Characterization of gene promoters in pig: conservative elements, regulatory motifs and evolutionary trend

Kai Wei[1,2], Lei Ma[1] and Tingting Zhang[1]

[1] College of Life Science, Shihezi University, Shihezi, Xinjiang, China
[2] Center of Life and Food Sciences Weihenstephan, Technische Universität München, Freising, Byern, Germany

ABSTRACT

It is vital to understand the conservation and evolution of gene promoter sequences in order to understand environmental adaptation. The level of promoter conservation varies greatly between housekeeping (HK) and tissue-specific (TS) genes, denoting differences in the strength of the evolutionary constraints. Here, we analyzed promoter conservation and evolution to exploit differential regulation between HK and TS genes. The analysis of conserved elements showed CpG islands, short tandem repeats and G-quadruplex sequences are highly enriched in HK promoters relative to TS promoters. In addition, the type and density of regulatory motifs in TS promoters are much higher than HK promoters, indicating that TS genes show more complex regulatory patterns than HK genes. Moreover, the evolutionary dynamics of promoters showed similar evolutionary trend to coding sequences. HK promoters suffer more stringent selective pressure in the long-term evolutionary process. HK genes tend to show increased upstream sequence conservation due to stringent selection pressures acting on the promoter regions. The specificity of TS gene expression may be due to complex regulatory motifs acting in different tissues or conditions. The results from this study can be used to deepen our understanding of adaptive evolution.

Submitted 18 February 2019
Accepted 29 May 2019
Published 25 June 2019

Corresponding authors
Lei Ma, mlei@shzu.edu.cn
Tingting Zhang, zting@shzu.edu.cn

Academic editor
Guoliang Li

Additional Information and Declarations can be found on page 13

DOI **10.7717/peerj.7204**

Subjects Bioinformatics, Evolutionary Studies, Genetics, Genomics
Keywords Sequence conservation, Regulatory motif, Housekeeping promoter, Tissue-specific promoter, Evolutionary dynamics

INTRODUCTION

Housekeeping (HK) genes are consistently expressed in different tissues and conditions to maintain basic life activities (*Butte, Dzau & Glueck, 2001*; *Zhu et al., 2008*). They may be the minimum collection of genes for normal cellular physiological processes (*Kouadjo et al., 2007*). Tissue-specific (TS) genes are, in contrast to HK genes, are expressed in specific tissues or conditions and show fluctuant expression levels in different tissues, developmental stages or environments (*Kouadjo et al., 2007*; *Thorrez et al., 2011*). Some previous studies reported that significant difference in gene structure, function and evolution between HK and TS genes. For example, HK genes evolve on average more slowly than TS genes (*Zhang & Li, 2004*), the entropy of TS genes is significantly less than HK genes (*Thomas et al., 2015*) and the introns, untranslated regions (UTRs) and coding

sequences (CDS) of the HK genes are shorter, indicating a selection for compactness in these genes (*Eisenberg & Levanon, 2003*; *Zhu et al., 2008*).

The correct performance of function is mainly dependent on complex gene expression regulation which ensure that different genes were expressed in specific tissues, developmental stages and different conditions (*Wray et al., 2003*). Promoters are the regulatory center in this process, due to a large number of cis-regulatory elements located upstream of a transcription start site (TSS) (*Halees, 2003*). The key elements related with conservation and gene expression regulation in promoters include short tandem repeat (STR), G-quadruplex sequence (G4), also known as potential quadruplex-forming sequences (PQS) and CpG island, transcription factor binding site, which are often interacted and integrated into combined regulatory motifs to regulate some critical physiological functions (*Abe & Gemmell, 2014*; *Wittkopp & Kalay, 2012*; *Gemayel et al., 2010*). Some studies indicated divergence between promoters of HK and TS genes in structure, conversation and regulation in human and mouse. For example, regulatory motifs of HK and TS promoters showed differently positional bias and conservation in mouse (*Bellora, Farré & Albà, 2007*; *Farré et al., 2007*).

In previous studies, empirical results have indicated that nucleotide substitution in regulatory motifs could be one of the causes of phenotypic differentiation (*Horton et al., 2014*; *Andersson, 2009*; *Xu et al., 2014*). The comparisons of upstream promoter sequence across different species have suggested significantly different evolutionary constraints exhibited by promoters of HK and TS genes. In addition, promoters of genes encoding trans-acting factors, such as transcription factors and/or developmental regulatory factors, tend to exhibit especially strong upstream promoter sequence conservation (*Lee, Kohane & Kasif, 2005*; *Iwama & Gojobori, 2004*), indicating that the mutations of cis-regulatory elements may change gene expression in different tissues or conditions. Therefore, the evidence of conservation and selection in promoters of different types of gene can contribute to identify HK and TS genes (*She et al., 2009*). In addition, evolutionary dynamics analysis of promoters can contribute to understanding regulatory patterns and evolutionary trends of HK and TS genes (*De Jonge et al., 2007*).

The pig (*Sus scrofa*) is an important meat resource and biomedical model. Surveying pig conservation and regulatory patterns in promoters may help pave the way for a greater understanding of the regulatory divergence and evolutionary dynamics in pig HK and TS promoters. Here, we analyzed differences in the conversation of promoters and expression patterns exhibited by HK and TS genes. And the evolutionary dynamics of HK and TS promoters were compared to further understand the reasons for the differences in regulatory patterns. Thus, it is of interest to investigate how evolutionary selection acts on promoters to cause divergent regulation of HK and TS genes.

MATERIALS AND METHODS

Data preparation and definition of HK and TS genes

Gene datasets were defined from pig transcriptome data from 14 RNA-seq projects which includes 21 tissues (heart, spleen, liver, kidney, lung, musculus longissimus dorsi, occipital

Table 1 The structural comparison between HK and TS genes.

Structure	HK gene	TS gene	*P*-value[c]
Total intron length[a]	28,108 ± 173[b]	67,167 ± 691	3.50*E*-182
5′ UTR length	156 ± 3	132 ± 4	2.70*E*-56
3′ UTR length	658 ± 13	499 ± 18	1.30*E*-37
Average exon length per gene	261 ± 3	206 ± 2.63	1.60*E*-19
CDS length	2,181 ± 10	1,475 ± 44	8.40*E*-134
Number of exons	9.2 ± 0.1	15.2 ± 0.68	7.30*E*-61
Transcript length	3,312 ± 13	1,817 ± 40	2.10*E*-79

Notes:
[a] The length was measured in nucleotides.
[b] The value gives the average and standard error of mean.
[c] The *P*-value was calculated based on the Mann–Whitney test. UTR, untranslated region; CDS, coding sequence.

cortex, hypothalamus, frontal cortex, cerebellum, endometrium, mesenterium, greater omentum, backfat, gonad, ovary, placenta, testis, blood, uterine and lymph nodes) and a total of 131 samples (Table S1). The SRA files of transcriptome data were downloaded from the SRA database of NCBI and then converted to fastq files using fastq-dump in SRA Toolkit (*Kodama et al., 2012*). Reads of average quality score above 20 were extracted by IlluQC.pl (*Patel & Jain, 2012*). The filtered reads were mapped to pig reference genome (Sus Sscrofa10.2) using Tophat 2.0.14 (*Trapnell, Pachter & Salzberg, 2009*). The mapped reads were then submitted to an assembler Cufflinks 2.2.1 to assemble into transcripts and estimate their abundances (*Trapnell et al., 2010*). The Fragments per Kilobase of exon per Million fragments mapped (FPKM) were calculated to estimate expression level of transcripts.

