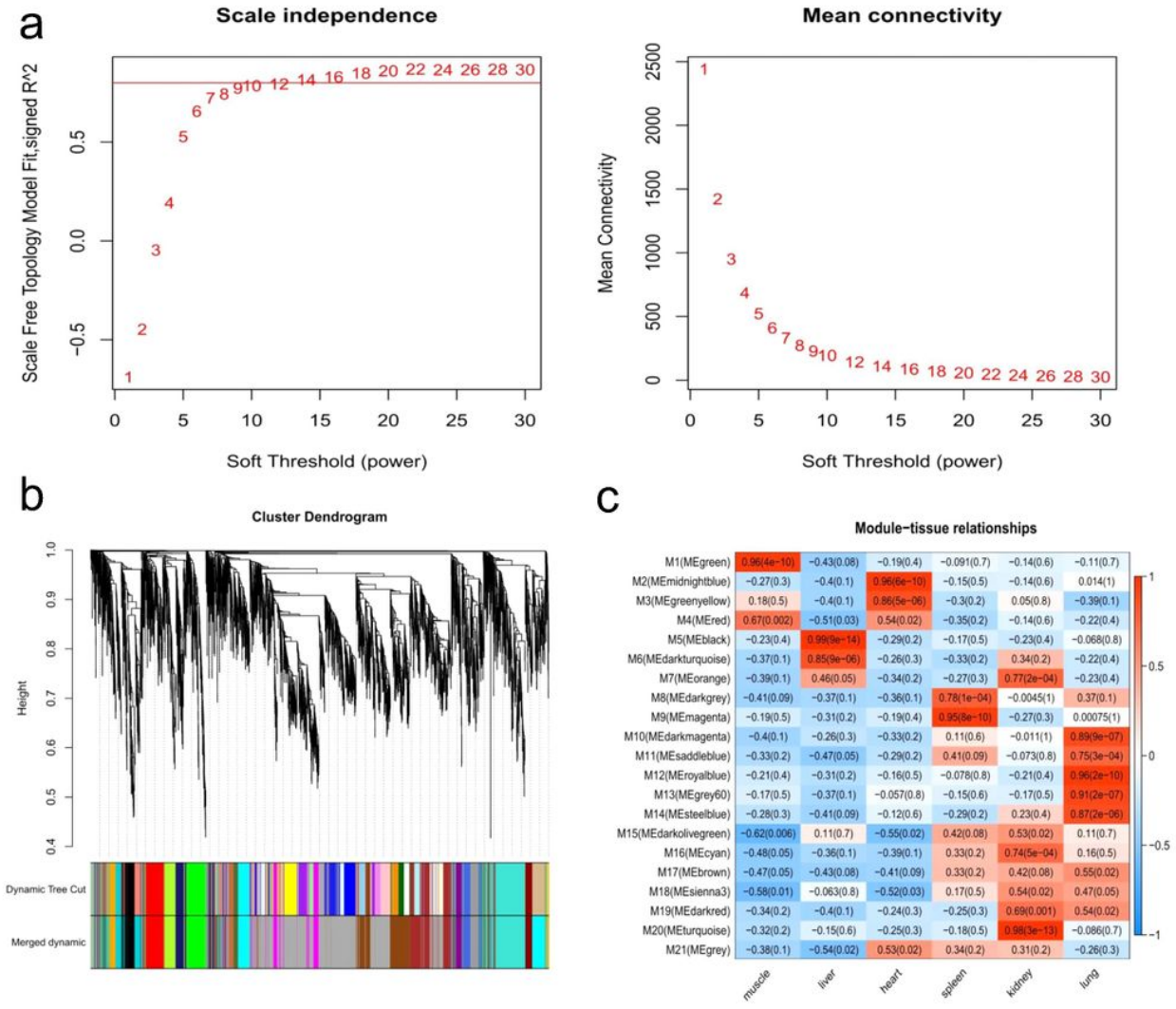

Figure 2

Weighted gene co-expression network analysis of Tibetan pigs (a) Network topology of different soft-thresholding power of Tibetan pig Co-expression Network. The left panel displays the influence of soft-thresholding power (x-axis) on scale-free fit index (y-axis). The right panel shows the influence of soft-thresholding power (x-axis) on the mean connectivity (degree, y-axis). (b) Gene clustering module of Tibetan pig co-expression network. The dissimilarity was based on topological overlap. The “Merged dynamic” is the result of merging modules with a correlation higher than 0.9. The y-axis is the distance determined by the extent of topological overlap. (c) Heatmap of the correlation between module eigengenes and the six tissues of Tibetan Pig. The x-axis is the six tissues of Tibetan pigs, and the y-axis is the module eigengene (ME). In the heatmap, red represents high adjacency (positive correlation) and blue represents low adjacency (negative correlation). In brackets is the p-value of the correlation test.

