

RESEARCH ARTICLE

Open Access

# Abundance, arrangement, and function of sequence motifs in the chicken promoters

Hideaki Abe<sup>1\*</sup> and Neil J Gemmell<sup>1,2</sup>

## Abstract

**Background:** Eukaryotic promoters are regions containing various sequence motifs necessary to control gene transcription. Much evidence has emerged showing that structural and/or contextual changes in regulatory elements can critically affect *cis*-regulatory activity. As sequence motifs can be key factors in maintaining complex promoter architectures, one effective approach to further understand the evolution of promoter regions in vertebrates is to compare the abundance and distribution patterns of sequence motifs in these regions between divergent species. When compared with mammals, the chicken (*Gallus gallus*) has a very different genome composition and sufficient genomic information to make it a good model for the exploration of promoter structure and evolution.

**Results:** More than 10% of chicken genes contained short tandem repeat (STR) in the region 2 kb upstream of promoters, but the total number of STRs observed in chicken is approximately half of that detected in human promoters. In terms of the STR motif frequencies, chicken promoter regions were more similar to other avian and mammalian promoters than these were to the entire chicken genome. Unlike other STRs, nearly half of the trinucleotide repeats found in promoters partly or entirely overlapped with CpG islands, indicating potential association with nucleosome positions. Moreover, the chicken promoters are abundant with sequence motifs such as poly-A, poly-G and G-quadruplexes, especially in the core region, that are otherwise rare in the genome. Most of sequence motifs showed strong functional enrichment for particular gene ontology (GO) categories, indicating roles in regulation of transcription and gene expression, as well as immune response and cognition.

**Conclusions:** Chicken promoter regions share some, but not all, of the structural features observed in mammalian promoters. The findings presented here provide empirical evidence suggesting that the frequencies and locations of STR motifs have been conserved through promoter evolution in a lineage-specific manner. Correlation analysis between GO categories and sequence motifs suggests motif-specific constraints acting on gene function.

**Keywords:** Promoter, Transcription factor binding site, Short tandem repeat, CpG island, G-quadruplex, Gene ontology, Transcription, GC content, Selection, *Cis*-regulatory element

## Background

Promoters are well-characterized transcriptional *cis*-regulatory sequences in complex genomes [1]. They generally locate immediately upstream of a transcription start site (TSS) and have a variety of sequence motifs that participate in gene regulation [2]. The key elements include transcription factor binding sites (TFBSs), short tandem repeat (STR), G-quadruplex (G4), and CpG island (CGI), which are frequently co-localized and otherwise integrated into combined motifs. Given these *cis*-regulatory DNA

sequences contain TFBSs and/or other regulatory modules that play a critical role in transcription, mutations that either alter affinity of TFBSs or disrupt spacing between existing TFBSs have the potential to affect *cis*-regulatory activity [3,4]. Throughout the last decade, empirical data have accumulated suggesting that mutations in regulatory elements could be a major cause of phenotypic divergence [5-7].

Recent studies have shown that eukaryotic promoters are rich in repetitive sequences — approximately 25% of yeast (*Saccharomyces cerevisiae*) genes contain at least one STR in their regulatory elements [8]. In humans, the Short Tandem Repeats in Regulatory Regions Table (STaRRRT)

\* Correspondence: habe2031@gmail.com

<sup>1</sup>Department of Anatomy, University of Otago, Dunedin, New Zealand  
Full list of author information is available at the end of the article

has shown that 5,264 STRs are present in the upstream regulatory region of 4,441 genes [9]. Recently, genes driven by repeat-containing promoters are reported to have significantly higher rates of transcriptional divergence than those without repeat elements, as corroborated by *in vitro* experiments showing that the gain or loss of STR units within promoters yields quantitative differences in gene expression [8,10,11]. For example, the EWS/FLI protein, which belongs to EST-type transcription factor (TF), directly binds to GGAA-motifs in the glutathione S-transferase M4 promoter, and transcriptional activity is highly dependent upon the number of repeats [12-14]. These facts, together with a higher conservation rate of STRs in regions proximal to TSSs than that in distal regions [15], strongly support a significant role for tandem repeats in differential transcriptional regulation.

Other DNA sequence motifs that affect chromatin structure have the potential to impact on gene expression by changing the accessibility of transcription and regulatory proteins to the DNA. G4 is one such sequence motif that has a four-stranded DNA structure held together by four or more tandem guanine tracts [16]. Recent work has shown that ~60% of the genes in warm-blooded animals represented by human, mouse, and chicken have at least one potential G-quadruplex sequence (PQS) within the 5 kb region upstream of TSS [17]. Some of the G4 sequences so far examined appear to act as silencer elements in the promoter regions [18]. The clearest evidence for a role of G4 structure in transcriptional regulation comes from empirical study of the onco-gene *c-myc* [19]. Disruption of G4 motifs in the *c-myc* promoter resulted in increased gene expression, whereas stabilization of the G4 decreased transcription, raising the strong possibility that G4 formation affects the deposition of regulatory proteins and histones on double-stranded DNA [17,18].

CpG islands are CG-rich stretches that have been found in approximately half of mammalian promoters at or near the TSS [20]. In vertebrates, promoters with CGI are characterized by the presence of many TSS and high transcriptional activity in multiple tissues, whereas promoters without CGIs are defined by a single TSS and show tightly regulated expression in specific tissues [21,22]. Correlations between gene ontologies (GOs) and CGI length hint at the important role of CGI in higher-order chromatin structures via methylation [23]. Most of the CGIs in chicken promoters remain hypomethylated, contributing to nucleosome-free regions over the promoter [24]. Open CGI/CG-rich promoters would naturally lack nucleosome scaffolds that are required to adopt an open conformation, and different histone modification patterns have been observed between genes with or without CGI promoters [25].

Stochastic and spatial data on the aforementioned sequence motifs that may modify chromatin structure and affect transcription are essential for understanding the

nature of regulatory complexity in higher organisms. In this study we investigate the enrichment and arrangement of several sequence motifs within chicken (*Gallus gallus*) promoters to shed light on compositional structure in avian promoter sequences and their association with gene functions. Birds are hypothesized to have diverged from a common ancestor with mammals around 300 million years ago [26]. Thus, it is of interest to investigate whether previous findings on the distribution and abundance of sequence motifs in mammalian promoters would be consistent with those derived from chicken promoters, given the significant divergence times and very different evolutionary trajectories these lineages have followed.