A total of 3,136 HK genes were defined according to strict criteria (File S1): (i) the transcripts must be detected in all 21 tissues; (ii) the expression variance across tissues were tested by Kolmogorov–Smirnov uniform test, $P > 0.1$ was chosen as the cutoff to extract candidate transcripts; (iii) no abnormal expression in any single tissue; that is, the expression values were restricted within the fourfold range of the average across tissues; and (iv) all transcripts from same candidate gene must met the above criteria. In addition, transcripts with expression restricted to one to three tissues were classified as TS genes, including 1,316 TS genes (File S1). In order to compare the conservative elements and regulatory motifs between HK and TS genes, the two kb upstream sequences of genes were obtained as promoters from Ensemble BioMart (*Chen et al., 2010*; *Kinsella et al., 2011*).

Structure analysis

The structure data of genes, including intron length, 5′ and 3′ UTR length, exon length, CDS length and Transcript length, were obtained from the Ensembl BioMart (*Kinsella et al., 2011*). The length of various parts between HK and TS genes were compared by Mann–Whitney test (Table 1).

Gene ontology analysis

The functional enrichment of HK and TS genes was performed using DAVID, ver. 6.8 (*Huang, Sherman & Lempicki, 2009a*, *2009b*). All expressed genes in the data were used as background to control accuracy of results. The false discovery rates (FDR) values were

calculated to estimate the level of overrepresentation of the selected genes in gene ontology (GO) categories (*Storey, 2002*). FDR less than 0.01 were used as the cut-off value to acquire significant GO terms.

Identification of conservative elements

To understand distribution of GC in promoters, we identified CpG islands using the Newcpgreport software (*Labarga et al., 2007*). The default parameters were chosen to identify CpG islands: (i) the GC content in a 100 bp window exceeded 50%, (ii) the length of CpG island exceeded 200 bp, and (iii) the ratio of observed to expected (O/E) number of CpG islands were must bigger than 0.6 (*Gardiner-Garden & Frommer, 1987*).

Short tandem repeats were detected in HK and TS promoter sequences using the Phobos 3.3.12 software (*Mayer, Leese & Tollrian, 2010*). We identified STRs according to following criteria: (i) the STRs identified were must perfect repeats, (ii) repeats units were 2, 3, 4, 5 and 6, (iii) STRs were selected with number of repeat units exceeded six and (iv) the overlapped STRs were counted separately. The mononucleotide repeats were not considered due to repeat number could not be identified.

The Quadruplex forming G-Rich Sequences Mapper was used to detect PQSs in promoters (*Kikin, D'Antonio & Bagga, 2006*). The search parameters were set as follows: (i) maximum length of PQSs cannot exceed 30, (ii) the minimum number of units in a PQS was four and (iii) the minimum loop size was set as zero. Note that these settings cause some elements to be counted twice in both STRs and PQSs.

Regulatory motifs discovery by the MEME suite

The protein binding sites and interaction domains are very important features for the regulation of gene expression. The regulatory motifs were found using MEME Suite (*Bailey et al., 2009*). The following options of input parameters were used: (i) 100 bp bin windows were set to search motifs, (ii) zero or one occurrence per sequence model was chosen to improve the sensitivity and quality of the motif search, (iii) the maximum and minimum width of the motifs were 15 and 6, respectively, (iv) the given promoter sequences or on its reverse complement sequences were searched, (v) the number of motifs was set to five and (vi) 0-order model of sequences was used as the background model (*Abe & Gemmell, 2014*).

The JASPAR database was used to search biological functions of motifs (*Khan et al., 2018*).

Evolutionary features analysis

The evolutionary dynamics of HK and TS CDSs were compared by calculating the substitution ratio. The non-synonymous substitution rate (dN) and synonymous substitution rate (dS) were estimated using the Nei–Gojobori method embedded in MEGA 7.0 (Z-test, $P < 0.05$) (*Kumar, Stecher & Tamura, 2016*; *Wei, Zhang & Ma, 2018*). The CDSs of HK and TS genes were downloaded from Ensembl BioMart. The orthologous sequences of mouse (*Mus musculus*) were used as outgroups to perform multiple sequence alignments. The following criteria were used: (i) the Overall Average option was chosen, (ii) pairwise deletion was selected to treat Gaps/Missing data. In addition, the

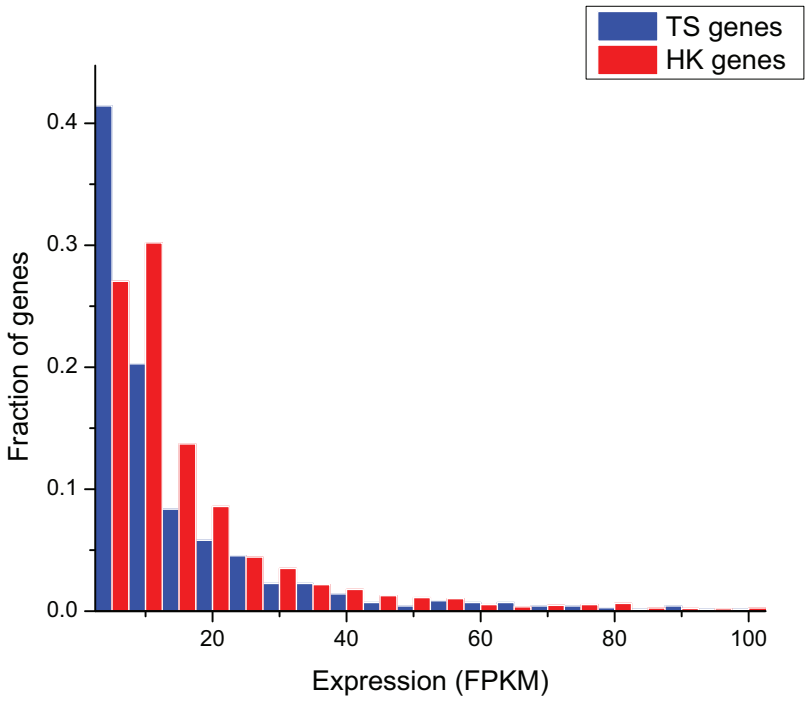

Figure 1 **The comparison of expression level between HK and TS genes.**
Full-size DOI: 10.7717/peerj.7204/fig-1

orthologous sequences were downloaded from Ensembl BioMart (*Kinsella et al., 2011*). The dN/dS ratios were calculated to estimate the selective pressure (*Hurst, 2002*; *Dasmeh et al., 2014*). In addition, the nucleotide substitution rate of promoters were calculated to estimate conservation of promoters.

