

# SCIENTIFIC REPORTS



OPEN

## Differentially expressed genes in the caecal and colonic mucosa of Landrace finishing pigs with high and low food conversion ratios

Zhen Tan<sup>1</sup>, Yuan Wang<sup>1</sup>, Ting Yang<sup>1</sup>, Kai Xing<sup>1</sup>, Hong Ao<sup>2</sup>, Shaokang Chen<sup>3</sup>, Fengxia Zhang<sup>1</sup>, Xitong Zhao<sup>1</sup>, Jianfeng Liu<sup>1</sup> & Chuduan Wang<sup>1</sup>

The feed conversion ratio (FCR) is an essential economic trait for pig production, and is directly related to feed efficiency. Studies identifying the differential expression of functional genes involved in biological and molecular mechanisms in the intestine in relation to growth performance are rare. In this study, RNA-Seq was used to identify transcriptomes in caecal and colonic mucosal tissues in order to determine the differential expression of genes from two full-sibling pairs and two half-sibling pairs of Landrace finishing pigs with opposing FCR phenotypes. In total, 138 (comparison of high and low FCR in caecal mucosa), 64 (comparison of high and low FCR in colonic mucosa), and 165 (contrast between the caecal and colonic mucosa) differentially expressed genes were identified. Some of these genes were functionally related to energy and lipid metabolism, particularly short chain fatty acids metabolism, as well as gastrointestinal peristalsis and ion transport. Functional annotation were performed to identify differentially expressed genes, such as *GUCA2A*, *GUCA2B*, *HSP70.2*, *NOS2*, *PCK1*, *SLCs*, and *CYPs*, which may positively influence feed efficiency in Landrace pigs. These differentially expressed genes need to be further tested for candidate genes that are related to feed efficiency.

The performance of livestock animals is influenced by genetics and the environment, including diet and rearing conditions<sup>1</sup>. Feed accounts for more than 60% of the costs of pig production. Therefore, improving feed efficiency (FE) is an important method for reducing costs in the pig industry. FE can be measured using the feed conversion ratio (FCR). The FCR is the feed intake divided by the weight gained during a specified period. Thus, an animal with a high FCR value is less efficient at converting feed into body mass than one with a low FCR. Previous studies have indicated that the heritability values for FCR range from 0.13 to 0.31<sup>2,3</sup>.

Gene expression profiling in the liver and adipose tissue in pigs has revealed that the biological processes of lipid metabolism, mitochondrial activity, and glucose synthesis are associated with FE<sup>4</sup>. In cattle, differentially expressed genes (DEGs) between high and low residual feed intake (RFI) groups are related to cell growth and differentiation, lipid metabolism, and carbohydrate metabolism<sup>5</sup>.

Genetic lines of chickens with different feed efficiency have also been associated with changes in gut physiology and gut microbial composition<sup>6</sup>. There are a variety of ways through which the microflora can have a negative impact on the host, including the use of excessive amounts of energy, and the diversion of energy towards the immune system by inducing inflammatory responses<sup>7</sup>. Some nutrients, such as resistant starch (RS), cannot be digested completely and need to be fermented by caecal and colonic microbes to produce short-chain fatty acids (SCFAs)<sup>8,9</sup>.

In this study, we report, for the first time, a complete dataset detailing the transcriptome of caecal and colonic mucosal tissues identified from female near-market weight Landrace pigs with high or low FCR using RNA-Seq. We identified genes that were significantly differentially expressed between locations or groups to investigate whether they are associated with growth traits and/or microbial digestion. The putative candidate genes identified

<sup>1</sup>National Engineering Laboratory for Animal Breeding, MOA Key Laboratory of Animal Genetics and Breeding, Department of Animal Genetics and Breeding, China Agricultural University, Beijing, China. <sup>2</sup>The State Key Laboratory of Animal Nutrition, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, China. <sup>3</sup>Beijing General Station of Animal Husbandry, Beijing, China. Correspondence and requests for materials should be addressed to C.W. (email: [cdwang@cau.edu.cn](mailto:cdwang@cau.edu.cn))

Group	Sample	Total reads (million)	Total mapped reads (%)	Unique matched reads (%)	Multi matched reads (%)	CDS exons (%)	5'UTR exons (%)	3'UTR exons (%)	Intron (%)	Intergenic (%)
Hce	H1caecum	53.58	72.70	68.45	4.26	64.57	1.45	12.28	8.04	13.66
	H2caecum	52.68	71.05	66.83	4.22	69.85	1.73	10.91	5.52	11.99
	H3caecum	51.18	71.55	67.36	4.19	64.88	1.44	11.93	8.04	13.71
	H4caecum	58.62	71.04	66.91	4.12	58.72	1.05	12.76	11.57	15.91
Lce	L1caecum	61.19	72.94	68.69	4.25	67.88	1.68	11.45	6.44	12.55
	L2caecum	48.72	73.00	68.68	4.32	66.26	1.59	12.00	6.67	13.48
	L3caecum	66.97	72.90	68.62	4.28	68.15	1.67	11.26	6.36	12.56
	L4caecum	58.57	70.56	66.37	4.20	65.84	1.54	11.12	7.98	13.51
Hco	H1colon	69.50	71.48	67.12	4.36	68.89	1.72	11.04	6.00	12.35
	H2colon	77.40	70.15	65.82	4.33	69.16	1.77	10.53	6.22	12.32
	H3colon	69.08	70.01	65.71	4.30	64.42	1.54	10.86	9.06	14.13
	H4colon	54.37	73.02	68.90	4.12	65.32	1.38	12.35	7.42	13.54
Lco	L1colon	64.75	69.24	64.83	4.41	68.20	2.00	9.26	7.71	12.83
	L2colon	60.30	72.25	67.90	4.35	64.69	1.51	11.38	8.25	14.17
	L3colon	69.99	69.84	65.49	4.34	68.86	1.84	9.51	7.22	12.56
	L4colon	63.21	72.54	68.28	4.28	66.03	1.61	11.80	7.22	13.35

**Table 1.** RNA sequencing of mRNA from the caecal and colonic mucosa of Landrace female finishing pigs with high and low FCR. Uses *Sus scrofa* 10.2 as the reference genome annotation to classify the mapping tags to the different regions. Hce, caecal samples of HFCR group; Lce, caecal samples of LFCR group. Hco, colonic samples of HFCR group; Lco, colonic samples of LFCR group.

could lead to an improved understanding of digestion in the large intestine while providing new insights into growth traits.