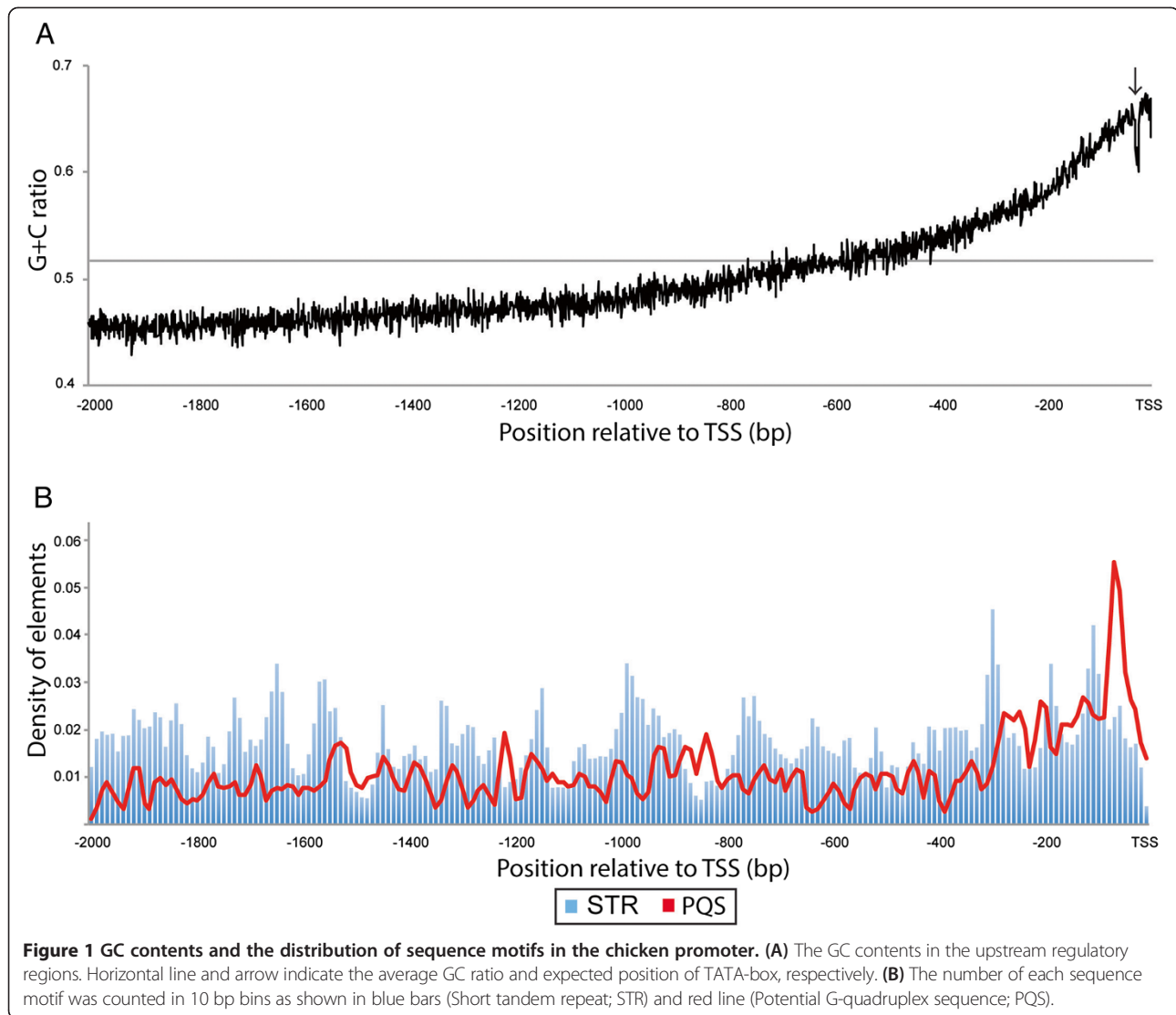
## Results

### GC content and CpG island density

Chicken promoter sequences had considerable variation in their GC contents, ranging from 31.9 to 73.6%, and their overall average was 51.5% (Additional file 1). The average G + C ratio increased gradually as it approached the TSS (Figure 1A). A sharp decline detected in the core promoter region (-31 to -23 bp) was presumably attributed to the presence of a TATA box. The Newcpgreport software identified 2,809 CGIs in promoter regions, and the number of CGIs-containing genes reached to 2,251 (58.3%). This was almost the same with that reported elsewhere [27], even though search conditions were not exactly the same. The distribution of CpG observed/expected (O/E) ratio was slightly lower (mode O/E = 0.95 ~ 1.00; Additional file 2) as compared with the previous report [22] (mode O/E = 1.1 ~ 1.2). This might be due to the fact that some of CGIs identified in this study were truncated at TSS. We referred to the chicken promoters harboring long CGIs (>800 bp in total) as "long CGI" (LCGI) promoters, because they occupied the top10% of the genes in terms of CGI length. "No CGI" (NCGI) promoter was hereafter used to refer to promoters without CGI.

### Abundance of short tandem repeats and G-quadruplexes in chicken promoters

Table 1 summarizes the frequencies of STR motifs in chicken and other mammalian promoters as well as those identified in the entire chicken genome. The total number of STRs observed in chicken promoters was almost equivalent to that observed in both duck and zebra finch promoters. In all avian species, approximately 10% of genes had at least 6 perfect repeat units of 2-6 nucleotide STR in 2 kb upstream region from TSS, whereas mouse and human promoters contained a much larger number of STRs in the same regions. The number of STR motifs counted per human promoter sequence was more than double that of chicken promoter. A rank correlation analysis showed that the STR motif frequency in chicken promoter shares a significant level of similarity with all



the promoter sets examined here (Table 1). However, the comparison of STR motif frequencies between chicken promoter and the entire chicken genome exhibited the lowest similarity values in all statistical parameters (i.e., correlation coefficient, Kendall's  $\tau$  distance, and 2-sided  $p$ -value for Kendall  $\tau$  rank correlation), confirming the paucity of STRs previously described for this genome [28]. Several STR motifs also showed taxon-specific differences in their distribution. For example, avian promoters were distinguishable from those of mammals on the basis of a lower frequency of AG/CT motifs. The duck promoter contained an extremely low number of GC-rich STR motifs such as CG and CCG.

Chicken STRs were not equally distributed, but rather varied over the region of promoters (Figure 1B). A total of 302 PQSs were identified in the chicken promoters but unlike STRs, PQSs were especially accumulated in the core promoter region. The number of PQS identified

in this study was much fewer than that previously reported in transcriptional regulatory region of chicken genome [29]. This was probably due to differences in the stringency of PQS screening as well as in length of target promoter region.