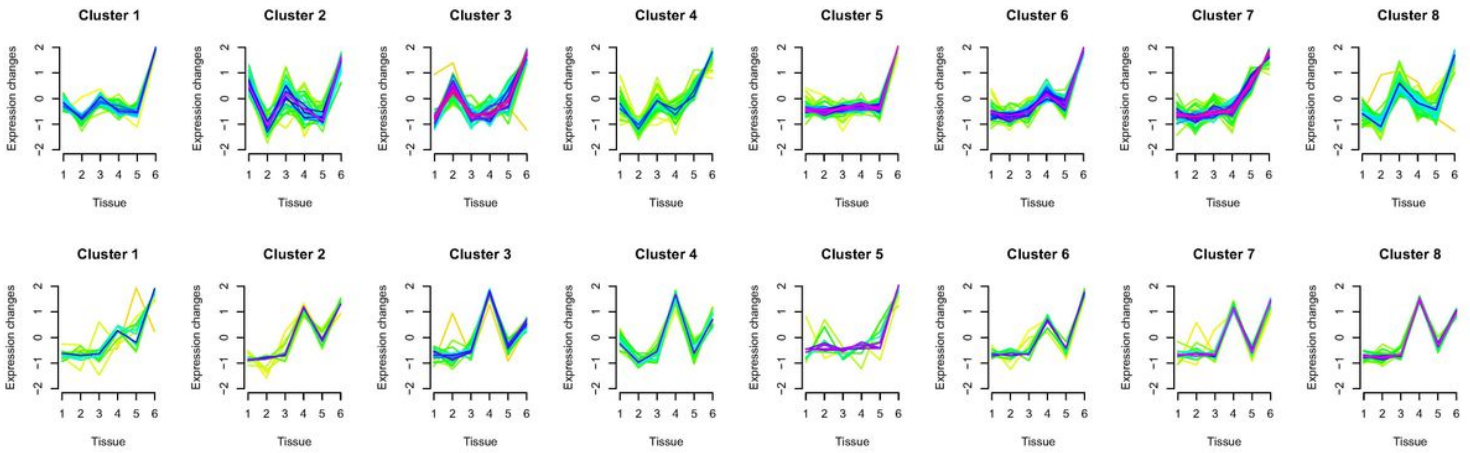

Figure 3

Multi-tissue expression patterns of genes in key modules of lung tissue of two pig breeds (a) Multi-tissue expression patterns of key module genes in Tibetan pigs lung tissue. (b) Multi-tissue expression patterns of key module genes in Rongchang pigs lung tissue. The 1, 2, 3, 4, 5, and 6 in the figure represent muscle, liver, heart, spleen, kidney, and lung tissues, respectively. And Yellow or green colored lines correspond to genes with low membership value; red and purple colored lines correspond to genes with high membership value.