## Results

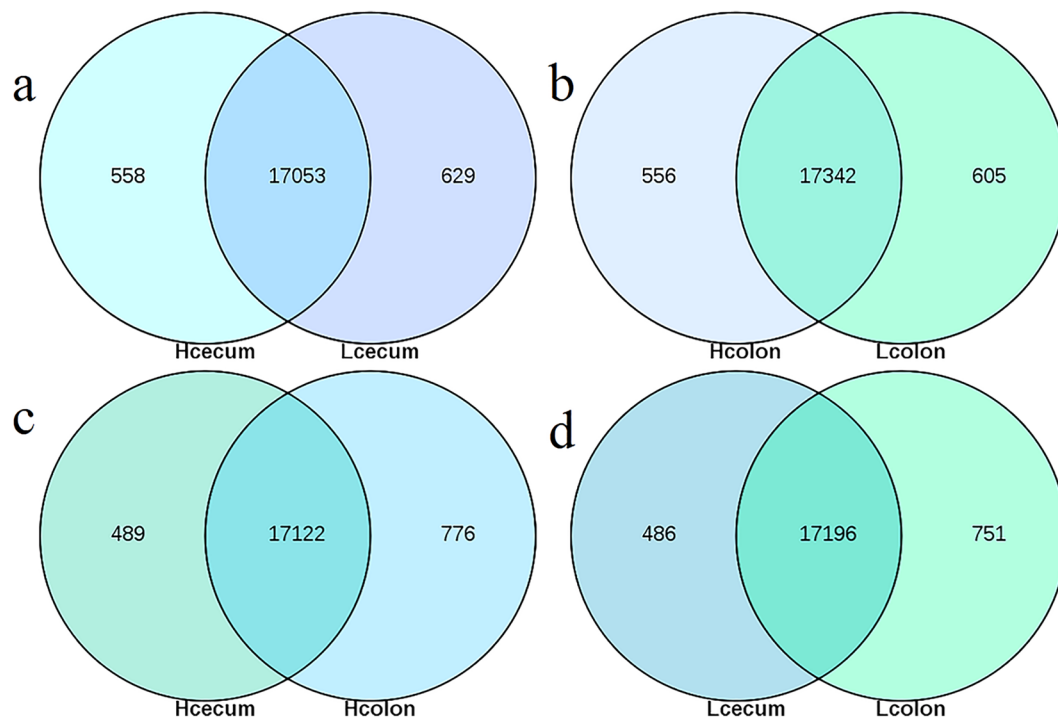
**Overview of RNA-Seq data and differentially expressed genes analyzed.** cDNA libraries were constructed for each RNA-Seq sample ( $n = 16$ ), sequencing was conducted on an Illumina HiSeq. 4000 platform, and 150-bp paired-end reads were generated. As shown in Table 1, RNA-Seq analysis yielded between 48.72 and 77.40 million total clean reads per sample, and the percentage of mapped reads ranged from 69.24 to 73.02%. The percentages of mapped reads were similar between samples, and more than half were located within an exon, with percentages ranging from 58.72 to 69.85%. Other reads mapped within the untranslated region, introns, and intergenic regions, and only the uniquely mapped reads were analysed.

Expression levels of all genes were calculated using the RPKM method and Tophat 2 software, and the results of all groups are presented in Supplementary Table S2. The total number of expressed genes in the caecum and colonic mucosa were similar between groups (17,611–17,947) in four cDNA libraries. Most genes were expressed in both groups, and there were more than 17,000 common genes in all four comparisons between the caecal and colonic mucosa in the high and low FCR pigs (Fig. 1).

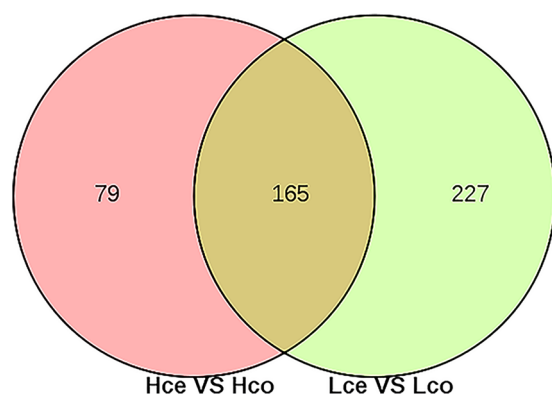
Differentially expressed genes (DEGs) were identified in each comparison, with criteria of a  $\geq 2$ -fold difference and a  $q$ -value of less than 0.05 ( $|\log_2FC| \geq 1$ ,  $q < 0.05$ ). The up- and down-regulated DEGs are listed for each comparison, and the details of all DEGs in the comparisons are shown in Supplementary Tables S3 and S4. In total, 138 genes were different in group caecal mucosa of high FCR compared with caecal mucosa of low FCR (Hce vs Lce), 64 genes were different in group colonic mucosa of high FCR compared with colonic mucosa of low FCR (Hco vs Lco), 244 genes were different in group caecal mucosa of high FCR compared with colonic mucosa of high FCR (Hce vs Hco), and 392 genes were different in group caecal mucosa of low FCR compared with colonic mucosa of low FCR (Lce vs Lco). Subsequently, 165 genes from the intersection of Hce vs Hco and Lce vs Lco were considered as the differentially expressed genes between caecal and colonic mucosal tissues (Fig. 2). Tables 2 and 3 list 40 and 13 annotated differentially expressed genes in caecal and colonic mucosa respectively, with either higher or lower expression in high FCR compared to low FCR pigs.

**Functional annotation clustering of DEGs.** Due to the limited annotation of the reference genome (*S. scrofa* 10.2), the DEGs were converted to human orthologs, and the online program Visualization and Integrated Discovery (DAVID) was used for gene ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis. GO enrichment analyses contained 'biological process', 'molecular function', and 'cell component'. Distinct functional separation between the DEGs and GO terms was observed through a graphical network drawn by ToppCluster software for each comparison. STRING was used to build protein-protein interactions nets of the DEGs encoding proteins.

For GO enrichment and KEGG pathway analysis of DEGs between caecal mucosa in the high and low FCR groups, multiple significant pathways and GO terms are shown in Supplementary Table S5. The DEGs were primarily enriched in categories related to cellular processes, single-organism processes, biological regulations, metabolic processes, responses to stimuli, cells, cell parts, organelles, binding, and catalytic activity (Fig. 3). Furthermore, KEGG pathway analysis identified significant pathways that were mostly related to immunity and disease. Fatty acid metabolism was also mapped, but did not reach significance ( $P = 0.078$ ).



**Figure 1.** Venn diagrams representing gene expression in four groups of pigs. Levels of gene expression in the caecal and colonic mucosa of pigs with high or low FCR: Hcaecum and Lcaecum (a), Hcolon and Lcolon (b), Hcaecum and Hcolon (c), and Lcaecum and Lcolon (d) are shown. The total number of expressed genes in each group is shown. The number of common genes is shown in the overlapping segments.



**Figure 2.** Venn diagrams showing DEGs between the caecal and colonic mucosa in the high and low FCR groups.

The functional network for DEGs was generated by ToppCluster as shown in Fig. 4. Most terms and pathways for DEGs were in categories associated with immunity and disease, which is consistent with previous results obtained through DAVID. The down-regulated genes encoded proteins associated with structure or function. The up-regulated genes were mostly related to immunity and disease. We identified a gene, complement component 4 binding protein alpha (C4BPA), which was found in groups of both up-regulated and down-regulated genes. This was because there were two transcripts belonging to the C4BPA gene, with one classified as being up-regulated and another classified as being down-regulated, when comparing the Hce group to the Lce group. C4BPA has been demonstrated to be involved in the pathway of complement and coagulation cascades, while the C4BP (C4b-binding protein) is known as a soluble complement inhibitor in the classical pathway of complement systems, and is important for body homeostasis<sup>10</sup>.

A protein interaction network analysis was generated from the DEGs by STRING (Fig. 5), for Hce compared with Lce. The nitric oxide synthase 2 (NOS2) gene participated in multiple interactions in this network of the comparison. This gene encodes a nitric oxide synthase that is expressed in the liver, and its expression is induced by a combination of lipopolysaccharide and certain cytokines. The genes in this network were related to some signalling pathways, such as cytokine, interferon, interferon gamma, and citrulline and aspartate metabolism<sup>11</sup>.