#### Heterogeneity in the pattern of STR expansion between avian and mammalian promoters

The pattern of STR unit expansion was quite different between avian and mammalian promoters. All avian promoters examined here exhibited a similar trend of STR expansion with significantly larger number of STR units in tetra-, penta-, and hexanucleotide (hereafter tetranucleotide) repeats against dinucleotide repeats (Figure 2; Mann-Whitney U-test [chicken];  $z = 5.56$ ,  $p < 0.001$ , [duck];  $z = 7.33$ ,  $p < 0.001$ , [zebra finch];  $z = 10.75$ ,  $p < 0.001$ ). While human and mouse promoters were characterized with much longer dinucleotide repeats as compared with the tetranucleotide repeats

**Table 1 Comparison of STR motif distributions among chicken and avian/mammalian promoters**

Assembly Motif	Promoter										Genome*	
	Chicken		Duck		Zebra finch		Mouse		Human		Chicken	
	<i>galGal4</i>		<i>BGI_duck_1.0</i>		<i>taeGut3.2.4</i>		<i>mm10</i>		<i>hg19</i>		<i>galGal3</i>	
	Count	Freq.	Count	Freq.	Count	Freq.	Count	Freq.	Count	Freq.	Count	Freq.
AC	405	(25.9)	433	(31.5)	378	(23.2)	6006	(40.0)	3189	(35.6)	97994	(26.1)
AT	276	(17.6)	294	(21.4)	270	(16.6)	739	(4.9)	887	(9.9)	52845	(14.1)
AG	174	(11.1)	174	(12.7)	186	(11.4)	2572	(17.1)	1356	(15.1)	98988	(26.4)
AGG	131	(8.4)	98	(7.1)	170	(10.4)	426	(2.8)	192	(2.1)	13509	(3.6)
AAT	68	(4.3)	61	(4.4)	40	(2.5)	264	(1.8)	413	(4.6)	18974	(5.1)
CCG	66	(4.2)	1	(0.1)	87	(5.3)	200	(1.3)	462	(5.2)	722	(0.2)
AAC	46	(2.9)	38	(2.8)	16	(1.0)	356	(2.4)	283	(3.2)	13218	(3.5)
AGC	40	(2.6)	26	(1.9)	72	(4.4)	73	(0.5)	67	(0.7)	31737	(8.5)
AAAT	30	(1.9)	49	(3.6)	21	(1.3)	432	(2.9)	393	(4.4)	6017	(1.6)
AAAG	26	(1.7)	21	(1.5)	10	(0.6)	583	(3.9)	223	(2.5)	2769	(0.7)
CG	26	(1.7)	0	(-)	36	(2.2)	177	(1.2)	207	(2.3)	687	(0.2)
ATCCC	18	(1.2)	6	(0.4)	14	(0.9)	1	(0.0)	0	(-)	ND	
AAAC	17	(1.1)	33	(2.4)	14	(0.9)	563	(3.7)	183	(2.0)	6290	(1.7)
AAGG	15	(1.0)	11	(0.8)	13	(0.8)	274	(1.8)	122	(1.4)	1094	(0.3)
AAG	13	(0.8)	9	(0.7)	5	(0.3)	246	(1.6)	75	(0.8)	15000	(4.0)
ACC	13	(0.8)	8	(0.6)	8	(0.5)	150	(1.0)	224	(2.5)	8154	(2.2)
AGAGG	13	(0.8)	5	(0.4)	7	(0.4)	54	(0.4)	0	(-)	ND	
ATCC	12	(0.8)	5	(0.4)	32	(2.0)	41	(0.3)	29	(0.3)	436	(0.1)
ACGGC	12	(0.8)	0	(-)	12	(0.7)	0	(-)	0	(-)	ND	
AAGGG	11	(0.7)	22	(1.6)	6	(0.4)	23	(0.2)	4	(0.0)	ND	
STR/seq**		0.114		0.101		0.104		0.555		0.231		-
vs.chicken promoter												
Correlation	<i>r</i>	-		0.985		0.984		0.852		0.912		0.841
Kendall	$\tau$	-		0.595		0.738		0.581		0.656		0.487
	<i>p</i>	-		< 0.001		< 0.0001		< 0.001		< 0.0001		< 0.05

\*Data derived from [50]. ND: no data available.

\*\*The number of STR motif counted per promoter sequence.

(Mann–Whitney U-test [human];  $z = 6.14$ ,  $p < 0.001$ , [mouse];  $z = 27.72$ ,  $p < 0.001$ ). The number of dinucleotide repeat units was significantly lower in chicken promoters than that of human (Mann–Whitney U-test;  $z = 8.69$ ,  $p < 0.001$ ), but the reverse was true for tetranucleotide repeats, which are prevalent in chicken versus human (Mann–Whitney U-test;  $z = 3.69$ ,  $p < 0.001$ ).

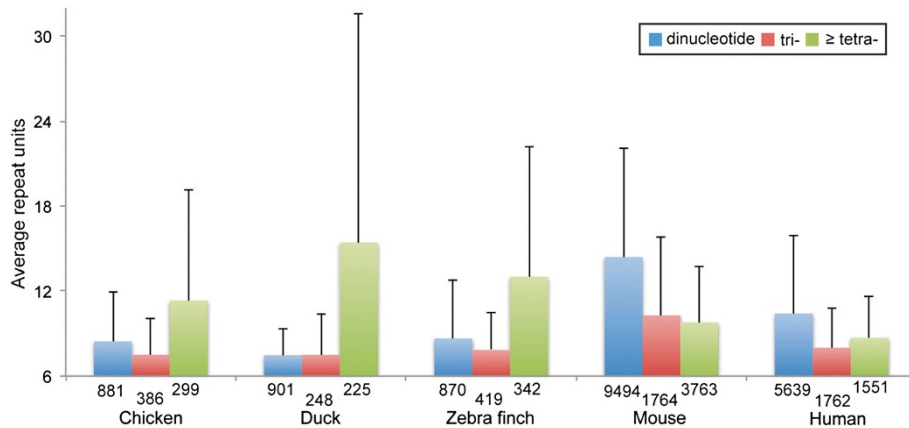
#### Distribution of sequence motifs in conjunction with CpG islands

Some chicken promoters contained multiple sequence motifs, which either co-existed with or were integrated with CGIs. The maximum number of CGI, STR, and PQS identified in a single promoter was 5, 5, and 4, respectively. The relative abundance of STRs did not change the rate of multiple CGIs, and *vice versa*. However, the co-existence of PQSs significantly affected the CGI number in promoter

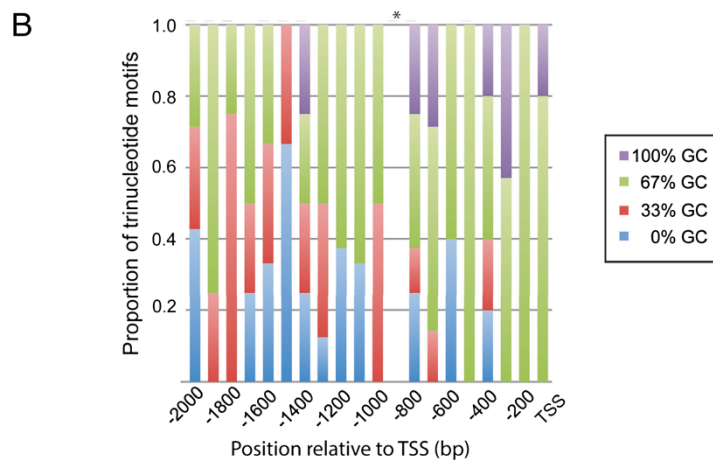
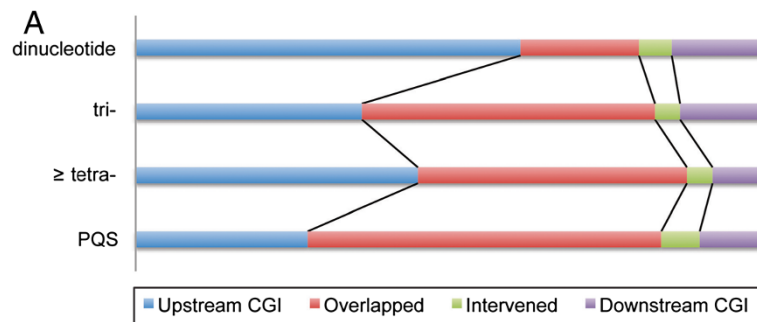
regions (Fisher's exact test (FET);  $p < 0.001$ ; Additional file 3). The number of STRs that overlapped with CGI motifs was significantly different between di- and trinucleotide repeats (FET;  $p < 0.001$ ; Figure 3A). Most of dinucleotide STRs were located upstream of CGIs, whereas a higher proportion of trinucleotide repeats was found to be overlapped with CGIs. We anticipated that almost all of PQSs would be found to overlap with CGIs, but actually 43.7% of PQSs located in up- or down-stream of the CGI.