Statistical analyses involved in present study were performed in R (www.r-project.org).

RESULTS

Identification of HK and TS genes

In our previous study, 3,136 genes were defined as HK genes, which maintain relatively stable expression level in all 21 tissues (File S1; *Wei, Zhang & Ma, 2018*). The 1,316 genes defined as TS genes contained 2,214 transcripts expressing in one to three tissues (File S1).

The comparison of gene expression in ERP002055 sequencing project indicates that the average expression level of HK genes (FPKM = 17.10 ± 3.63) was significantly higher than TS genes (FPKM = 6.43 ± 64.08) (Mann–Whitney test, $P < 0.01$) (Fig. 1).

The structural and functional comparison of HK and TS genes

There are significant differences between HK and TS in gene structural length (Mann–Whitney test, $P < 0.01$, Table 1). The total length and intron length of TS genes are significantly longer than HK genes, but other structures are significantly shorter than HK genes, such as UTR and CDS. These results indicated that the structure of HK genes is more compact than TS genes. Combined with expression level analysis, the high

expression characteristics of HK genes may require a flexible gene structure, that is, a more compact gene structure enables it to initiate expression quickly, and it takes less time and energy in the expression process.

In addition, TS genes displayed a higher number of exons and transcripts compared with HK genes (Mann–Whitney test, $P < 0.01$), which may be related to the spatiotemporal dependence of TS genes that express different splicing isoforms at different developmental stages of the cell or in different environmental conditions.

The GO enrichment analysis of biological processes revealed that the functions of HK genes are mainly concentrated on the basal metabolism of cells, such as energy metabolism, cellular transport and synthesis and decomposition of macromolecules (Table S2). The principal functions of TS genes are related to tissue specificity, such as many genes enriched to tissue differentiation and development, and many genes are associated with cellular immune response (Table S3). The results showed that HK genes and TS genes have their own specific functional characteristics, and their roles in cells are significantly different. TS genes are genes that distinguish between tissues. HK genes mainly provides the necessary substances and energy in the cells to perform basic life activities. HK and TS genes gradually form unique functional characteristics in the long-term evolutionary process, and their mutual cooperation is the basis for the orderly operation of cell life activities.

GC content and CpG island density in HK and TS promoters

Promoter sequences of pig HK and TS genes increased gradually as it approached the TSS in their GC contents (Fig. 2A), ranging from 0.30 to 0.75, and their averages were 0.46 and 0.45, respectively. GC contents in HK promoters were significantly higher than TS promoters as it approached the TSS (Mann–Whitney test, $P < 0.01$) (Fig. 2B).

The 1,556 CpG islands were identified in HK promoters with a density of 0.47 per promoter. TS promoters contained 393 CpG islands with a density of 0.30. Figure 2C shows that the density of CpG islands in HK promoters is higher than TS promoters (Mann–Whitney test, $P < 0.01$). In addition, the analysis showed that the length of CpG islands in HK promoters is longer than TS promoters (Mann–Whitney test, $P < 0.01$) (Fig. 2D). The higher GC and CpG island content may indicate HK promoters are more stable than the TS promoters. HK genes with high density CpG islands have higher transcriptional activity across tissues, that is, it has a higher level of expression, while TS genes may be restricted by strict expression in specific tissues (*Fenouil et al., 2012*; *Vavouri & Lehner, 2012*).

Abundance of STR and PQS in HK and TS promoters

Table 2 summarized the frequencies of STR motifs in HK and TS promoters. The similar STR motifs were detected in HK and TS promoters. However, STR motifs density in HK promoters was significantly higher than TS promoters (Table 2, Mann–Whitney test, $P < 0.01$). Figure 3A indicated STR density of HK promoters significantly higher than TS promoters (Mann–Whitney test, $P < 0.01$). In addition, the distribution of PQS between HK and TS promoters were no significant difference. But PQS content in the proximal part of promoter was higher than the distal part of the promoter (Fig. 3B).