Gene	Locus	FPKM_Hce	FPKM_Lce	Log2(Fold change)	q_value	Gene description
GUCA2A	6:156476367-156478316	1822.97	641.96	-1.51	6.16E-03	Guanylin
HSP70.2	GL896522.1:2-1799	1450.72	417.43	-1.80	3.40E-03	<i>Sus scrofa</i> heat shock protein 70.2
ISG12(A)	7:122215477-122220545	1395.64	417.83	-1.74	3.40E-03	interferon, alpha-inducible protein 27
UBD	7:25330260-25332807	1242.79	287.87	-2.11	3.40E-03	ubiquitin D
IGLV-11	14:52228869-52229344	1169.88	534.43	-1.13	1.52E-02	immunoglobulin lambda variable 11
S100G	X:13996225-14092759	1043.02	312.73	-1.74	6.16E-03	S100 calcium binding protein G
ASS1	1:304454128-304508998	877.10	234.95	-1.90	3.40E-03	argininosuccinate synthase 1
IGLV-2	14:52404074-52404828	564.54	191.16	-1.56	3.40E-03	immunoglobulin lambda variable 2
MS4A12	2:10736890-10753331	530.87	254.17	-1.06	1.90E-02	membrane spanning 4-domains A12
CLDN4	3:10776891-10780656	236.74	101.29	-1.22	3.40E-03	claudin 4
HBB	9:5632714-5634344	235.47	93.61	-1.33	3.40E-03	hemoglobin, beta
GZMB	7:7983196-79834605	204.96	97.37	-1.07	1.52E-02	granzyme B
UBE2L6	2:12987516-13002993	177.71	83.94	-1.08	3.40E-03	ubiquitin conjugating enzyme E2L 6
ISG15	GL895967.2:58448-59139	139.46	29.59	-2.24	3.40E-03	ISG15 ubiquitin-like modifier
GZMA	16:36385657-36393941	117.24	48.25	-1.28	3.40E-03	granzyme A
GNLY	3:61952961-61957608	114.36	52.31	-1.13	3.40E-03	antimicrobial peptide NK-lysin precursor
GBP2	4:139612346-139633862	112.97	44.49	-1.34	3.40E-03	guanylate binding protein 2, interferon-inducible
NOS2	12:46022453-46053026	103.76	35.61	-1.54	3.40E-03	nitric oxide synthase 2
SLC30A10	10:11615849-11627111	85.04	36.22	-1.23	3.40E-03	solute carrier family 30 member 10
ENTPD8	1:314009380-314012067	78.46	39.17	-1.00	3.40E-03	ectonucleoside triphosphate diphosphohydrolase 8
PLEKHG6	5:66793344-66811338	66.27	28.88	-1.20	6.16E-03	pleckstrin homology and RhoGEF domain containing G6
ANPEP	7:60240144-60262914	64.18	22.43	-1.52	1.10E-02	alanyl aminopeptidase, membrane
SLA-7	7:27613956-27617558	63.35	27.62	-1.20	3.40E-03	MHC class I antigen 7
KIFC2	4:415602-422897	58.43	24.88	-1.23	3.40E-03	kinesin family member C2
OAS1	14:41230164-41288614	58.41	25.10	-1.22	4.19E-02	2'-5'-oligoadenylate synthetase 1
7SK	7:134400748-134401079	54.67	22.78	-1.26	2.47E-02	7SK RNA
IGSF23	6:47050121-47057994	51.73	15.06	-1.78	3.40E-03	immunoglobulin superfamily member 23
PARP15	13:147507154-147531686	45.23	6.90	-2.71	3.40E-03	poly(ADP-ribose) polymerase family member 15
C4BPA	9:74195615-74207951	39.17	17.67	-1.15	3.74E-02	complement component 4 binding protein alpha
SLC14A1	1:105418611-105444533	37.42	8.04	-2.22	3.40E-03	solute carrier family 14 member 1
OASL	14:43414433-43431814	30.69	7.16	-2.10	3.40E-03	2'-5'-oligoadenylate synthetase-like
FRMD1	1:2657896-2672308	11.75	3.31	-1.83	3.40E-03	FERM domain containing 1
IGKV-3	3:59934903-59935620	110.74	732.09	2.72	3.40E-03	immunoglobulin kappa variable 3
IGLV-4	14:52392751-52393255	68.83	552.68	3.01	3.40E-03	immunoglobulin lambda variable 4
CH242-307A4.1	3:67957189-67973255	48.89	337.90	2.79	3.40E-03	lithostathine-like precursor
IGLV-5	14:52384767-52385271	27.55	171.00	2.63	3.40E-03	immunoglobulin lambda variable 5
PIGR	9:73889671-73899096	71.17	150.36	1.08	3.40E-03	polymeric immunoglobulin receptor precursor
REG3G	3:68078690-68081405	5.92	129.41	4.45	3.40E-03	regenerating islet-derived 3 gamma
C4BPA	9:74110908-74154259	45.63	104.31	1.19	3.40E-03	complement component 4 binding protein alpha
CHAC1	1:145522852-145525603	2.49	64.03	4.68	3.40E-03	ChaC glutathione specific gamma-glutamylcyclotransferase 1

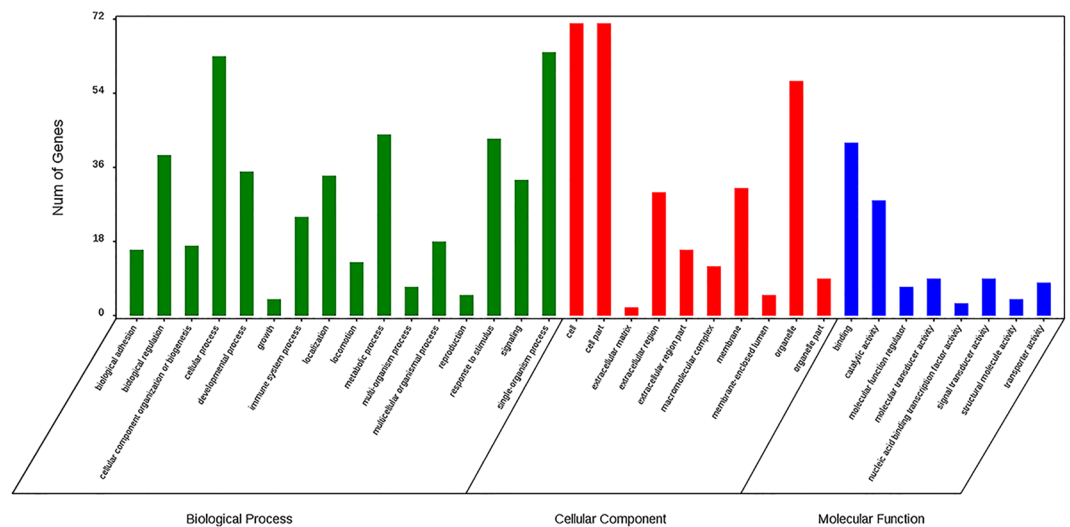
**Table 2.** List of 40 differentially expressed genes (DEGs) between caecal mucosa with high and low FCR.

DEGs in the colonic mucosa of the high and low FCR groups were functionally annotated, and were found to be related to cellular processes, single-organism processes, metabolic processes, cells, cell parts, organelles, binding, and catalytic activity (Fig. S2). Multiple significant GO terms are shown in Supplementary Table S6. Certain terms were significant and were related to metabolism ( $P < 0.05$ ), such as cell recognition, ion transport (a biological process), beta-galactosidase activity, calcium channel activity, and carbohydrate binding (molecular function). Alpha-linolenic acid metabolism was mapped in the KEGG pathway, but did not reach significance ( $P = 0.097$ ).