#### Low repeat number and positional bias of trinucleotide repeats in chicken promoter

The average number of trinucleotide repeat units were the smallest both in avian and mammalian promoters in all STR periods. All chicken trinucleotide repeats were divided into four groups that had different numbers of guanine or cytosine bases in each repeat unit (hereafter



**Figure 2** The average number of STR units detected in avian and mammalian promoters. Note that only perfect repeats with at least six units are considered here. The number of STR loci is described under each bar. The total number of promoter sequences for five organisms are: chicken (13,754), duck (13,583), zebra finch (15,633), mouse (27,046), and human (38,687).



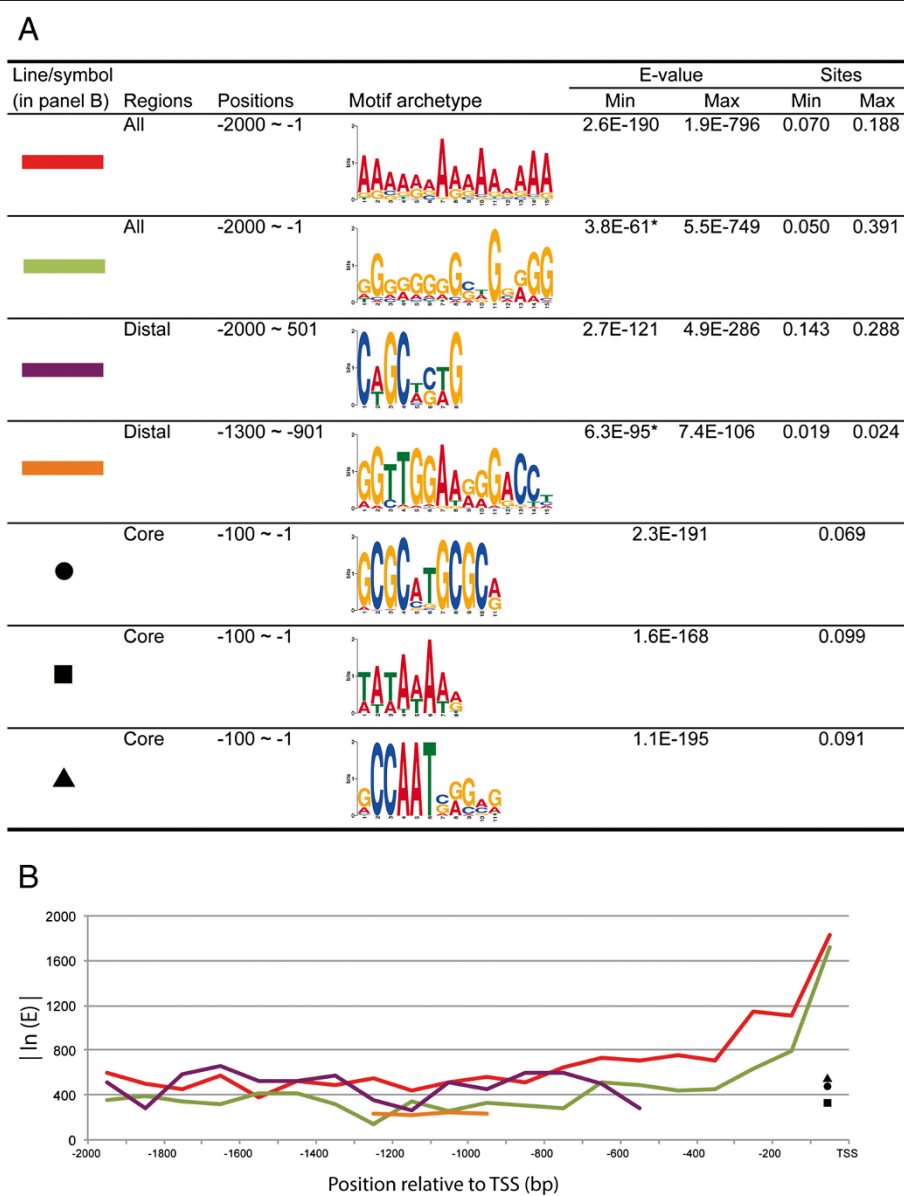
**Figure 3** Biases in the location of sequence motifs found in the chicken promoter. **(A)** The locations of STRs and PQSs are shown in conjunction with a CpG island (CGI). Upstream and downstream CGI indicate the locations of sequence motifs. "Overlapped" means that at least part of a sequence motif is overlapped with a CGI. "Intervened" is used when a sequence motif is found to be flanked by two or more CGIs (not overlapped). There is significant difference in the locations between di- and trinucleotide repeats (Fisher's exact test;  $p < 0.001$ ). **(B)** The distribution of trinucleotide repeats in chicken promoter. Trinucleotide repeats are classified into four categories by the number of guanine or cytosine in each repeat unit. Note that no trinucleotide repeat was found in the position at -901 to -1000 bp (asterisk).

referred as 100%GC, 67%GC, 33%GC, and 0%GC). Figure 3B shows different pattern of trinucleotide STR distribution between 100%GC and low GC group (33% and 0%GC; FET;  $p < 0.01$ ). The motifs with 100%GC distributed mostly in the proximal and core regions, while 33% and 0%GC motifs were predominantly found in the distal part of promoter.