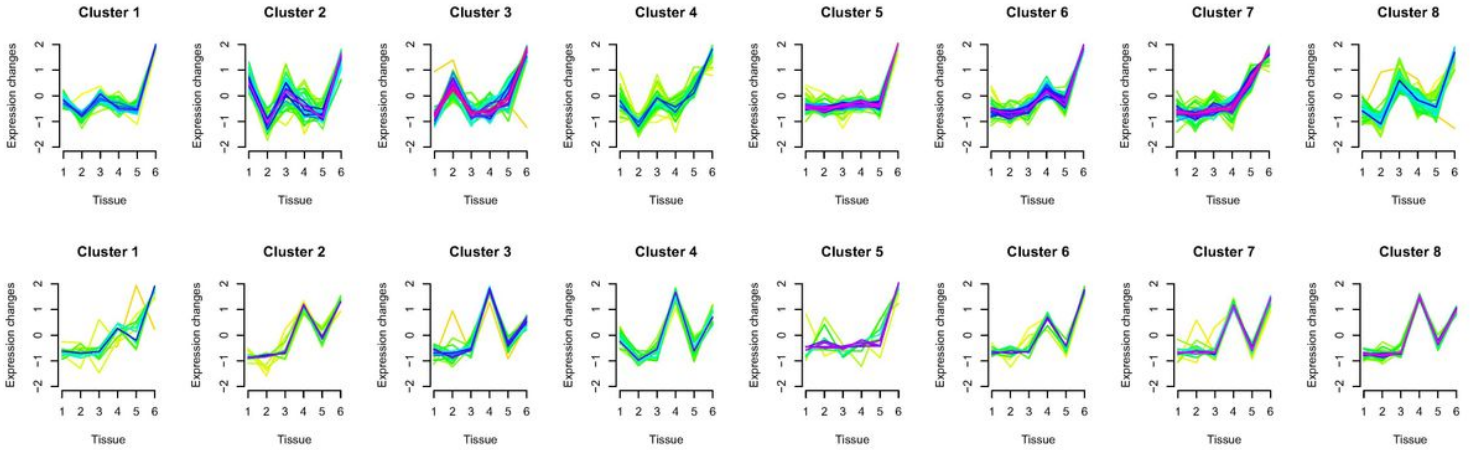

Figure 3

Multi-tissue expression patterns of genes in key modules of lung tissue of two pig breeds (a) Multi-tissue expression patterns of key module genes in Tibetan pigs lung tissue. (b) Multi-tissue expression patterns of key module genes in Rongchang pigs lung tissue. The 1, 2, 3, 4, 5, and 6 in the figure represent muscle, liver, heart, spleen, kidney, and lung tissues, respectively. And Yellow or green colored lines correspond to genes with low membership value; red and purple colored lines correspond to genes with high membership value.