We found differences the common DEGs present in the caecal mucosa to those in the colonic mucosa in both high and low FCR groups. Subsequently, GO terms and KEGG pathways of DEGs were determined between the

gene	locus	FPKM_Hco	FPKM_Lco	log2(fold_change)	q_value	Gene description
TRPV6	18:7636964-7652271	14.71	1.73	-3.09	3.40E-03	transient receptor potential cation channel subfamily V member 6
SPINK1	2:155635124-155640264	32976.60	10332.90	-1.67	3.40E-03	serine peptidase inhibitor, Kazal type 1
IGLV-2	14:52404074-52404828	303.47	112.61	-1.43	3.40E-03	immunoglobulin lambda variable 2
TMEM120A	GL894019.2:55427-61934	42.33	16.05	-1.40	3.40E-03	transmembrane protein 120A
CYP2B6	6:44929853-44978800	56.83	22.22	-1.35	1.71E-02	cytochrome P450 family 2 subfamily B member 6
FRMD1	1:2657896-2672308	33.53	14.14	-1.25	3.44E-02	FERM domain containing 1
CD209	2:71917404-71931695	24.01	55.44	1.21	3.40E-03	CD209 molecule
SLC2A13	5:73929816-74023533	39.91	105.25	1.40	3.40E-03	solute carrier family 2 member 13
CH242-307A4.1	3:67957189-67973255	3.18	9.41	1.56	1.32E-02	lithostathine-like precursor
IGLV-5	14:52384767-52385271	22.54	90.57	2.01	3.40E-03	immunoglobulin lambda variable 5
IGKV-3	3:59934903-59935620	39.11	169.95	2.12	6.16E-03	immunoglobulin kappa variable 3
IGLV-4	14:52392751-52393255	26.33	226.01	3.10	3.40E-03	immunoglobulin lambda variable 4
CHAC1	1:145522852-145525603	1.16	10.11	3.13	3.40E-03	ChaC glutathione specific gamma-glutamylcyclotransferase 1

**Table 3.** List of 13 differentially expressed genes (DEGs) between colonic mucosa with high and low FCR.

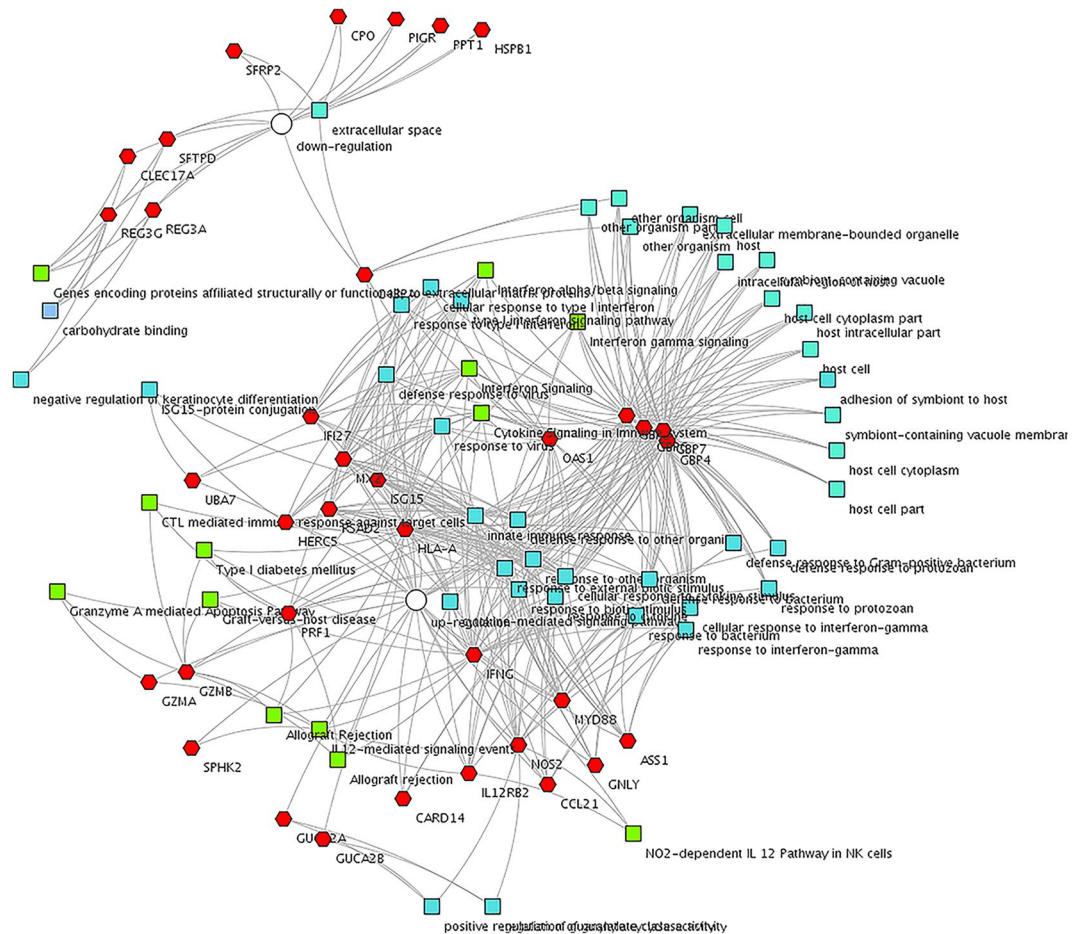


**Figure 3.** Enriched GO terms and the number of corresponding DEGs for each term in the Hce vs Lce group.

caecal and colonic mucosa. The primary GO terms clustered were cellular process, single-organism process, and metabolic process in 'biological process'; cells, cell parts, and organelles in 'cellular component'; and binding and catalytic activity in 'molecular function' (Fig. S3). Pathway analysis identified significant pathways that were mostly related to nutrient metabolism, and the PPAR signalling pathway was also mapped, reaching significance ( $P < 0.05$ ) (Table S7).

Results for the DEG functional network are presented in Fig. S4, and pathways involving up-regulated DEGs were found to be associated with ascorbate and aldarate metabolism. The down-regulated DEGs were related to trans-membrane transport and drug metabolism. A protein interaction network of DEGs was also generated, and a few protein-coding genes were found to interact with others (Fig. S5). We found that forkhead box A2 (FOXA2) was associated with three other genes. This gene encodes a member of the forkhead class of DNA-binding proteins. These hepatocyte nuclear factors are transcriptional activators of liver-specific genes, such as albumin and transthyretin, and they also interact with chromatin. FOXA2 plays a pivotal role in regulating intestinal epithelial cell function<sup>12</sup>.





**Figure 4.** Relationship between DEGs, pathways, and GO terms of Hce vs Lce. Functional association analysis performed by ToppCluster based on pathway networks showing enriched terms from Gene Ontology and pathways. The top part of the figure depicts significant enrichments for up-regulated DEGs in Hce compared with Lce; the lower part depicts significant enrichments for down-regulated DEGs in Hce compared with Lce. Red hexagon: DEG; green square: pathways; blue square: biological processes; grey square: molecular functions.

**DEGs compared with the QTL database.** Using the Pig Quantitative Trait Locus (QTL) Database (Pig QTLdb), production and production-associated QTLs with a confidence interval of less than 5 Mb were selected<sup>13</sup>. Four DEGs in Hco vs Lco, 15 DEGs in Hce vs Lce, and eight DEGs in ce vs co were available (Table S8). The DEGs found in each comparison represented additional candidate genes affecting pig production.