#### Conserved motifs identified in the chicken promoter

The MEME Suite [30] was used to detect conserved motifs that might affect gene regulation. The most and second most common blocks in the chicken promoter

were poly-A and poly-G prevalent in promoter regions (Figure 4A and B). Both polypurine repeats were relatively constant in the motif frequencies through the distal promoter (-2000 to -500 bp), but gradually increased through the proximal region (-500 to -100 bp). Both motifs were characterized by a steep increase in the core promoter region (-100 bp to TSS). The other motif C[A/T]GC[A/T][C/G][A/T]G also appeared in the distal promoter, but was seldom seen either in proximal or in core regions. This motif was compared to known motifs in JASPAR Vertebrates [31] and UniPROBE Mouse database [32] by TOMTOM

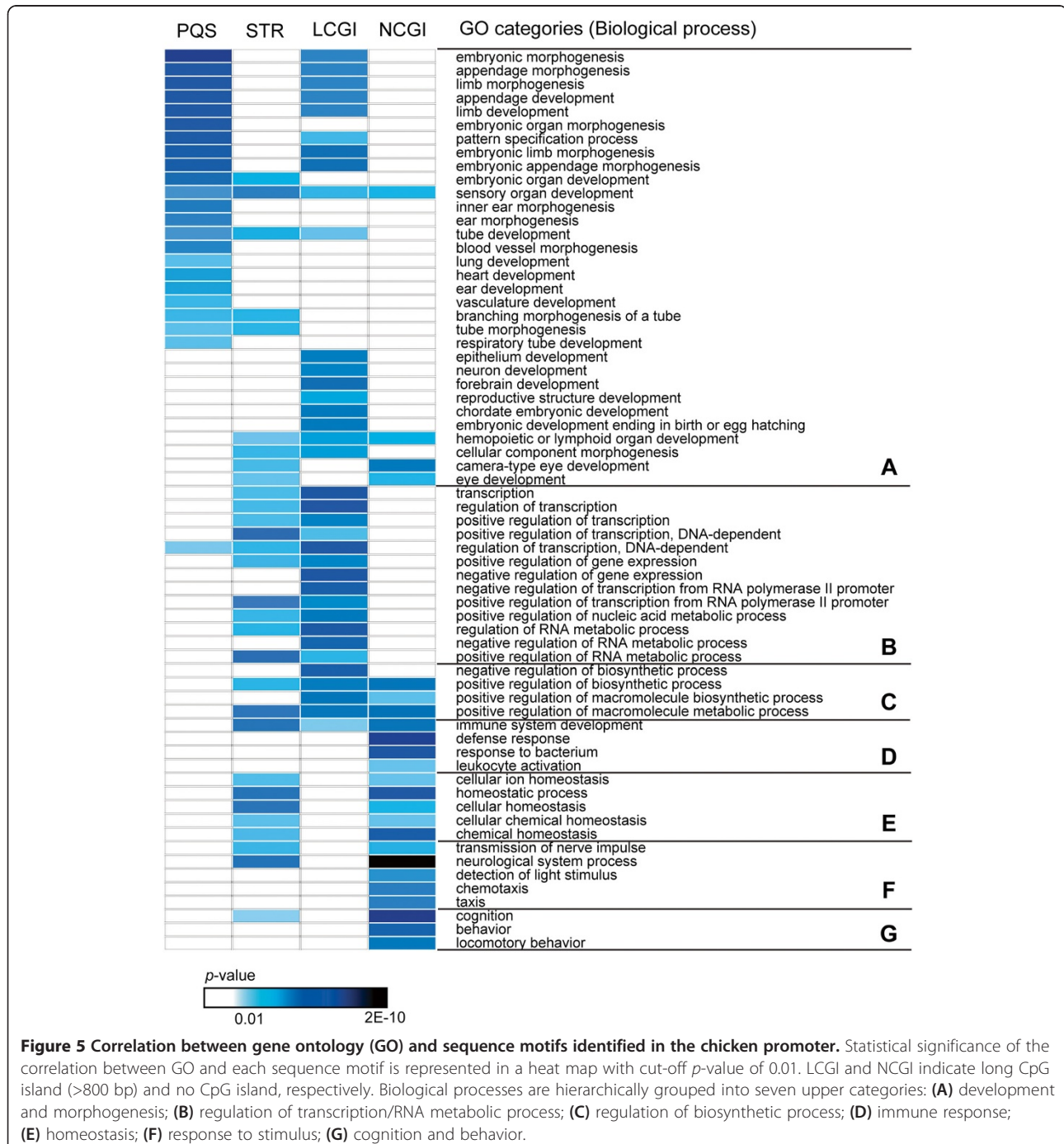


**Figure 4 Conserved motifs identified by MEME Suite software.** Chicken promoter sequences were divided into 100 bp bins and used for MEME search. **(A)** Each archetype of motifs is shown with the maximum and minimum E-values and the proportion of motif identified in a dataset. Cut-off E-value was set in 1.0E-100 except two cases (shown with asterisk). **(B)** The distribution of conserved motifs in the upstream regulatory regions.

[33] ver 4.9.1. As a result, the following 10 TF motifs were detected with significant level of similarities (FET;  $p < 0.01$ ); Zfp691, TFAP4, Zic1, NHLH1, Zbtb3, Zic3, ZEB1, Osr2, Tcf3, and Gfi1b. Approximately 10% of chicken promoters contained a TATA box in their core region and the number and location of TATA boxes in the chicken promoters were comparable to those reported in the genome-wide analysis of mammalian promoters, showing -30 and -31 from TSS as the preferred sites [34].

#### Gene functions associated with sequence motifs

For functional annotation analysis using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) program [35], chicken promoters were grouped into four sets of genes depending on the presence or absence of sequence motifs (i.e., PQS, STR, LCGI, and NCGI). The heat map shown in Figure 5 clearly illustrates a bias in biological processes that exhibit significant probabilities. PQSs were predominantly detected in



the genes associated with development and morphogenesis, while genes with STRs were less correlated with the particular GO terms. LCGI promoters were strongly associated with gene functions related to regulation of transcription and gene expression, whereas NCGI promoters associated with other gene functions such as immune response and cognition. A full list of GO categories found to be correlated with the sequence motifs is presented in Additional file 4.

## Discussion

In the present study, we show that chicken promoter sequences share some, but not all features with the human and mouse promoters. Although the frequency and variety of STR motif were highly conserved even between avian and mammalian promoters, chicken promoters had the least similarity with the entire chicken genome in terms of STR motif frequency. This finding is partly supported by previous data that showed the predominant STR motifs found in promoter region (TSS to -500 bp) were quite distinct from those detected in the other part of genes (i.e., 5' untranslated region (UTR), coding, intron, and 3' UTR) both in human and mouse [36]. The AC/GT was the most common motif in all the promoter sequences examined in this study, but AG/CT motifs were predominantly found in human 5' UTR and coding regions [36]. The inconsistency of predominant STR motifs between promoter and adjacent non-promoter regions seems to support the previous suggestion that STR motifs in promoters can alter gene expression as they expand or contract, with particular attention to secondary structures [37]. Considering that (AC/GT)<sub>n</sub> dinucleotide repeats have a propensity to form Z-DNA and occasionally block the movement of RNA polymerase when it occurs downstream of the TSS [38], repeat expansion or contraction of this motif might be constrained by the conformation of other sequence motifs that participate in transcriptional regulation. Hence, the AC/GT motif may be regarded as the most frequently used "tuning knob", as demonstrated by Bayele *et al.* [39], and this role appears to be evolutionarily conserved in vertebrates.