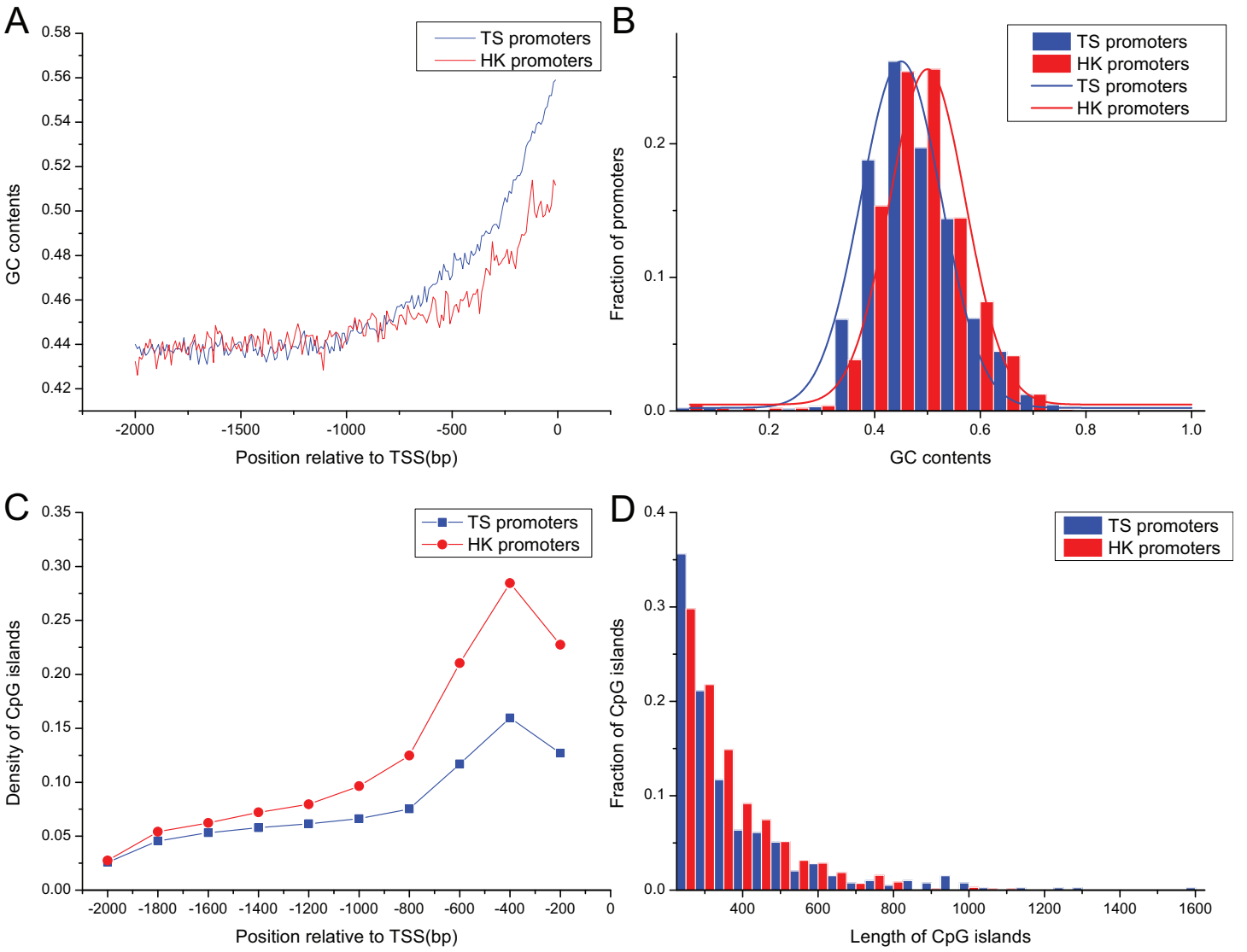

Figure 2 The distribution of content and length of GC and CpG islands between HK and TS promoters. (A) The tendency of GC contents in the promoters (the mean and standard error are 0.46 ± 0.0015 and 0.45 ± 0.0024 in HK and TS promoters, respectively), (B) the distribution of GC contents, (C) the distribution of CpG islands (the averaged density in HK and TS promoters are 0.47 ± 0.0013 and 0.30 ± 0.0016, respectively), (D) the distribution of length of CpG islands (the averaged length in HK and TS promoters are 352 ± 9.82 and 234 ± 3.36, respectively).

Full-size DOI: 10.7717/peerj.7204/fig-2

Regulatory motifs identified in the HK and TS promoters

Motif density and types of TS promoters were significantly higher than HK promoters (Mann–Whitney test, $P < 0.01$). A total of 38 types of regulatory motifs were identified in HK promoters, a total of 74,322, with a density of 23 motifs per promoter (Table 3; Table S4). There were 115 types of regulatory motifs in the TS promoters, a total of 67,123, with a density of 51 motifs per promoter (Table 4; Table S5). These results are consistent with variable expression levels and patterns of TS genes in different tissues and conditions. In HK and TS promoters, some motifs are zinc finger factors, especially in

Table 2 The comparison of STR between HK and TS promoters.

STR	TS promoters		HK promoters	
	Number of STR	Frequency of STR	Number of STR	Frequency of STR
AC	76	0.058	393	0.13
AG	30	0.023	160	0.051
AT	34	0.026	145	0.046
CG	5	0.0035	10	0.0030
AAC	20	0.015	74	0.024
AAG	1	0.00064	14	0.0046
AAT	4	0.0029	24	0.0076
ACC	0	0	7	0.0023
AGG	3	0.0026	5	0.0015
AGC	2	0.0013	10	0.0030
CCG	12	0.0089	12	0.0038
ACAG	0	0.00032	5	0.0015
AAGG	0	0.00032	19	0.0061
AATC	1	0.00064	0	0
AAAC	3	0.0022	14	0.0046
AAAG	2	0.0016	31	0.0099
AAAT	8	0.0061	24	0.0076
AGAT	1	0.00096	2	0.00076
AGGG	1	0.00064	5	0.0015
ATCC	0	0	7	0.0023
AAAAG	1	0.00032	5	0.0015
AAAAT	1	0.00032	5	0.0015
STR/seq[a]		0.15		0.31

Note:
[a] STR/seq is the number of STR motif counted per promoter sequence.

HK promoters. The functions of HK motifs are partially similar with TS motifs, for examples some C2H2 zinc finger factors but different motifs are chosen to bind the same transcription factor.

In addition, there are 22 and 99 specific regulatory motifs in HK and TS promoters, respectively. But only 16 types of regulatory motifs were shared between them. These results indicated a large number of specific regulatory motifs in TS promoters which may help TS genes to adapt to different conditions.

Divergence of HK and TS promoter sequences

The promoters of genes show sequence divergence (*Lee, Kohane & Kasif, 2005*; *Iwama & Gojobori, 2004*; *Suzuki et al., 2004*). The level of promoter sequence divergence is positively correlated with the evolutionary rate of the encoded protein (*Castillo-Davis, Hartl & Achaz, 2004*; *Chin, Chuang & Li, 2005*). To investigate evolutionary dynamic of HK and TS promoters, the number of non-synonymous substitutions per non-synonymous site (dN), the number of synonymous substitutions per synonymous site (dS) and dN/dS ratio

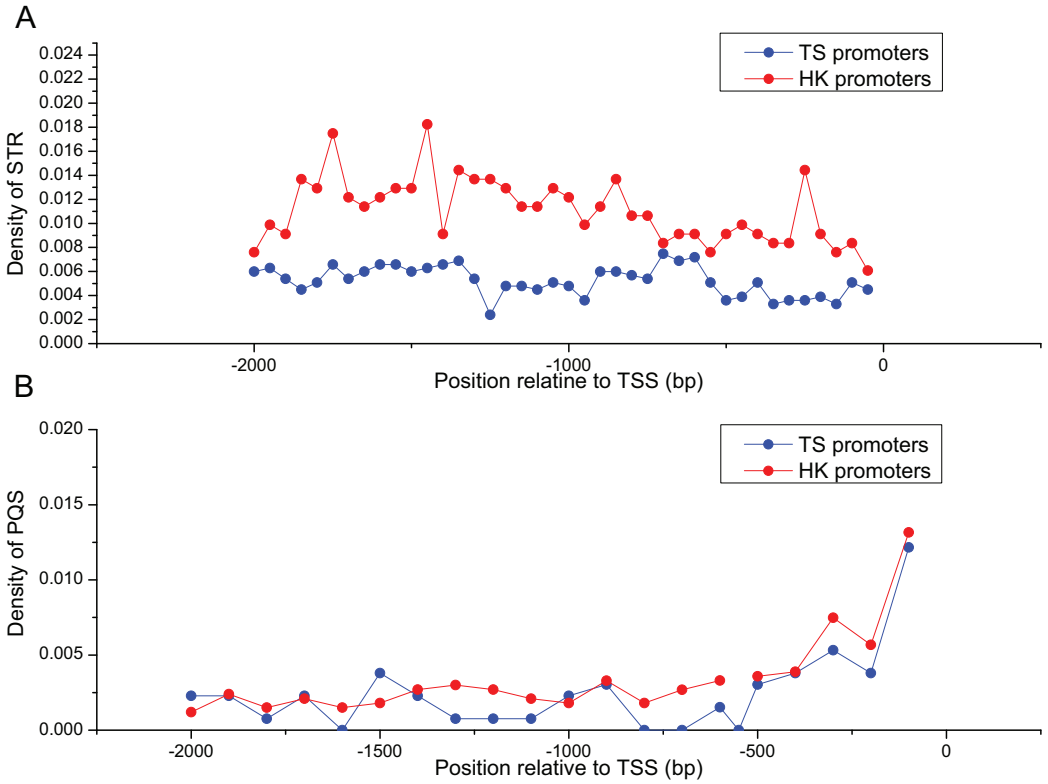

Figure 3 The distribution of STR and PQS between HK and TS promoters. (A) The distribution of STR density in the promoters, (B) the distribution of PQS density in the promoters (the mean and standard error are 0.0028 ± 0.00061 and 0.0024 ± 0.00058). Full-size DOI: 10.7717/peerj.7204/fig-3

were calculated for HK and TS CDS using mouse (*Mus musculus*) as an outgroup. And the promoter nucleotide substitution rate (dP) was also estimated to understand the evolutionary trend of promoters in pig (Files S2 and S3).