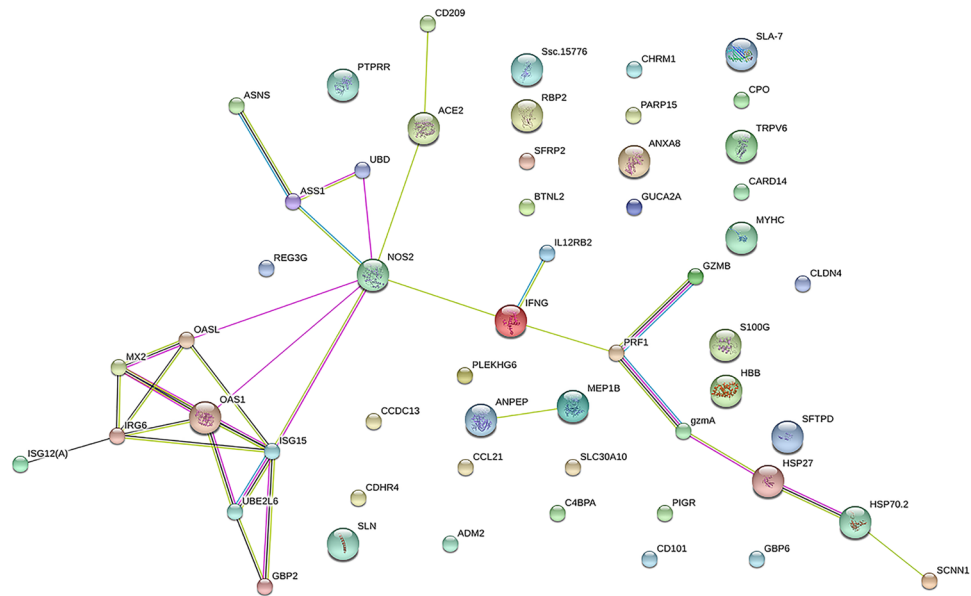
## Discussion

In this study, systematic transcriptome profiling of caecal and colonic mucosal tissues from two full-sib pairs and two half-sib pairs of pigs with opposing FCRs was performed using high-throughput RNA-Seq technology. This could reduce false-positive results caused by genetic background noise and the number of replicates<sup>13,14</sup>.

The phenotypic trait, feed efficiency (FE), is an important economic trait for meat production in the pig industry. Thus, FE should be improved to reduce the costs of pig production. To date, FE can be measured by two highly correlated indicators, i.e., FCR and RFI, and a low FCR/RFI level signifies improvement in FE<sup>15,16</sup>. It is important to investigate the mechanisms of FE for pig breeding. Studies have suggested that energy metabolism in the liver and muscle tissues is essential for the regulation of FE of pigs<sup>3,17,18</sup>. Additionally, FE has also been related to the intestinal innate immune response and enhanced antimicrobial enzyme activity, and lower serum IL-8, myeloperoxidase and endotoxin levels have been found to be associated with high FE<sup>19,20</sup>.

Considerable colonisation of the intestine by microorganisms, especially in the large intestine, provides a homeostatic environment that affects the function of the gastrointestinal tract, disease resistance, health status, and animal performance<sup>9,21</sup>. The short-chain fatty acids (SCFA) produced in the large intestine by microbial fermentation of dietary fibre are rapidly absorbed from the large intestinal mucosa to provide the host with energy for basal metabolism<sup>22</sup>. Additionally, the caecal mucosa plays an important role in urea recycling in the gut of monogastric animals<sup>23</sup>.

In the study of caecal and colonic microbiota using the same animals<sup>24</sup>, Firmicutes and Bacteroidetes were the most abundant phyla in caecal and colonic microbiota of pigs in both groups, consistent with other studies<sup>25,26</sup>, and the dominant genera were *Prevotella* and *Bacteroides*. By comparing the high FCR and low groups in caecal



**Figure 5.** STRING analysis shows that DEGs are involved in known and predicted protein-protein interactions. STRING is used to analyse DEGs in the mucosa of Hce and Lce. The network nodes stand for those genes shown in Tables S3 and S4. Different coloured lines represent seven types of evidence used to predict associations. Red line: fusion evidence; green line: neighbourhood evidence; blue line: co-occurrence evidence; purple line: experimental evidence; yellow line: text mining evidence; light blue line: database evidence and black line: co-expression evidence.

and colonic microbiota respectively, the differentially predicted genes were most enriched in carbohydrate metabolism, followed by amino acid metabolism and energy metabolism.

*Prevotella* sp. CAG:604 were the most different species in both low groups of caecal and colonic microbiota. *Prevotella* sp. CAG:604 contains some genes that encode proteins involved in nutrient and energy metabolism, such as BN731\_01873, ychF, gpmI, queF, speA, and fmt. The presence of fructo-oligosaccharides and starch in the lower intestine may cause a higher abundance of *Prevotella*<sup>27</sup>. There were species of *Lactobacillus* enriched in the high group of both caecal and colonic microbiota, with *Lactobacillus* spp. often considered as probiotics. Several species of *Lactobacillus* belong to lactic acid bacteria (LAB), which convert carbohydrates to lactic acid in homofermentation or heterofermentation, or to acetic acid in heterofermentation<sup>28</sup>. The microorganisms actively participate in the process of material digestion in the lower intestine, and these nutrients are absorbed through the intestinal mucosa into the circulatory system. The genes expressed in the mucosa mainly maintain the healthy homeostasis of the intestinal environment, and some participate in the identification, catalysis and synthesis of substances.

We identified DEGs between pigs with different FCRs by profiling the caecal and colonic mucosa transcriptome, and identified nutrient metabolism pathways through GO and pathway analysis that were related to FE variation in pigs.

Previous studies of FE have identified many DEGs by RNA-Seq that are considered important for improving the growth of pigs<sup>3,18</sup>. In contrast to previous studies, the current study provides the first statistical analysis of growth traits to detect DEGs generated from intestinal mucosal tissue samples by RNA-Seq data. Based on transcriptome data, we compared the entire expression profile and analysed DEGs in caecal and colonic mucosal tissues in high and low FCR Landrace pigs. DEGs related to nutrient metabolism and growth, or to certain immune responses and diseases, may represent potential candidate genes that affect animal growth.

Although the top DEGs identified in the present study corresponded to immunity and disease, mucosal-associated bacterial communities help the host to displace pathogens, digest nutrients lost from the intestines, synthesise vitamins, and mature the immune system<sup>29,30</sup>. The mucosa is the first barrier involved in protecting the internal environment of the host. In addition, we identified several genes and pathways reported to be involved in important metabolic processes.

From the DEGs up-regulated in the caecal mucosa of the high FCR group compared with the low FCR groups, we identified several genes involved in growth traits (Table S3, Figs 3 and 4). Guanylate cyclase activator 2A (GUCA2A) and guanylate cyclase activator 2B (GUCA2B) are related to ion and fluid homeostasis, irritable bowel syndrome (IBS), abdominal pain<sup>31,32</sup>, growth, and intestinal barrier integrity<sup>33</sup>, and were expressed in the gastrointestinal tract<sup>34</sup>. GUCA2A, GUCA2B, and GUCY2C were reported to bind to the GC-C signalling pathway via several steps, to sustain epithelial barrier integrity and affect IBD<sup>35</sup>. GUCA2B has advantages in the regulation of feeding, energy homeostasis, body mass, and metabolism<sup>36,37</sup>. In addition, GUCA2A binds to and increases the activity of guanylate cyclase, which is involved in glycogenolysis, and relaxation of smooth muscle. The gene GUCA2A also participates in biological processes of digestion by GO term annotation and GUCA2A and CUCA2B were involved in pathways of miscellaneous digestion events. As such, the higher expression of

genes *GUCA2A* and *CUCA2B* could be beneficial for feed digestion and intestinal protection from external stimuli. Therefore, *GUCA2A* and *GUCA2B* might be good candidate genes for growth traits.