However, there are some points of difference in the STR motif frequencies and their expansion between avian and mammalian promoters. In mammalian promoters, the number of tetranucleotide STR units was significantly lower than those of avian promoters. One possible explanation of this pattern is that the expansion of tetranucleotide repeats might have been subjected to purifying selection to preserve some functions in human promoters. Such a scenario is also consistent with the frequently reported associations between tetranucleotide expansion and human diseases [40]. For instance, alteration in array length of TAAA affected the level of *nadA* transcription through modulation of the binding of the transcription factor IHF [41]. Another study on human

prostate cancer also suggested an involvement of TAAA tandem repeats as mediators of the expression of *PCa3* gene [42]. Further investigation is needed to elucidate any lineage-specific preference for STR expansion in vertebrate promoters.

Another important finding is that the density of STRs in the chicken promoter region is much lower than that estimated for human promoter. Taking into account that avian genomes contain much less STRs than mammals [43,44], it may simply reflect the difference in the occurrence rate of slippage-like indels across organisms, as suggested by Kruglyak *et al.* [45]. In this case the higher GC contents observed in chicken promoters [46] as well as relatively small genome size [47] are plausible reasons for the lower occurrence rate of slippage events.

Several previous works have shown *cis*-regulatory motifs to be constrained in a position- and/or distance-specific manner [48,49]. STRs are among the most plausible factors contributing to the changing spaces between functional elements in promoters. Our data clearly showed that the distribution and expansion of STR in chicken promoters are largely different among repeat unit classes. Dinucleotide tandem repeats were mainly found in non-CpG sites, whereas a higher rate of trinucleotide repeats were overlapped with CGI, with fewer repeats. Therefore, it is tempting to speculate that the expansion of trinucleotide STRs in chicken promoter is constrained either by a position and distance limitation or by direct targeting of TF. Furthermore, trinucleotide repeats showed skewed distribution between high and low ratio of guanine/cytosine in the repeat unit (Figure 3B). This finding is of great interest since the previous study on STR abundance in the chicken genome demonstrated that the rate of STR polymorphisms increases in high GC group (67% and 100% GC), exclusively in trinucleotide tandem repeats [50]. This discrepancy may be explained by the significant role of the trinucleotide tandem repeats as an enhancer/modulator of transcription in the core promoter region [51]. A previous *in vitro* experiment also supports the significance of 100% GC trinucleotide repeats as a key modulator of transcription, indicating that the insertion of (CGG)<sub>12</sub> into the *CYC1-lacZ* promoter increased gene expression about 10-fold, even other trinucleotide repeats of (CTG)<sub>12</sub> and (GAA)<sub>12</sub> had little effect [52]. All these facts imply that trinucleotide motifs with high guanine or cytosine contents, especially those found in the proximal and core promoter regions may have a pivotal role in the maintenance of an open chromatin structure, which will constrain STR expansion. Indeed, several studies clearly illustrated that GC-rich trinucleotide repeats are highly flexibility and possess a greater propensity to bend towards the major groove [53,54].

Motif identification using the MEME software revealed several conserved motifs either in all or in particular part



of chicken promoters. The poly-A was the most ubiquitous motif among them. Previous studies indicated that polypurine motifs are the most common STRs in the human genome and are particularly enriched in promoter regions [55]. It was suggested that poly-A might act to alter the stability or dynamics of nucleosomes, somehow enhancing the ability of gene activator proteins to bind nearby DNA target sites [56]. This hypothesis is well supported by our observation that poly-A are especially abundant in the core promoter region where maintenance of open chromatin structure is necessary. In contrast, the abundance of poly-G in core promoter is likely to provide potential binding sites for Sp1, which is a crucial TF for the expression of some genes. For example, the human *vascular endothelial growth factor* (*VEGF*) promoter contains a 39-bp poly-G sequence, located -85 to -50 bp relative to the TSS, including three potential Sp1 binding sites [57]. These independent studies give strong indications that sequence motifs – TFBS associations within the core promoter region may hold the key to deciphering the complexity of gene expression. In chicken promoters, we also detected another conserved motif, C[A/T]GC[A/T][G/C][A/T]G in the distal promoter. Cooper *et al.* reported that negative elements to human promoter activity were identified -1000 to -500 bp upstream of the TSS by their deletion analyses [58]. Therefore, it is possible that the conserved motifs that are unique to the distal part of the promoter may have some role as negative regulators of promoter activity.

CGIs are deeply involved in gene regulatory processes [59]. In particular, the length of CGIs is a pivotal factor in determining the number of RNA polymerase II binding sites in mammalian promoters [21]. In this study, LCGI promoters were strongly associated with the biological processes such as “regulation of transcription” (GO: 0045449, FET;  $p < 10^{-7}$ ) and “transcription” (GO: 0006350, FET;  $p < 10^{-7}$ ), whereas NCGI promoters were significantly involved in gene function that linked with “neurological system process” (GO: 0050877; FET;  $p < 10^{-9}$ ) and “defence response” (GO: 0006952; FET;  $p < 10^{-7}$ ). In addition, some of GO categories related with development and morphogenesis were moderately associated with LCGI promoters. These findings are analogous to the results obtained in mammalian promoters [21,22], indicating that an association between CGI lengths and particular gene functions is conserved, at least within warm-blooded vertebrates. In other words, we find that both the pattern of tissue-specific gene expression [60] and the motif-specific expression patterns are evolutionary conserved across the warm-blooded vertebrates. Another intriguing finding is that chicken promoters with PQS motifs are generally correlated with GO categories related to development and morphogenesis. It is notable that some specific biological processes such as (inner) ear morphogenesis and heart development were not significantly correlated with LCGI

but with PQS mediated promoters. This observation suggests that some PQSs that are non-overlapping with CGI might play a decisive role in gene expression through the fine-tuning of transcriptional activity. This, together with the recent finding that cell proliferation/cell-cycle could be regulated by presence of PQS – TFBS combinations in mammalian promoters [61] hint at the importance of positional context of sequence motifs. The case-by-case approach should be employed to reveal the underlying role of PQSs as transcriptional regulators in chicken promoters.

## Conclusions

This paper has provided novel findings from the investigation of sequence motifs in chicken promoter. The STR motif frequency in chicken promoters is similar with both those of other avian and mammalian promoters, but relatively divergent from that of the rest of the chicken genome. We have also revealed that the pattern of STR unit expansion is largely different between avian and mammalian promoters. These findings indicate that STR sequence motifs in promoter regions are strongly conserved and may play roles in transcription regulation, but that lineage specific pressures on motif expansion may exist. Although GC content in chicken promoter is higher than in mammals, the same pattern of correlations between biological processes and CGI lengths can be found in this study. Moreover, we have shown that PQSs are exclusively recognized in a set of genes involved in development and morphogenesis. Searching for lineage-specific patterns of various sequence motifs in promoter regions will certainly extend our understanding of the relationship between structural complexity of promoters and functional consequences.