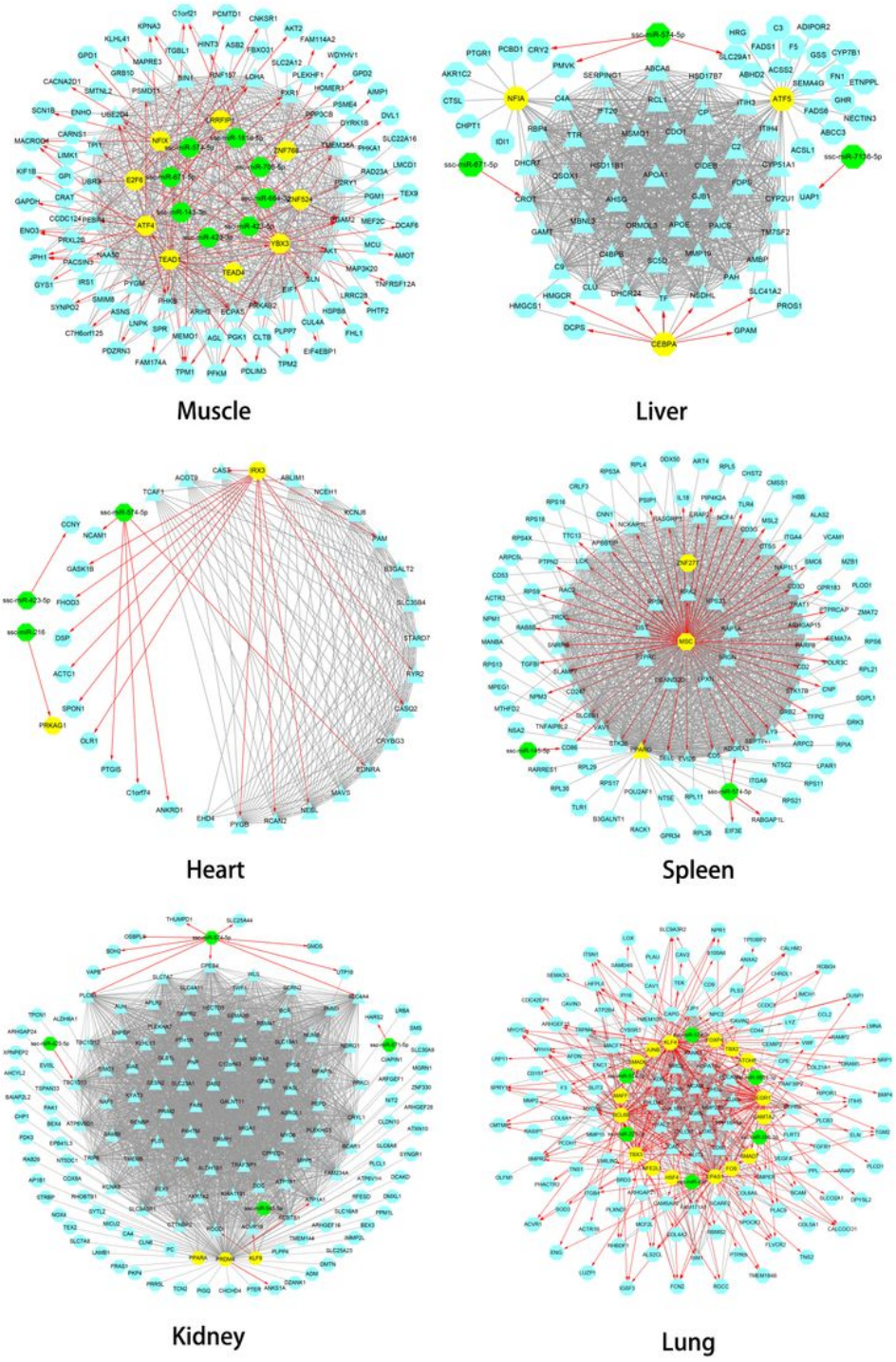

Figure 4

Gene regulatory network of six tissues of Tibetan pigs In each network in the figure, the yellow dots represent TFs, the green dots represent miRNAs, and the hub genes are represented by triangles. The red edges with arrows represent the regulatory relationship between TFs and miRNAs and target genes. The gray edge indicates that there is only a co-expression relationship between the two genes.