The results of other studies have suggested that the *HSP70.2* (heat shock 70 kDa protein 1B) gene might be a candidate gene involved in obesity and type 2 diabetes<sup>38,39</sup>. As well as being involved in the immune system and cellular response, these genes are involved in ATP binding, ATPase activity and ATP metabolic processes, which provide the direct energy source for the body.

*NOS2* is also considered to be a candidate gene, and controls gastrointestinal peristalsis<sup>11</sup>. By annotation of the GO term and pathway, *NOS2* was found to be closely related to growth, as well as carboxylic acid binding, cAMP-dependent protein kinase regulator activity, cadherin binding in molecular function; responses to lipids, protein catabolic processes, nitrogen compound transport, responses to bacteria in biological processes; and pathways of arginine and proline metabolism. These activities are related to the metabolism of substances and indirectly affect the growth efficiency of animals.

Gene expression in caecal mucosa of high FCR group was compared with the low FCR group. Down-regulated genes were related to immunity and disease (Table 2), and up-regulated genes were connected with intestinal health and homeostasis. These genes (such as *GUCA2A*, *HSP70.2*, *NOS2*, *ASS1*, *HBB*, *SLC30A10*) can directly or indirectly affect nutrient absorption and utilization by animals, and therefore feed efficiency.

These different genes can be initially used as candidate genes to distinguish the difference in feed efficiency of pigs, although intensive research needs to be done in future to confirm this.

Within the up-regulated DEGs in the colonic mucosa of high FCR animals compared with that of the low FCR animals, the serine protease inhibitor Kazal type 1 referred to as pancreatic secretory trypsin inhibitor (*SPINK1*) that is known to inhibit trypsin and regulate homeostasis is secreted by the pancreatic acinar cells into the pancreatic juice<sup>40</sup>. This gene has been correlated with cancer growth and progression<sup>41</sup>.

Furthermore, high FPKM values were found in the caecal mucosa in both FCR groups. The cytochrome P450s 2B6 (*CYP2B6*) gene is involved in the synthesis of multiple drugs<sup>42</sup>, in the biotransformation of endogenous and exogenous compounds, and in the metabolism of many toxicants<sup>43</sup>.

There were not many different genes in the colonic mucosa of high FCR group compared with low group after we annotated these and removed genes of low expression. The down-regulated genes were also related to the immune system and disease, and interestingly, two genes *IGLV-2* and *FRMD1* were up-regulated, and five genes *CH242-307A4.1*, *IGLV-5*, *IGKV-3*, *IGLV-4* and *CHAC1* were down-regulated in both caecal and colonic locations.

Comparison of the caecal mucosa with the colonic mucosa in both the high and low FCR groups revealed differences between locations. Analysis of GO enrichment, KEGG pathway, protein-protein interactions, and QTL location (Tables S3, S8, Figs S3, S4), revealed that CYPs and solute carrier (SLC) gene families, and some metabolism-related genes, were up-regulated in the caecum compared with the colon. CYPs are of prime importance in the oxidation of unsaturated fatty acids and in catalysing metabolic reactions<sup>44</sup>. SLCs contain many glucose transporters, and are known as tumour suppressors<sup>45</sup>. SLCs directly facilitate the solute flux gradient, and indirectly power the solute electrochemical gradient<sup>46</sup>.

Phosphoenolpyruvate carboxykinase-1 (*PCK1*) was reported to participate in lipid metabolism<sup>47</sup>, is the rate-limiting enzyme in hepatic gluconeogenic pathways, and is closely related to obesity<sup>48</sup>. *PCK1* was also involved in pathways of glucose metabolism and pyruvate metabolism.

In summary, annotated DEGs in the caecal and colonic mucosal tissues of high and low FCR pigs were analysed and compared. The results revealed that the up-regulated DEGs related to loss of starch from the small intestine, some nucleotides, and drug metabolism. Up-regulated DEGs might also be directly or indirectly involved in FE regulation of pigs. Down-regulated DEGs were mostly related to immunity and disease in caecal mucosa or colonic mucosa tissues. *GUCA2A*, *GUCA2B*, *HSP70.2*, *NOS2*, *PCK1*, SLCs, and CYPs are possible candidate genes for FE in pigs. Overall, the current investigation utilizing RNA-Seq will add to the knowledge of gene expression in the pig intestine.

In conclusion, 138 (Hcaecum vs Lcaecum), 64 (Hcolon vs Lcolon), and 165 (caecum vs colon) differentially expressed genes were identified in the caecal and colonic mucosa of two Landrace pig groups with diverging FCR. Some of these genes were functionally related to energy and lipid metabolism, in particular short chain fatty acids, as well as gastrointestinal peristalsis and biotransformation. GO enrichment, KEGG pathway, protein-protein interaction, and QTL location analysis were performed to identify DEGs, such as *GUCA2A*, *GUCA2B*, *HSP70.2*, *NOS2*, *PCK1*, SLCs, and CYPs, which may be potential candidate genes associated with growth traits in Landrace pigs. As only four pairs of pigs were used in this study, these results need be validated using a larger pig cohort in future. The findings of this study extend our understanding of the molecular mechanisms of feed efficiency that regulate growth trait in pigs.

## Methods

**Animals, phenotypes, and tissues.** In this experiment, we used female Landrace pigs provided by Tianjin Ninghe primary pig breeding farm (Ninghe, China). Pigs (total 120) were housed in an environmentally controlled room (10 pigs in each pen), and given feed and water *ad libitum* throughout the experiment. Pedigree information was available for all pigs. Feed intake and body weight were recorded from 120 to 165 day old pigs using a Velos (Nedap co., LTD, Groenlo, the Netherlands) automated individual feeding system. The feed conversion ratio (FCR) was calculated for individuals for the 45 day trial period. The FCR values of all individuals were ordered, and high (20 pigs) and low (20 pigs) end individuals were tested using one way analysis of variance (ANOVA). The analysis found significant differences in FCR (see Supplementary Fig. 1), and two full-sib pairs and two half-sib pairs were selected in such a way that one sibling per pair had a low FCR and the other a high FCR (Table S1), in order to reduce the effects of genetic background as much as possible. Individuals with low



feed efficiency and high FCR were classified as the low group, and those with high feed efficiency and low FCR were classified as the high group.

The selected pigs were euthanised in the morning on which they reached 166 days of age, and mucosal samples were scraped from the caecum and colon using glass microscope slides within 20 min of euthanasia. All samples were collected in sterile tubes and stored in liquid nitrogen until further analyses.

All experimental protocols were approved by the Animal Welfare Committee of China Agricultural University (permit number: DK996). Pigs were slaughtered in accordance with the approved slaughtering guidelines (GB/T 17236-2008) from the Quality Supervision, Inspection, and Quarantine Committee of the People's Republic of China<sup>13</sup>. All efforts were made to minimize animal suffering during the study. Intestinal mucosal tissue of the caecum and colon were isolated aseptically and frozen in liquid nitrogen immediately after slaughter until required for RNA isolation<sup>30</sup>.