Evolutionary dynamic analysis showed that the vast majority dN and dN/dS of CDS, were less than one, showing a power-law distribution, indicating that most of the CDS were under the purifying selection pressure and in negative selection (Figs. 4A and 4C; Table S6). The dS showed an approximately normal distribution and was significantly greater than dN (Mann–Whitney test, $P < 0.01$). About 20% of CDS had dS greater than one (Fig. 4B). In addition, dP of TS promoters (0.64) was significantly higher relative to that of HK promoters (Fig. 4D; Table S6), which indicated HK promoters with increased conservation and suffered more stringent selection pressure than TS promoters.

Interestingly, the nucleotide substitution rate of promoters showed significantly positive correlation with the CDS (for HK genes, dP and dN, $r = 0.23$, $P < 10^{-32}$; dP and dN/dS, $r = 0.16$, $P < 10^{-38}$; dP and dS, $r = 0.38$, $P < 10^{-37}$; and for TS genes dP and dN, $r = 0.27$, $P < 10^{-36}$; dP and dN/dS, $r = 0.23$, $P < 10^{-32}$; dP and dS, $r = 0.44$, $P < 10^{-41}$). Therefore, promoters showed a similar tendency with the CDS.

The nucleotide substitution rate of HK promoters was significantly smaller than that of TS promoters. The structure of HK promoters became more stable and evolved slower than TS promoters, which were determined by the importance of HK genes in cells

Table 3 The top 10 of regulatory motifs in HK promoters.

Motif	Length	Number of motifs	*E*-value	Description
GCYRCAGC	8	3,637	3.8*E*-403	C2H2 zinc finger factors
GCCHGGGA	8	2,991	1.8*E*-353	Rel homology region (RHR) factors
GCTGTRGC	8	2,256	1.4*E*-313	C2H2 zinc finger factors
TCCCWGGC	8	2,521	1.1*E*-298	Rel homology region (RHR) factors
TCSTTAAC	8	1,896	1.7*E*-263	Tryptophan cluster factors
GGAACTYC	8	1,882	1.4*E*-242	Rel homology region (RHR) factors
AAAAWAAA	8	4,404	3.4*E*-229	C2H2 zinc finger factors
GTGGTGTA	8	1,282	9*E*-228	C4 zinc finger factors
TACACCAC	8	1,310	8.4*E*-226	C4 zinc finger factors
CATATGS	7	2,634	4.7*E*-224	Basic helix-loop-helix factors

Note:
The top10 regulatory motifs in HK promoters were listed in table. N or X: A G C T; V: A C T; H: A C T; D: A G T; B: C G T; M: A C; R: A G; W: A T; S: C G; Y: C T; K: G T.

Table 4 The top 10 of regulatory motifs in TS promoters.

Motifs	Length	Number of motifs	*E*-value	Description
GCYACAGC	8	806	2.40*E*-112	C2H2 zinc finger factors
GCCHGGGA	8	948	3.50*E*-99	Fork head/winged helix factors
GCTGTRGC	8	703	1.10*E*-97	C2H2 zinc finger factors
AAAAWAAA	8	1,668	1.80*E*-92	C2H2 zinc finger factors
TATWTAT	7	1,055	8.80*E*-85	MADS box factors
TCSTTAAC	8	584	7.90*E*-77	Tryptophan cluster factors
TACACCAC	8	413	6.60*E*-71	C4 zinc finger factors
TTTTTYTT	8	1,630	9.80*E*-74	C2H2 zinc finger factors
CATATGS	7	896	1.60*E*-67	Basic helix-loop-helix factors
CCACTGAG	8	517	1.90*E*-63	Nuclear receptors with C4 zinc

Note:
The top10 regulatory motifs in TS promoters were listed in table. N or X: A G C T; V: A C T; H: A C T; D: A G T; B: C G T; M: A C; R: A G; W: A T; S: C G; Y: C T; K: G T.

(*Nei & Kumar, 2000*). The evolution of TS promoters is significantly faster than that of HK promoters, indicating weaker selection pressures can help specific tissues adapt to different environmental conditions (*Zhang & Li, 2004*).

DISCUSSION

The present study characterized conservative motifs and regulatory elements of gene promoters in pig. In addition, combined with the analysis of evolutionary dynamics, we investigated the difference of HK and TS genes in regulation of gene expression.

In the long-term evolution and environmental adaptation process, HK and TS genes gradually form specific genomic structure, respectively. HK genes showed more compact structures than TS genes. This may be due to different properties of gene expression (*Chang et al., 2011*; *Hong et al., 2013*). TS genes showed shorter transcript length, but a higher number of transcripts and exons indicated that more alternative splicing occurs in