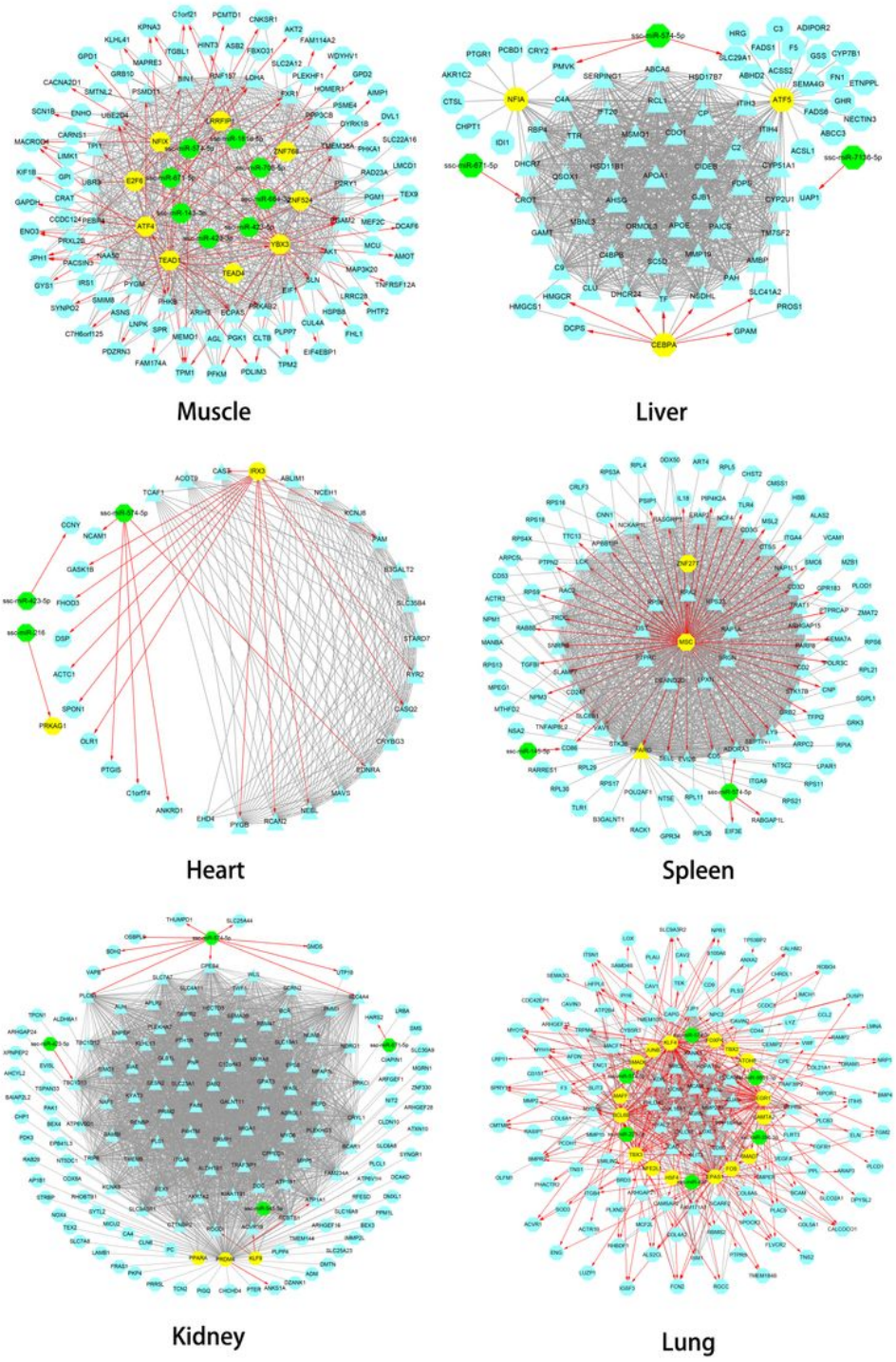

Figure 4

Gene regulatory network of six tissues of Tibetan pigs In each network in the figure, the yellow dots represent TFs, the green dots represent miRNAs, and the hub genes are represented by triangles. The red edges with arrows represent the regulatory relationship between TFs and miRNAs and target genes. The gray edge indicates that there is only a co-expression relationship between the two genes.

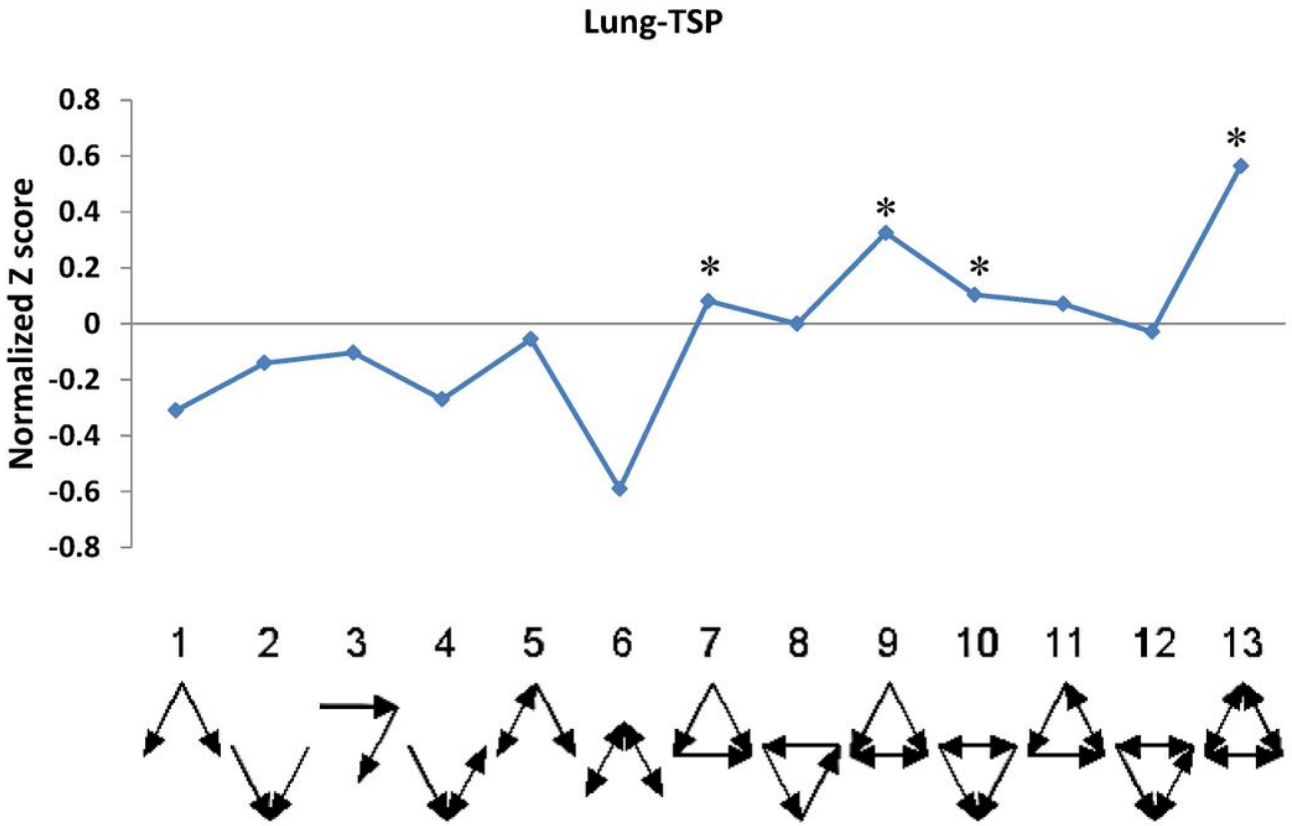

Figure 5

The triad significance profile (TSP) of Tibetan pig lung gene regulatory network The ordinate in the figure is the normalized Z value, and the abscissa is 13 motifs types. And the point marked with “*” is that the frequency of the corresponding motif in lung tissues gene regulatory network is significantly different from that of random networks (p<1E-04). The motifs are FFL (7), Regulated mutual (9), Regulating mutual (10) and Clique (13) in order.