**RNA isolation.** Total RNA was extracted from the caecal and colonic mucosa using Trizol (Takara Biotechnology Co. Ltd., Beijing, China) and following the manufacturer's instructions. RNA integrity was checked using 1% agarose gel electrophoresis and the RNA concentration was measured with an Agilent 2100 Bioanalyser (Agilent, Santa Clara, CA).

**Library preparation and RNA sequencing.** A total amount of 1.5 µg RNA per sample was used as input material to prepare RNA samples. Sequencing libraries were generated using NEB Next<sup>®</sup> Ultra<sup>™</sup> RNA Library Prep Kit for Illumina<sup>®</sup> (NEB, USA) following the manufacturer's instructions, and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Samples were fragmented using divalent cations under elevated temperature in NEB Next First-Strand Synthesis Reaction Buffer (5X). First-strand cDNA was synthesised using random hexamer primers and M-MuLV Reverse Transcriptase (RNaseH-). Second-strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEB Next Adaptors with hairpin loop structure were ligated to prepare for hybridization. To select cDNA fragments with the appropriate length, the library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, USA). Subsequently, 3 µL USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. Following this, PCR was performed using Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primers. Finally, products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyser 2100 system.

The libraries were sequenced on an Illumina HiSeq platform, and 150 bp paired-end reads were generated. Sequenced RNA-Seq data for the caecal and colonic mucosa of 16 female Landrace pigs were available from NCBI Sequences Read Archive, under accession numbers SRP065563 and SRP065827.

**Read mapping on the *Sus scrofa* reference genome.** Before reads could be mapped on the reference genome, the raw data had to be quality controlled. Clean reads were obtained using the software FASTQC, which removed reads containing adapter sequences, reads with poly-N sequences (in which the percentage of unknown bases "N" was greater than 10%), and low quality reads (threshold quality score < 20, the percentage of read bases whose error rate was less than 1%). The *S. scrofa* genome was downloaded from the Ensembl database ([ftp://ftp.ensembl.org/pub/release-85/fasta/sus\\_scrofa/dna/](ftp://ftp.ensembl.org/pub/release-85/fasta/sus_scrofa/dna/)) as the reference genome for assembly using Tophat v2.1.1 software<sup>49</sup>. The number of mapping reads of exons, introns, and intergenic positions was calculated using RSeQC (version 2.6.3) in the genome<sup>50</sup>.

**Differential gene expression analysis.** Cufflinks (version 2.2.1) was used for transcriptome assembling, and the Cuffmerge script from Cufflinks was used to merge the transcript files into a single transcriptome annotation. Differential gene expression was then analysed by Cuffdiff<sup>49</sup>. The expression of each gene was represented by counting the number of fragments per kilobase exon per million fragments mapped (FPKM value). Finally,  $|\log_2(\text{fold change})| \geq 1$  and  $q \leq 0.05$  were set as the threshold for differentially expressed gene (DEG) selection<sup>49</sup>.

**DEGs functional annotation clustering.** DEGs were converted to human homologous Ensembl Genes and then utilised for functional enrichment analysis<sup>51</sup>. GO and pathway enrichment analysis of DEGs was implemented using the DAVID Bioinformatics Resources v6.8 (<http://david.abcc.ncifcrf.gov/>)<sup>52,53</sup>. P-values < 0.05 were deemed to show significant enrichment by DEGs. Figures for DEG GO enrichment were drawn online using OmicsShare tools (<http://www.omicsshare.com/forum/>). The protein-protein interaction network was analysed using STRING (<http://www.string-db.org/>)<sup>54</sup>. ToppCluster was used to create the potential important network(s) of pathways and DEGs; a P-value cut-off of 0.05 was selected for ToppCluster analysis<sup>55</sup>.

**DEGs comparison with the animal QTL database.** QTL mapping of DEGs was performed by comparative analysis of DEGs and porcine QTL chromosome positions, which were collected from the animal QTL database (<http://www.animalgenome.org/QTLdb>). Furthermore, the DEGs mapped to metabolism QTLs were refined<sup>56</sup>.

## References

1. Aggrey, S. E., Karnuah, A. B., Sebastian, B. & Anthony, N. B. Genetic properties of feed efficiency parameters in meat-type chickens. *Genet Sel Evol* **42**, 25 (2010).
2. Gilbert, H. *et al.* Correlated responses in sow appetite, residual feed intake, body composition, and reproduction after divergent selection for residual feed intake in the growing pig. *J Anim Sci* **90**, 1046 (2012).
3. Jing, L. *et al.* Transcriptome analysis of mRNA and miRNA in skeletal muscle indicates an important network for differential Residual Feed Intake in pigs. *Sci Rep-UK* **5**, p11953 (2015).