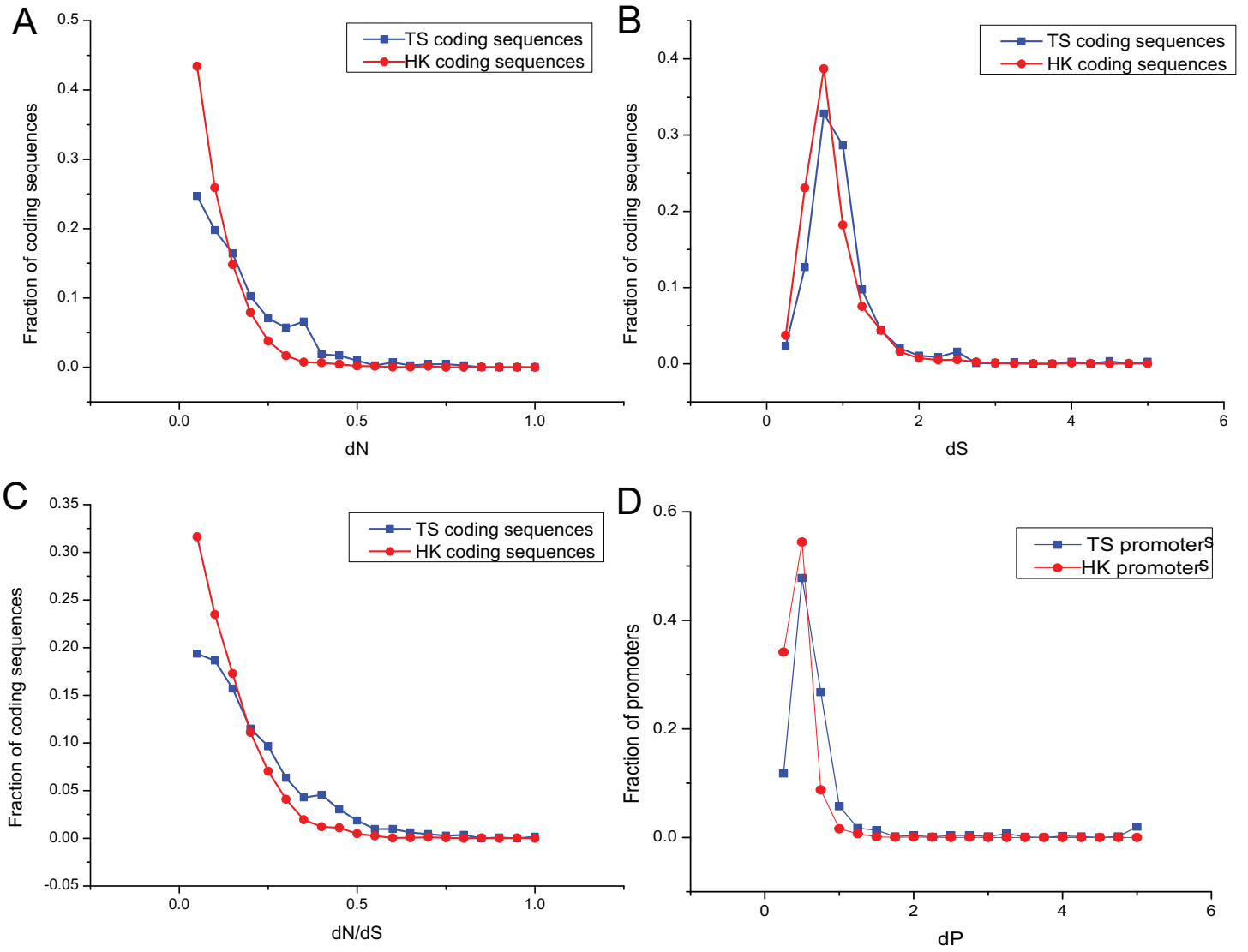

Figure 4 The overlap of regulatory motifs between HK and TS promoters. Evolutionary dynamics of promoters and coding sequence. (A–C) The distribution of dN, dS and dN/dS in coding sequences, (D) the distribution of dP in promoters. Full-size DOI: 10.7717/peerj.7204/fig-4

expression to adapt to different environments. This may contribute to the expression of HK genes activated at any time to maintain the basic life activities of the cells (*Eisenberg & Levanon, 2003*, *2013*). For example, genes associated with ribosome complex are continuously expressed in the cells to meet the requirements of the body protein (*Brandman et al., 2012*). However, TS genes only express at specific developmental stages of a particular tissue, and their ultimate goal is to adapt to temporal and spatial development of tissues (*Holder & Klein, 1999*; *Lawson & Zhang, 2008*). For example, EPHB3 (EPH receptor B3) gene is expressed in the nervous system, which is mainly involved in the development of neurons (*Holder & Klein, 1999*).

In the process of evolution, the HK promoters are under strict purifying selection pressure, and the gene expression level tends to be stable in different tissues and environments to

maintain life, while constrained forces of TS promoters in evolution is much smaller than HK promoters. In addition, nucleotide substitution rate of TS promoters is significantly higher than HK promoters. The adaptability is mainly reflected in the phenotypic changes, so the adaptability of the organism is mainly reflected in the selective expression of TS genes under different environmental conditions (*Hill et al., 1998*). This also explains the reason that the higher nucleotide substitution of TS promoters. The evolution of TS promoters and selective expression are the embodiment of environmental adaptability, while the evolution of HK promoters and the stability of expression aim to maintain the basic cellular function and in different tissues and conditions (*Urrutia & Hurst, 2001*).

The regulatory elements on promoter are important factors which can contribute to species adaptation to changing environments. The HK promoters of pig shows higher sequence conservation than TS promoters, mainly due to the strict purifying selection pressure act on HK promoters to maintain the stability of HK gene expression in different environments. The expression of TS genes is selective, and it is selectively expressed and fluctuating under different conditions, which requires the promoter to initiate different regulatory pathways under different conditions. So the expression of genes can be regulated at any time to adapt to the current environment (*Larsen et al., 2013*; *Urrutia & Hurst, 2001*).

The conserved sequences (STR, PQS and CpG island) in the HK promoters are higher than TS promoters. Genes driven and regulated by repeat sequence promoters are indicated to show significantly higher rates of transcription than those without repeat elements as reported by experiments showing that knockout of STR elements in promoters show significant differences in gene expression compared with promoters without having knocked out STR (*Vinces et al., 2009*; *Valipour et al., 2013*). Promoters with CpG islands show high transcriptional activity in multiple tissues (*Elango & Yi, 2011*; *Sharif et al., 2010*). The relationship between gene ontologies and CpG islands length suggest the important role of CpG islands in chromatin structures by methylation (*Robertson, 2002*). The regulation of HK genes is relatively simple compared to TS gene regulation because HK genes are continuously expressed under any conditions (*Bao, Li & Zhao, 2012*; *Bellora, Farré & Albà, 2007*). TS genes are differentially expressed at different developmental stages and conditions, and are effector genes that adapt to different environments. They have different isoforms and expression levels under different conditions and need a large number of different regulatory motifs to bind different transcription factors to regulate gene expression (*Murakami, Kojima & Sakaki, 2003*). For example, the UCL1 (Urothelial cancer associated 1 conserved region) gene, which is specifically expressed in the bladder, is regulated under normal conditions by the transcription factor C/EBPα binding to the promoter, but transcription factor HIF-1α (Hypoxia-inducible factor 1 alpha) plays a major role in the regulation of UCA1 gene expression under conditions of cellular hypoxia (*Wang et al., 2006*, *2008*).

CONCLUSIONS

In the long-term evolution process, HK genes and TS genes showed significant differences in evolutionary constraint and evolutionary trend. HK promoters are more conservative

than TS promoters. TS genes exhibited more complex regulatory patterns than HK genes. The adaptation of organisms to different environments may be achieved through the regulation of genes by TS motifs.

ACKNOWLEDGEMENTS

We thank all of the contributors of the RNA-seq data sets and the anonymous reviewers for helpful suggestions on the manuscript.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The research was supported by the National Natural Science Foundation of China (31560310, 31760302 and 31272416). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Natural Science Foundation of China: 31560310, 31760302 and 31272416.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Kai Wei conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, approved the final draft.
- Lei Ma conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Tingting Zhang conceived and designed the experiments, contributed reagents/materials/analysis tools, approved the final draft.

Data Availability

The following information was supplied regarding data availability:
The details of the third-party datasets are available in Table S1.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.7204#supplemental-information.

REFERENCES

Abe H, Gemmell NJ. 2014. Abundance, arrangement, and function of sequence motifs in the chicken promoters. *BMC Genomics* **15(1)**:900 DOI 10.1186/1471-2164-15-900.