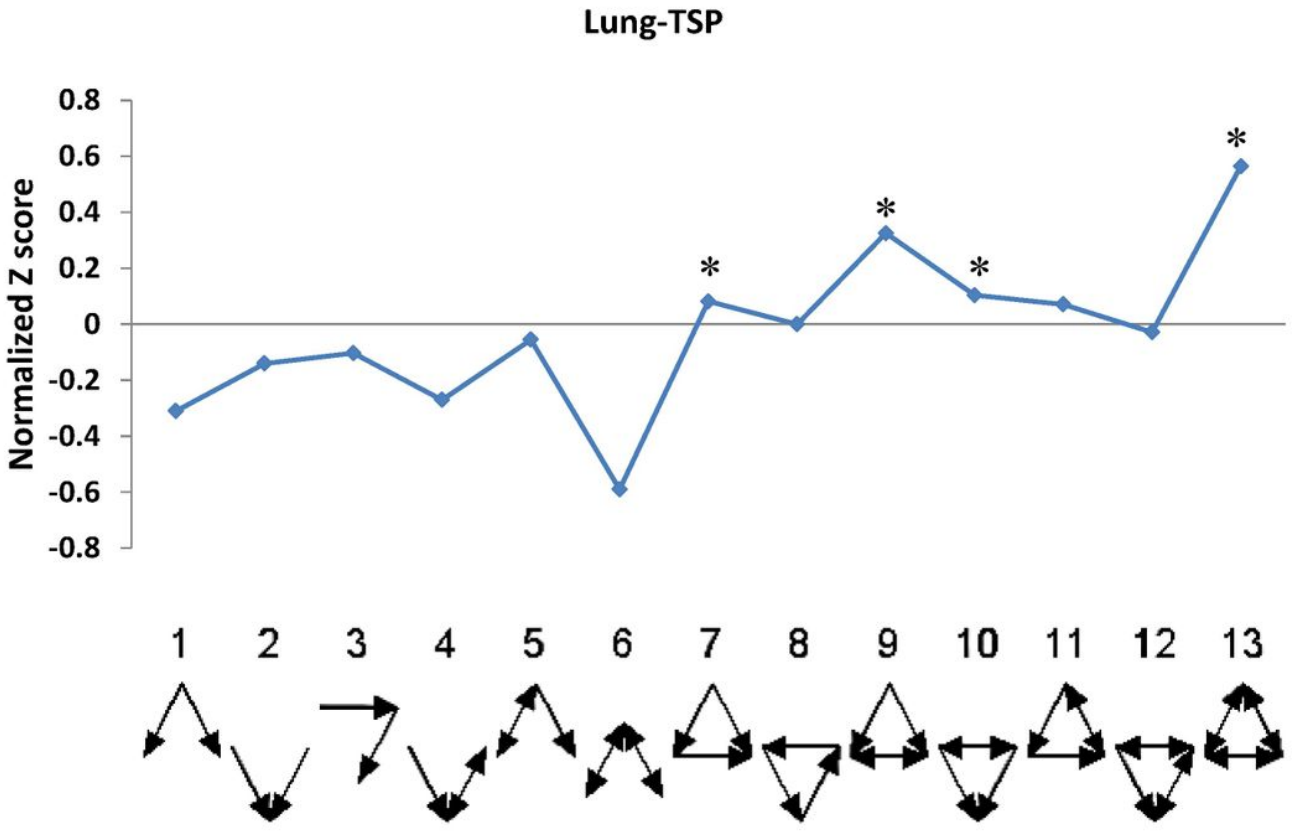

Figure 5

The triad significance profile (TSP) of Tibetan pig lung gene regulatory network The ordinate in the figure is the normalized Z value, and the abscissa is 13 motifs types. And the point marked with “*” is that the frequency of the corresponding motif in lung tissues gene regulatory network is significantly different from that of random networks (p<1E-04). The motifs are FFL (7), Regulated mutual (9), Regulating mutual (10) and Clique (13) in order.