## Methods

### Data preparation

In order to compare the STR motif distributions between chicken and other animals (duck, zebra finch, mouse, and human), the 2 kb upstream sequences from flanking genes were obtained either from Ensemble BioMart [62] or from the University of California, Santa Cruz (UCSC) Genome Browser [63,64]. In this study, we expediently defined the “promoter region” as 2 kb upstream of TSS, while acknowledging that *cis*-regulatory elements can also be found much further upstream, 3' of the gene, or inside of genes [65]. Chicken sequences 2 kb upstream of the TSS of RefSeq genes with annotated 5' UTRs were also downloaded from the UCSC browser and used for structural and function-related analyses. Consequently, we obtained 3,858 promoter sequences from the corresponding genes with functional annotations (up to 22.6% of total genes in the chicken genome (UCSC release *galGal4*)).

### Identification of sequence motifs

CGIs were identified using the Newcpgreport downloaded from the European Bioinformatics Institute (EMBL-EBI) Browser [66]. The traditional criteria were used to identify CGIs: (i) base composition of guanine and cytosine in a window (100 bp) exceeded 50%, (ii) minimum length was 200 bp, and (iii) the ratio of observed to expected number of CpG dinucleotides (CpG O/E) was more than 0.6 [67]. We did not try to seek 3' end of CGIs when they were overlapping with TSS and reached to 5' UTRs. Thus, some of CGIs identified in this study were truncated up to several hundreds bp in length.

The avian and mammalian promoter sequences were screened for STRs using the WebSat [68] and Phobos software [69] integrated into the STADEN package [70]. We searched STRs under following conditions: (i) only perfect repeats were considered, (ii) repeats periods were 2, 3, 4, 5, and 6, (iii) STRs with at least six repeat units were scored, and (iv) combined STRs with two or more motifs were counted separately. We did not take mononucleotide repeats into consideration, mainly due to their uncertainty in the repeat number. The data on STR frequency distributions were subjected to the non-parametric Kendall  $\tau$  trend analysis between chicken and other animals under the null hypothesis of no association between two data sets. The bottom 10% of minor motifs in chicken STR frequency were eliminated from the data set. In addition, information on STR occurrence in the entire chicken genome was obtained from the previous study [50] and compared with those detected within promoter regions to examine the heterogeneity of STR motif frequency between them.

PQSs were detected by the Quadruplex forming G-Rich Sequences (QGRS) Mapper software [71]. The details of search parameters were as follows: (i) max length of PQSs was 30, (ii) the minimum number of tetrads in a G4 was four, and (iii) the minimum loop size was set to zero. Note that these parameters led some motifs being double-counted in both STRs and PQSs. For example, the (GGGGT)<sub>6</sub> motif found in the *distal-less homeobox 3* (*DLX3*) gene was hit by both STRs and PQSs searches. The motifs that comprised of undisrupted poly-G were not counted as PQS motifs.

### Motifs discovery by the MEME Suite

The MEME Suite was used to find sequence motifs representing features such as DNA binding sites and protein interaction domains on the promoter regions [30]. The promoter sequences were divided into 100 bp bin to create query files. MEME has a large number of optional inputs to fine-tune its performance. The following options were used: (i) zero or one occurrence per sequence model (i.e., zoops) was chosen, (ii) the maximum width of the motifs was 15, (iii) motifs occurrences were on the given

DNA strand or on its reverse complement (i.e., revcomp), and (iv) the number of motifs was set to five. The probability reported by MEME is actually an approximation of the E-value of the log likelihood ratio, and the width of the motifs can also affect the statistical significance of the motifs. Thus, in general, motifs with a longer width tend to have lower levels of E-value in MEME analysis.

### Gene ontology analysis

The analysis of functional gene annotations was performed using the DAVID ver. 6.7 available on website [35,72]. We sorted chicken promoter sequences into four subgroups based on presence or absence of the aforementioned target motifs. The sequence motifs that shared characteristics with both PQSs and STRs were sorted into the STR. FET  $p$ -values were calculated to estimate the level of over-representation of the selected genes in GO categories [73], especially in the biological process. Probabilities less than 0.01 were used as cut-off value and considered to show significant level of correlation. Heat map analysis was also conducted through DAVID outcomes to visualize a matrix of enriched GO. R software ver. 3.0.2 was used to create heat map of the significances.

### Additional files

**Additional file 1: GC content of each chicken promoter sequence.**

The GC content of each promoter shown with the overall average GC content (horizontal bar).

**Additional file 2: The observed/expected (O/E) CpG ratio in the chicken promoter.** The maximum O/E CpG ratio plotted against the number of genes.**Additional file 3: Co-occurrence of CpG island and sequence motifs in a single promoter.** The number of chicken genes was statistically examined whether it showed significant excess of co-occurring motif pairs.**Additional file 4: A full list of genes found to be correlated with sequence motifs.** GO biological process categories enriched amongst genes, either of which has a sequence motif within a promoter.

### Abbreviations

bp: Base pairs; CGI: CpG island; DAVID: Database for annotation, visualization, and integrated discovery; EBI: European bioinformatics institute; FET: Fisher's exact test; G4: G-quadruplex; GO: Gene ontology; LCGI: Long CpG island; NCGI: No CpG island; O/E: Observed/expected; PQS: Potential G-quadruplex sequence; QGRS: Quadruplex forming G-rich sequences; STARRRT: Short tandem repeats in regulatory region table; STR: Short tandem repeat; TF: Transcription factor; TFBS: Transcription factor binding site; TSS: Transcription start site; UCSC: University of California, Santa Cruz; UTR: Untranslated region.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

HA conceived this study. He did the complete data analysis and wrote the first draft of the manuscript. NJG supervised the project and contributed to the data interpretation and writing of the manuscript. Both authors read and approved the final manuscript.

### Acknowledgements

This work was supported by FY2013 Researcher Exchange Program between the Japan Society for the Promotion of Science (JSPS) and the Royal Society of New Zealand (RSNZ). Authors would like to thank Sterling Sawaya for insightful suggestions on the research strategy and comments on an earlier version of this manuscript. We are also grateful for the comments made by anonymous reviewers that have helped further improve the manuscript.

### Author details

<sup>1</sup>Department of Anatomy, University of Otago, Dunedin, New Zealand. <sup>2</sup>Allan Wilson Centre for Molecular Ecology and Evolution, University of Otago, Dunedin, New Zealand.