Andersson L. 2009. Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. *Genetica* **136(2)**:341–349 DOI 10.1007/s10709-008-9312-4.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37(Web Server)**: W202–W208 DOI 10.1093/nar/gkp335.

Bao T, Li H, Zhao X. 2012. Distribution of nucleosome binding motifs around the functional sites of human housekeeping genes. In: *5th International Conference on Biomedical Engineering and Informatics.* Chongqing, China, 876–879 DOI 10.1109/BMEI.2012.6513116.

Bellora N, Farré D, Albà MM. 2007. Positional bias of general and tissue-specific regulatory motifs in mouse gene promoters. *BMC Genomics* **8(1)**:459 DOI 10.1186/1471-2164-8-459.

Brandman O, Stewart-Ornstein J, Wong D, Larson A, Williams CC, Li GW, Zhou H, King D, Shen PS, Weibezahn J, Dunn JG, Rouskin S, Inada T, Frost A, Weissman JS. 2012. A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell* **151(5)**:1042–1054 DOI 10.1016/j.cell.2012.10.044.

Butte AJ, Dzau VJ, Glueck SB. 2001. Further defining housekeeping, or "maintenance," genes focus on "A compendium of gene expression in normal human tissues". *Physiological Genomics* **7(2)**:95–96 DOI 10.1152/physiolgenomics.2001.7.2.95.

Castillo-Davis CI, Hartl DL, Achaz G. 2004. cis-Regulatory and protein evolution in orthologous and duplicate genes. *Genome Research* **14(8)**:1530–1536 DOI 10.1101/gr.2662504.

Chang C-W, Cheng W-C, Chen C-R, Shu W-Y, Tsai M-L, Huang C-L, Hsu IC. 2011. Identification of human housekeeping genes and tissue-selective genes by microarray meta-analysis. *PLOS ONE* **6(7)**:e22859 DOI 10.1371/journal.pone.0022859.

Chen Y, Cunningham F, Rios D, McLaren WM, Smith J, Pritchard B, Spudich GM, Brent S, Kulesha E, Marin-Garcia P, Smedley D, Birney E, Flicek P. 2010. Ensembl variation resources. *BMC Genomics* **11**:293 DOI 10.1186/1471-2164-11-293.

Chin C-S, Chuang JH, Li H. 2005. Genome-wide regulatory complexity in yeast promoters: separation of functionally conserved and neutral sequence. *Genome Research* **15(2)**:205–213 DOI 10.1101/gr.3243305.

Dasmeh P, Serohijos AW, Kepp KP, Shakhnovich EI. 2014. The influence of selection for protein stability on dN/dS estimations. *Genome Biology and Evolution* **6(10)**:2956–2967 DOI 10.1093/gbe/evu223.

De Jonge HJM, Fehrmann RSN, De Bont ES, Hofstra RMW, Gerbens F, Kamps WA, De Vries EGE, Van Der Zee AG, Te Meerman GJ, Ter Elst A. 2007. Evidence based selection of housekeeping genes. *PLOS ONE* **2(9)**:e898 DOI 10.1371/journal.pone.0000898.

Eisenberg E, Levanon EY. 2003. Human housekeeping genes are compact. *Trends in Genetics* **19(7)**:362–365 DOI 10.1016/S0168-9525(03)00140-9.

Eisenberg E, Levanon EY. 2013. Human housekeeping genes, revisited. *Trends in Genetics* **29(10)**:569–574 DOI 10.1016/j.tig.2013.05.010.

Elango N, Yi SV. 2011. Functional relevance of CpG island length for regulation of gene expression. *Genetics* **187(4)**:1077–1083 DOI 10.1534/genetics.110.126094.

Farré D, Bellora N, Mularoni L, Messeguer X, Albà MM. 2007. Housekeeping genes tend to show reduced upstream sequence conservation. *Genome Biology* **8(7)**:R140 DOI 10.1186/gb-2007-8-7-r140.

Fenouil R, Cauchy P, Koch F, Descostes N, Cabeza JZ, Innocenti C, Ferrier P, Spicuglia S, Gut M, Gut I, Andrau JC. 2012. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Research* **22(12)**:2399–2408 DOI 10.1101/gr.138776.112.

Gardiner-Garden M, Frommer M. 1987. CpG islands in vertebrate genomes. *Journal of Molecular Biology* **196(2)**:261–282 DOI 10.1016/0022-2836(87)90689-9.

Gemayel R, Vinces MD, Legendre M, Verstrepen KJ. 2010. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annual Review of Genetics* **44(1)**:445–477 DOI 10.1146/annurev-genet-072610-155046.

Halees AS. 2003. PromoSer: a large-scale mammalian promoter and transcription start site identification service. *Nucleic Acids Research* **31(13)**:3554–3559 DOI 10.1093/nar/gkg549.

Hill DW, Leiferman JA, Lynch NA, Dangelmaier BS, Burt SE. 1998. Temporal specificity in adaptations to high-intensity exercise training. *Medicine and Science in Sports and Exercise* **30(3)**:450–455 DOI 10.1097/00005768-199803000-00017.

Holder N, Klein R. 1999. Eph receptors and ephrins: effectors of morphogenesis. *Development* **126(10)**:2033–2044.

Hong S-J, Lee H-J, Oh J-H, Jung S-H, Min K-O, Choi S-W, Rhyu M-G. 2013. Age-related methylation patterning of housekeeping genes and tissue-specific genes is distinct between the stomach antrum and body. *Epigenomics* **5(3)**:283–299 DOI 10.2217/epi.13.17.

Horton BM, Hudson WH, Ortlund EA, Shirk S, Thomas JW, Young ER, Zinzow-Kramer WM, Maney DL. 2014. Estrogen receptor alpha polymorphism in a species with alternative behavioral phenotypes. *Proceedings of the National Academy of Sciences of the United States of America* **111(4)**:1443–1448 DOI 10.1073/pnas.1317165111.

Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37(1)**:1–13 DOI 10.1093/nar/gkn923.

Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4(1)**:44–57 DOI 10.1038/nprot.2008.211.

Hurst LD. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends in Genetics* **18(9)**:486–487 DOI 10.1016/S0168-9525(02)02722-1.

Iwama H, Gojobori T. 2004. Highly conserved upstream sequences for transcription factor genes and implications for the regulatory network. *Proceedings of the National Academy of Sciences of the United States of America* **101(49)**:17156–17161 DOI 10.1073/pnas.0407670101.

Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Chèneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, Sandelin A, Vandepoele K, Lenhard B, Ballester B, Wasserman WW, Parcy F, Mathelier A. 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Research* **46(D1)**:D1284 DOI 10.1093/nar/gkx1188.

Kikin O, D'Antonio L, Bagga PS. 2006. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Research* **34(Web Server)**:W676–W682 DOI 10.1093/nar/gkl253.

Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, Kersey P, Flicek P. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* **2011(0)**:bar030 DOI 10.1093/database/bar030.