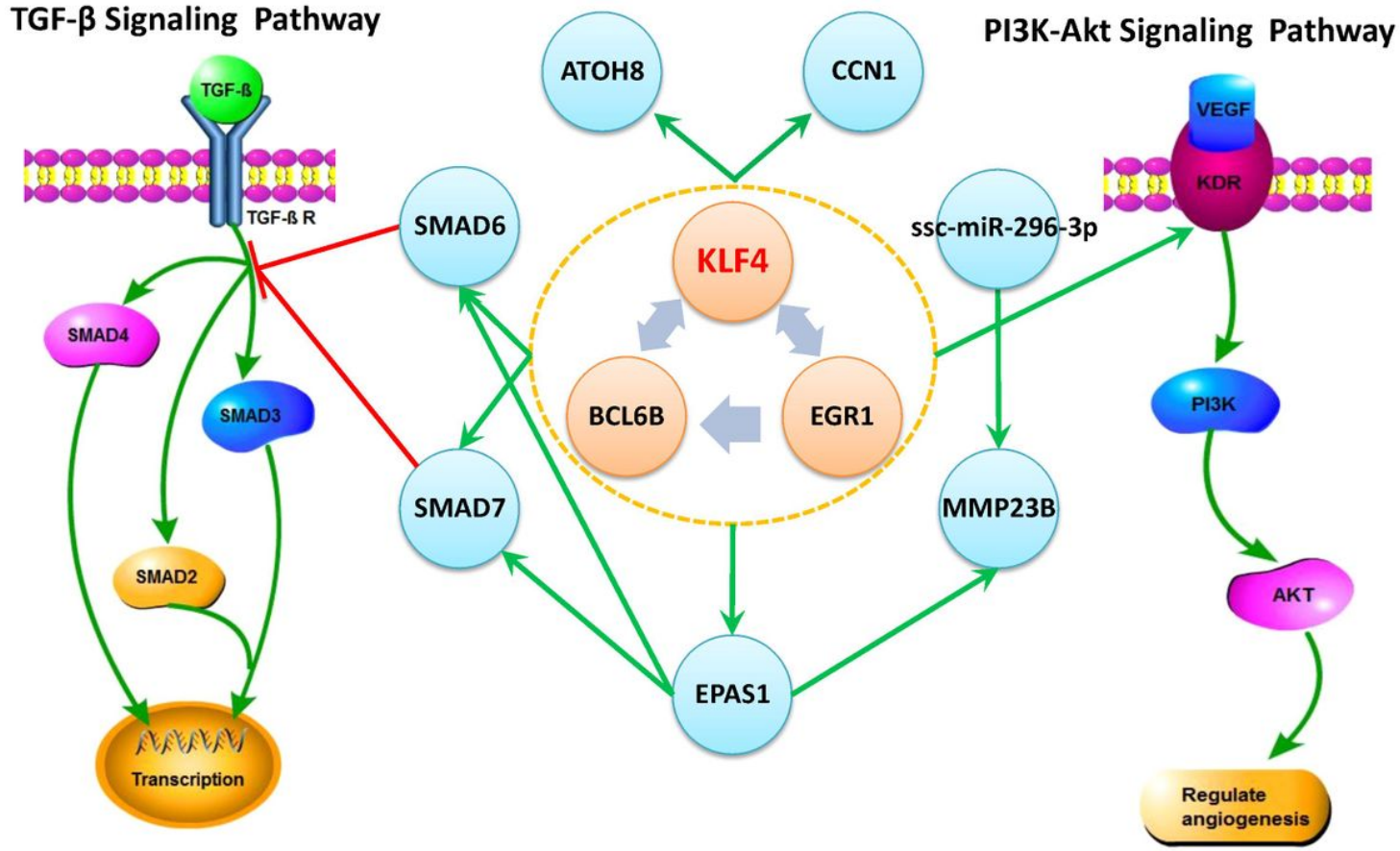

Figure 6

“KLF4-EGR1-BCL6B” transitive triplet and their regulated genes in Tibetan pig lung tissue The transitive triplet formed by KLF4-EGR1-BCL6B regulates EPAS1, SMAD6, SMAD7, KDR, ATOH8, CCN1 genes, and mediates the TGF-β and PI3K-Akt signaling pathways by regulating SMAD6, SMAD7 and KDR genes, respectively. The green edge in the figure represents regulation, and the red edge represents inhibition.