4. Sahana, G., Kadlecova, V., Hornshoj, H., Nielsen, B. & Christensen, O. F. A genome-wide association scan in pig identifies novel regions associated with feed efficiency trait. *J Anim Sci* **91**, 1041 (2013).
5. Chen, Y. *et al.* Global gene expression profiling reveals genes expressed differentially in cattle with high and low residual feed intake. *Anim Genet* **42**, 475 (2011).
6. Lumpkins, B. S., Batal, A. B. & Lee, M. D. Evaluation of the bacterial community and intestinal development of different genetic lines of chickens. *Poult Sci* **89**, 1614 (2010).
7. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027 (2006).
8. Knudsen, K. E. B., Hedemann, M. S. & Laerke, H. N. The role of carbohydrates in intestinal health of pigs. *Anim Feed Sci Tech* **173**, 41 (2012).
9. Sekirov, I., Russell, S. L., Antunes, L. C. M. & Finlay, B. B. Gut Microbiota in Health and Disease. *Physiol Rev* **90**, 859 (2010).
10. Ermert, D. & Blom, A. M. C4b-binding protein: The good, the bad and the deadly. Novel functions of an old friend. *Immunol Lett* **169**, 82 (2016).
11. Castro, M. *et al.* Involvement of neuronal nitric oxide synthase (nNOS) in the regulation of migrating motor complex (MMC) in sheep. *Vet J* **192**, 352 (2012).
12. Wolf, X. A., Serup, P. & Hyttel, P. Three-Dimensional Immunohistochemical Characterization of Lineage Commitment by Localization of T and FOXA2 in Porcine Peri-implantation Embryos. *Dev Dynam* **240**, 890 (2011).
13. Xing, K. *et al.* Identification of genes for controlling swine adipose deposition by integrating transcriptome, whole-genome resequencing, and quantitative trait loci data. *Sci Rep-Uk* **6** (2016).
14. Xing, K. *et al.* The liver transcriptome of two full-sibling Songliao black pigs with extreme differences in backfat thickness. *Journal of Animal Science and Biotechnology* **5**, 6 (2014).
15. Koch, R. M., Gregory, K. E., Chambers, D. & Swiger, L. A. Efficiency Of Feed Use In Beef Cattle. *J Anim Sci* **22**, 486 (1963).
16. Do, D. N., Strathe, A. B., Jensen, J., Mark, T. & Kadarmideen, H. N. Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. *J Anim Sci* **91**, 4069 (2013).
17. Le Naou, T., Le Floc'H, N., Louveau, I., Gilbert, H. & Gondret, F. Metabolic changes and tissue responses to selection on residual feed intake in growing pigs. *J Anim Sci* **90**, 4771 (2012).
18. Zhao, Y. *et al.* Transcriptome Analysis Reveals that Vitamin A Metabolism in the Liver Affects Feed Efficiency in Pigs. *G3 (Bethesda, Md.)* (2016).
19. Vigos, S., O'Doherty, J. V., Kelly, A. K., O'Shea, C. J. & Sweeney, T. The Effect of Divergence in Feed Efficiency on the Intestinal Microbiota and the Intestinal Immune Response in Both Unchallenged and Lipopolysaccharide Challenged Ileal and Colonic Explants. *Plos One* **11** (2016).
20. Mani, V. *et al.* Intestinal integrity, endotoxin transport and detoxification in pigs divergently selected for residual feed intake. *J Anim Sci* **91**, 2141 (2013).
21. Isaacson, R. & Kim, H. B. The intestinal microbiome of the pig. *Animal Health Research Reviews* **13**, 100 (2012).
22. Rios-Covian, D. *et al.* Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. *Front Microbiol* **7** (2016).
23. Stumpff, F. *et al.* Effects of dietary fibre and protein on urea transport across the cecal mucosa of piglets. *J Comp Physiol B* **183**, 1053 (2013).
24. Tan, Z. *et al.* Metagenomic Analysis of Cecal Microbiome Identified Microbiota and Functional Capacities Associated with Feed Efficiency in Landrace Finishing Pigs. *Front Microbiol* **8** (2017).
25. Pedersen, R. *et al.* Characterisation of Gut Microbiota in Ossabaw and Gottingen Minipigs as Models of Obesity and Metabolic Syndrome. *PLOS ONE* **8** (2013).
26. Kim, H. B. & Isaacson, R. E. The pig gut microbial diversity: Understanding the pig gut microbial ecology through the next generation high throughput sequencing. *Vet Microbiol* **177**, 242 (2015).
27. Metzler-Zebeli, B. U. *et al.* Grain-rich diets differently alter ruminal and colonic abundance of microbial populations and lipopolysaccharide in goats. *Anaerobe* **20**, 65 (2013).
28. Nicholson, J. K. *et al.* Host-Gut Microbiota Metabolic Interactions. *Science* **336**, 1262 (2012).
29. Bird, A. R., Conlon, M. A., Christophersen, C. T. & Topping, D. L. Resistant starch, large bowel fermentation and a broader perspective of prebiotics and probiotics. *Benef Microbes* **1**, 423 (2010).
30. Looft, T. *et al.* Bacteria, phages and pigs: the effects of in-feed antibiotics on the microbiome at different gut locations. *Isme J* **8**, 1566 (2014).
31. Castro, J. *et al.* Linaclotide Inhibits Colonic Nociceptors and Relieves Abdominal Pain via Guanylate Cyclase-C and Extracellular Cyclic Guanosine 3',5'-Monophosphate. *Gastroenterology* **145**, 1334 (2013).
32. Chey, W. D. *et al.* Linaclotide for Irritable Bowel Syndrome With Constipation: A 26-Week, Randomized, Double-blind, Placebo-Controlled Trial to Evaluate Efficacy and Safety. *Am J Gastroenterol* **107**, 1702 (2012).
33. Han, X. *et al.* Loss of Guanylyl Cyclase C (GCC) Signaling Leads to Dysfunctional Intestinal Barrier. *Plos One* **6** (2011).
34. Brenna, O. *et al.* Cellular localization of guanylin and uroguanylin mRNAs in human and rat duodenal and colonic mucosa. *Cell Tissue Res* **365**, 331 (2016).
35. Brenna, O. *et al.* The guanylate cyclase-C signaling pathway is down-regulated in inflammatory bowel disease. *Scand J Gastroenterol* **50**, 1241 (2015).
36. Begg, D. P. *et al.* Effect of Guanylate Cyclase-C Activity on Energy and Glucose Homeostasis. *Diabetes* **63**, 3798 (2014).
37. Kim, G. W. *et al.* Calorie-induced ER stress suppresses uroguanylin satiety signaling in diet-induced obesity. *Nutr Diabetes* **6** (2016).
38. Hooper, P. L. & Hooper, P. L. Inflammation, heat shock proteins, and type 2 diabetes. *Cell Stress Chaperon* **14**, 113 (2009).
39. Mardan-Nik, M. *et al.* Association of heat shock protein70-2 (HSP70-2) gene polymorphism with obesity. *Ann Hum Biol* **43**, 542 (2016).
40. RINDERKNECHT, H. Activation Of Pancreatic Zymogens - Normal Activation, Premature Intrapaneatic Activation, Protective Mechanisms Against Inappropriate Activation. *Digest Dis Sci* **31**, 314 (1986).
41. Mehner, C. *et al.* Serine protease inhibitor Kazal type 1 (SPINK1) drives proliferation and anoikis resistance in a subset of ovarian cancers. *Oncotarget* **6**, 35737 (2015).
42. Mo, S. *et al.* Substrate Specificity, Regulation, and Polymorphism of Human Cytochrome P450 2B6. *Curr Drug Metab* **10**, 730 (2009).
43. Ji, L. *et al.* Psoralen, a mechanism-based inactivator of CYP2B6. *Chem-Biol Interact* **240**, 346 (2015).
44. Lewis, D. 57 varieties: the human cytochromes P450. *Pharmacogenomics* **5**, 305 (2004).
45. Bhutia, Y. D. *et al.* SLC transporters as a novel class of tumour suppressors: identity, function and molecular mechanisms. *Biochem J* **473**, 1113 (2016).
46. Suh, A. M., Webb, A., Papp, A. C., Geier, E. G. & Sadee, W. Expression and splicing of ABC and SLC transporters in the human blood-brain barrier measured with RNAseq. *European journal of pharmaceutical sciences: official journal of the European Federation for Pharmaceutical Sciences* (2017).
47. Zhang, X. *et al.* Transcriptomic Analysis of Ovaries from Pigs with High And Low Litter Size. *Plos One* **10** (2015).
48. Vimalaswaran, K. S. *et al.* Lack of Association Between PCK1 Polymorphisms and Obesity, Physical Activity, and Fitness in European Youth Heart Study (EYHS). *Obesity* **18**, 1975 (2010).
49. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562 (2012).
50. Wang, L. *et al.* Measure transcript integrity using RNA-seq data. *Bmc Bioinformatics* **17** (2016).

51. Wang, T. *et al.* Deep Sequencing of the Transcriptome Reveals Inflammatory Features of Porcine Visceral Adipose Tissue. *Int J Biol Sci* **9**, 550 (2013).
52. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25 (2000).
53. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44 (2009).
54. Franceschini, A. *et al.* STRINGv9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* **41**, D808 (2013).
55. Kaimal, V., Bardes, E. E., Tabar, S. C., Jegga, A. G. & Aronow, B. J. ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. *Nucleic Acids Res* **38**, W96 (2010).
56. Hu, Z., Park, C. A., Wu, X. & Reecy, J. M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res* **41**, D871 (2013).

## Acknowledgements

We acknowledge the reviewers for their constructive suggestions for our manuscript. We are also indebted to Weiliang Zhou from Tianjin Ninghe primary pig breeding farm for providing the pigs for these experiments. This research was financially supported by the Beijing Innovation Consortium of Agriculture Research System (BAIC02-2016).

## Author Contributions

J.L., C.W., and Z.T. planned the project and designed the experiments. Z.T. conducted the experiments and carried out the data analysis with help from T.Y., K.X., and Y.W. Z.T., Y.W., K.X., F.Z., X.Z., H.A., and S.C. contributed reagents preparation and samples collection. Z.T. wrote the manuscript, which was critically reviewed by J.L. and C.W.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-14568-6>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017