Kodama Y, Shumway M, Leinonen R, On behalf of the International Nucleotide Sequence Database Collaboration. 2012. The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Research* **40(D1)**:D54–D56 DOI 10.1093/nar/gkr854.

Kouadjo KE, Nishida Y, Cadrin-Girard JF, Yoshioka M, St-Amand J. 2007. Housekeeping and tissue-specific genes in mouse tissues. *BMC Genomics* **8(1)**:127 DOI 10.1186/1471-2164-8-127.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33(7)**:1870–1874 DOI 10.1093/molbev/msw054.

Labarga A, Valentin F, Anderson M, Lopez R. 2007. Web services at the European bioinformatics institute. *Nucleic Acids Research* **35(Web Server)**:W6–W11 DOI 10.1093/nar/gkm291.

Larsen PF, Eg Nielsen E, Hansen MM, Wang T, Meier K, Pertoldi C, Loeschcke V. 2013. Tissue specific haemoglobin gene expression suggests adaptation to local marine conditions in North Sea flounder (*Platichthys flesus L.*). *Genes and Genomics* **35(4)**:541–547 DOI 10.1007/s13258-013-0101-9.

Lawson MJ, Zhang L. 2008. Housekeeping and tissue-specific genes differ in simple sequence repeats in the 5′-UTR region. *Gene* **407(1–2)**:54–62 DOI 10.1016/j.gene.2007.09.017.

Lee S, Kohane I, Kasif S. 2005. Genes involved in complex adaptive processes tend to have highly conserved upstream regions in mammalian genomes. *BMC Genomics* **6**:168 DOI 10.1186/1471-2164-6-168.

Mayer C, Leese F, Tollrian R. 2010. Genome-wide analysis of tandem repeats in Daphnia pulex–a comparative approach. *BMC Genomics* **11(1)**:277 DOI 10.1186/1471-2164-11-277.

Murakami M, Kojima T, Sakaki Y. 2003. Detection of tissue specific genes by putative regulatory motifs in human promoter sequences. *Genome Informatics* **14**:408–409 DOI 10.11234/gi1990.14.408.

Nei M, Kumar S. 2000. *Molecular evolution and phylogenetics*. Oxford: Oxford University Press, 52–72.

Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLOS ONE* **7(2)**:e30619 DOI 10.1371/journal.pone.0030619.

Robertson KD. 2002. DNA methylation and chromatin-unraveling the tangled web. *Oncogene* **21(35)**:5361–5379 DOI 10.1038/sj.onc.1205609.

Sharif J, Endo TA, Toyoda T, Koseki H. 2010. Divergence of CpG island promoters: a consequence or cause of evolution? *Development, Growth & Differentiation* **52(6)**:545–554 DOI 10.1111/j.1440-169X.2010.01193.x.

She X, Rohl CA, Castle JC, Kulkarni AV, Johnson JM, Chen R. 2009. Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. *BMC Genomics* **10(1)**:269 DOI 10.1186/1471-2164-10-269.

Storey J. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society* **64(3)**:479–498 DOI 10.1111/1467-9868.00346.

Suzuki Y, Yamashita R, Shirota M, Sakakibara Y, Chiba J, Mizushima-Sugano J, Nakai K, Sugano S. 2004. Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions. *Genome Research* **14(9)**:1711–1718 DOI 10.1101/gr.2435604.

Thomas D, Finan C, Newport MJ, Jones S. 2015. DNA entropy reveals a significant difference in complexity between housekeeping and tissue specific gene promoters. *Computational Biology and Chemistry* **58**:19–24 DOI 10.1016/j.compbiolchem.2015.05.001.

Thorrez L, Laudadio I, Van Deun K, Quintens R, Hendrickx N, Granvik M, Lemaire K, Schraenen A, Van Lommel L, Lehnert S, Aguayo-Mazzucato C, Cheng-Xue R, Gilon P, Van Mechelen I, Bonner-Weir S, Lemaigre F, Schuit F. 2011. Tissue-specific disallowance of housekeeping genes: the other face of cell differentiation. *Genome Research* **21(1)**:95–105 DOI 10.1101/gr.109173.110.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25(9)**:1105–1111 DOI 10.1093/bioinformatics/btp120.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28(5)**:511–515 DOI 10.1038/nbt.1621.

Urrutia AO, Hurst LD. 2001. Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. *Genetics* **159(3)**:1191–1199 DOI 10.1023/A:1013367100865.

Valipour E, Kowsari A, Bayat H, Banan M, Kazeminasab S, Mohammadparast S, Ohadi M. 2013. Polymorphic core promoter GA-repeats alter gene expression of the early embryonic developmental genes. *Gene* **531(2)**:175–179 DOI 10.1016/j.gene.2013.09.032.

Vavouri T, Lehner B. 2012. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biology* **13(11)**:R110 DOI 10.1186/gb-2012-13-11-r110.

Vinces MD, Legendre M, Caldara M, Hagihara M, Verstrepen KJ. 2009. Unstable tandem repeats in promoters confer transcriptional evolvability. *Science* **324(5931)**:1213–1216 DOI 10.1126/science.1170097.

Wang F, Li X, Xie X, Zhao L, Chen W. 2008. UCA1, a non-protein-coding RNA up-regulated in bladder carcinoma and embryo, influencing cell growth and promoting invasion. *FEBS Letters* **582(13)**:1919–1927 DOI 10.1016/j.febslet.2008.05.012.

Wang X-S, Zhang Z, Wang H-C, Cai J-L, Xu Q-W, Li M-Q, Chen Y-C, Qian X-P, Lu T-J, Yu L-Z, Zhang Y, Xin D-Q, Na Y-Q, Chen W-F. 2006. Rapid identification of UCA1 as a very sensitive and specific unique marker for human bladder carcinoma. *Clinical Cancer Research* **12(16)**:4851–4858 DOI 10.1158/1078-0432.ccr-06-0134.

Wei K, Zhang T, Ma L. 2018. Divergent and convergent evolution of housekeeping genes in human-pig lineage. *PeerJ* **6**:e4840 DOI 10.7717/peerj.4840.

Wittkopp PJ, Kalay G. 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics* **13(1)**:59–69 DOI 10.1038/nrg3095.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. 2003. The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution* **20(9)**:1377–1419 DOI 10.1093/molbev/msg140.

Xu Y, He B, Li R, Pan Y, Gao T, Deng Q, Sun H, Song G, Wang S. 2014. Association of the polymorphisms in the Fas/FasL promoter regions with cancer susceptibility: a systematic review and meta-analysis of 52 studies. *PLOS ONE* **9(3)**:e90090 DOI 10.1371/journal.pone.0090090.

Zhang L, Li W-H. 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Molecular Biology and Evolution* **21(2)**:236–239 DOI 10.1093/molbev/msh010.

Zhu J, He F, Hu S, Yu J. 2008. On the nature of human housekeeping genes. *Trends in Genetics* **24(10)**:481–484 DOI 10.1016/j.tig.2008.08.004.