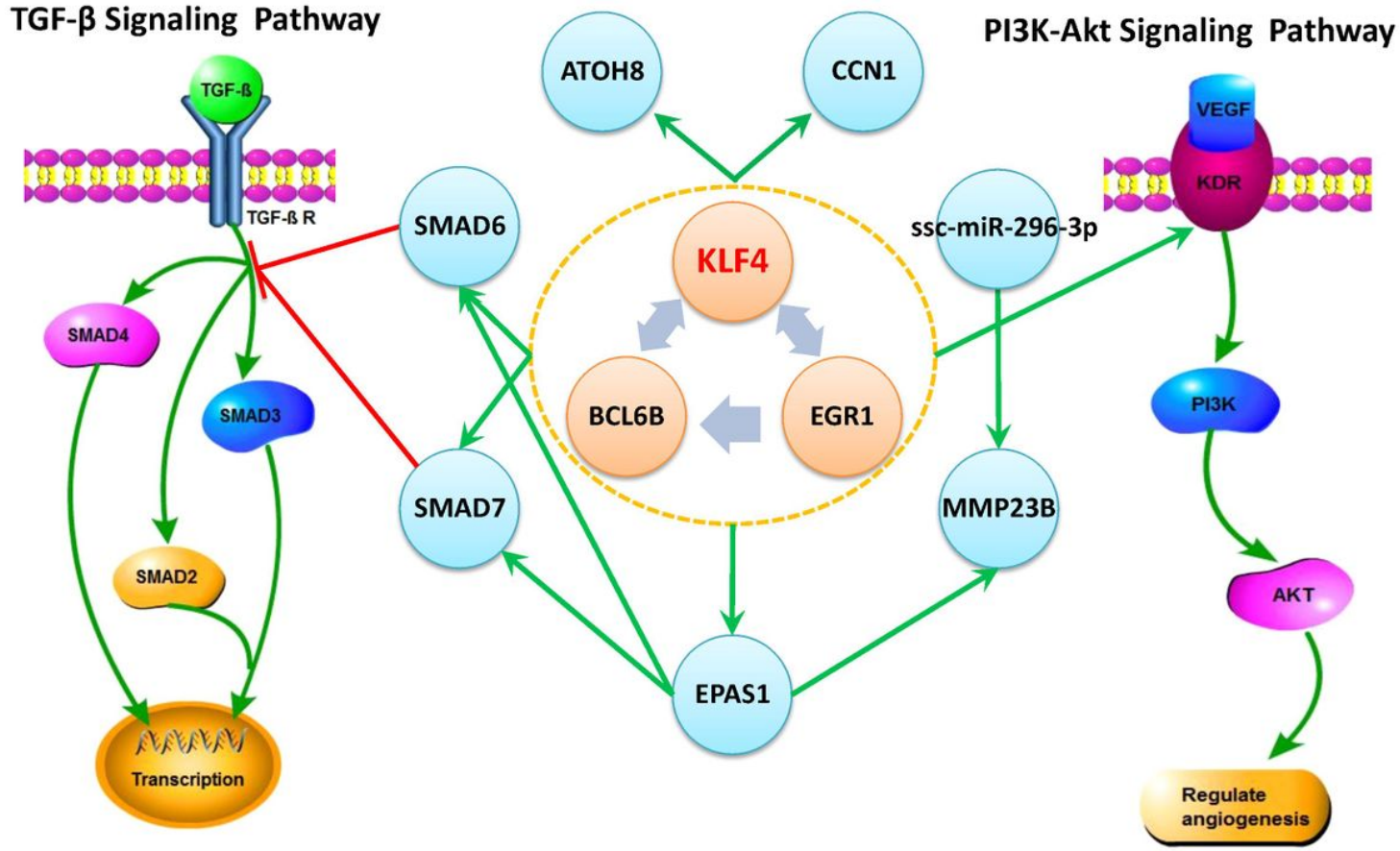

Figure 6

“KLF4-EGR1-BCL6B” transitive triplet and their regulated genes in Tibetan pig lung tissue The transitive triplet formed by KLF4-EGR1-BCL6B regulates EPAS1, SMAD6, SMAD7, KDR, ATOH8, CCN1 genes, and mediates the TGF-β and PI3K-Akt signaling pathways by regulating SMAD6, SMAD7 and KDR genes, respectively. The green edge in the figure represents regulation, and the red edge represents inhibition.

Supplementary Files

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

- FigureS1.jpg
- FigureS1.jpg
- FigureS2.jpg
- FigureS2.jpg

- Table17.docx
- Table17.docx