Received: 4 April 2014 Accepted: 8 October 2014

Published: 15 October 2014

### References

1. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA: **The evolution of transcriptional regulation in eukaryotes.** *Mol Biol Evol* 2003, **20**(9):1377–1419.
2. Halees AS: **PromoSer: a large-scale mammalian promoter and transcription start site identification service.** *Nucleic Acids Res* 2003, **31**(13):3554–3559.
3. Wittkopp PJ, Kalay G: **Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence.** *Nat Rev Genet* 2012, **13**(1):59–69.
4. Gemayel R, Vences MD, Legendre M, Verstrepen KJ: **Variable tandem repeats accelerate evolution of coding and regulatory sequences.** *Annu Rev Genet* 2010, **44**:445–477.
5. Horton BM, Hudson WH, Ortlund EA, Shirk S, Thomas JW, Young ER, Zinzow-Kramer WM, Maney DL: **Estrogen receptor alpha polymorphism in a species with alternative behavioral phenotypes.** *Proc Natl Acad Sci U S A* 2014, **111**(4):1443–1448.
6. Andersson L: **Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci.** *Genetica* 2009, **136**(2):341–349.
7. Xu Y, He B, Li R, Pan Y, Gao T, Deng Q, Sun H, Song G, Wang S: **Association of the polymorphisms in the Fas/FasL promoter regions with cancer susceptibility: a systematic review and meta-analysis of 52 studies.** *PLoS One* 2014, **9**(3):e90090.
8. Vences MD, Legendre M, Caldara M, Hagihara M, Verstrepen KJ: **Unstable tandem repeats in promoters confer transcriptional evolvability.** *Science* 2009, **324**(5931):1213–1216.
9. Bolton KA, Ross JP, Grice DM, Bowden NA, Holliday EG, Avery-Kiejda KA, Scott RJ: **STARR: a table of short tandem repeats in regulatory regions of the human genome.** *BMC Genomics* 2013, **14**(1):795.
10. Valipour E, Kowsari A, Bayat H, Banan M, Kazeminasab S, Mohammadparast S, Ohadi M: **Polymorphic core promoter GA-repeats alter gene expression of the early embryonic developmental genes.** *Gene* 2013, **531**(2):175–179.
11. Morris EE, Amria MY, Kistner-Griffin E, Svenson JL, Kamen DL, Gilkeson GS, Nowling TK: **A GA microsatellite in the Fli1 promoter modulates gene expression and is associated with systemic lupus erythematosus patients without nephritis.** *Arthritis Res Ther* 2010, **12**(6):R212.
12. Luo W, Gangwal K, Sankar S, Boucher KM, Thomas D, Lessnick SL: **GSTM4 is a microsatellite-containing EWS/FLI target involved in Ewing's sarcoma oncogenesis and therapeutic resistance.** *Oncogene* 2009, **28**(46):4126–4132.
13. Kovar H: **Downstream EWS/FLI1 - upstream Ewing's sarcoma.** *Genome Med* 2010, **2**:8.
14. Guillon N, Tirode F, Boeva V, Zynovyev A, Barillot E, Delattre O: **The oncogenic EWS-FLI1 protein binds *in vivo* GGAA microsatellite sequences with potential transcriptional activation function.** *PLoS One* 2009, **4**(3):e4932.
15. Sawaya S, Bagshaw A, Buschiazio E, Kumar P, Chowdhury S, Black MA, Gemmell N: **Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements.** *PLoS One* 2013, **8**(2):e54710.
16. Lipps HJ, Rhodes D: **G-quadruplex structures: *in vivo* evidence and function.** *Trends Cell Biol* 2009, **19**(8):414–422.
17. Zhang C, Liu HH, Zheng KW, Hao YH, Tan Z: **DNA G-quadruplex formation in response to remote downstream transcription activity: long-range sensing and signal transducing in DNA double helix.** *Nucleic Acids Res* 2013, **41**(14):7144–7152.
18. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH: **Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription.** *Proc Natl Acad Sci U S A* 2002, **99**(18):11593–11598.
19. Ambrus A, Chen D, Dai J, Jones RA, Yang D: **Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization.** *Biochemistry* 2005, **44**(6):2048–2058.
20. Akan P, Deloukas P: **DNA sequence and structural properties as predictors of human and mouse promoters.** *Gene* 2008, **410**(1):165–176.
21. Elango N, Yi SV: **Functional relevance of CpG island length for regulation of gene expression.** *Genetics* 2011, **187**(4):1077–1083.
22. Sharif J, Endo TA, Toyoda T, Koseki H: **Divergence of CpG island promoters: a consequence or cause of evolution?** *Dev Growth Differ* 2010, **52**(6):545–554.
23. Robertson KD: **DNA methylation and chromatin - unraveling the tangled web.** *Oncogene* 2002, **21**(35):5361–5379.
24. Li Q, Li N, Hu X, Li J, Du Z, Chen L, Yin G, Duan J, Zhang H, Zhao Y, Wang J, Li N: **Genome-wide mapping of DNA methylation in chicken.** *PLoS One* 2011, **6**(5):e19428.
25. Vavouri T, Lehner B: **Human genes with CpG island promoters have a distinct transcription-associated chromatin organization.** *Genome Biol* 2012, **13**(11):R110.
26. Furlong RF: **Insights into vertebrate evolution from the chicken genome sequence.** *Genome Biol* 2005, **6**(2):207.
27. Rao YS, Chai XW, Wang ZF, Nie QH, Zhang XQ: **Impact of GC content on gene expression pattern in chicken.** *Genet Sel Evol* 2013, **45**:9.
28. Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grutzner F, Belov K, Miller W, Clarke L, Chinwalla AT, Yang SP, Heger A, Locke DP, Miethke P, Waters PD, Veyrunes F, Fulton L, Fulton B, Graves T, Wallis J, Puente XS, Lopez-Otin C, Ordóñez GR, Eichler EE, Chen L, Cheng Z, Deakin JE, Alsop A, Thompson K, Kirby P, et al: **Genome analysis of the platypus reveals unique signatures of evolution.** *Nature* 2008, **453**(7192):175–183.
29. Du Z, Kong P, Gao Y, Li N: **Enrichment of G4 DNA motif in transcriptional regulatory region of chicken genome.** *Biochem Biophys Res Commun* 2007, **354**(4):1067–1070.
30. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS: **MEME SUITE: tools for motif discovery and searching.** *Nucleic Acids Res* 2009, **37**(Web Server issue):W202–W208.
31. Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E, Yusuf D, Lenhard B, Wasserman WW, Sandelin A: **JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles.** *Nucleic Acids Res* 2010, **38**(Database issue):D105–D110.
32. Newburger DE, Bulyk ML: **UniPROBE: an online database of protein binding microarray data on protein-DNA interactions.** *Nucleic Acids Res* 2009, **37**(Database issue):D77–D82.
33. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS: **Quantifying similarity between motifs.** *Genome Biol* 2007, **8**(2):R24.
34. Carrinci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Amer P, Chesi A, Gustinich S, Persichetti F, et al: **Genome-wide analysis of mammalian promoter architecture and evolution.** *Nat Genet* 2006, **38**(6):626–635.
35. da Huang W, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic Acids Res* 2009, **37**(1):1–13.
36. Lawson MJ, Zhang L: **Housekeeping and tissue-specific genes differ in simple sequence repeats in the 5'-UTR region.** *Gene* 2008, **407**(1–2):54–62.
37. Sawaya SM, Bagshaw AT, Buschiazio E, Gemmell NJ: **Promoter microsatellites as modulators of human gene expression.** In *Tandem Repeat Polymorphisms: Genetic Plasticity, Neural Diversity and Disease*. Edited by Hannan AJ. Austin/New York: Landes Bioscience and Springer Science + Business Media; 2012:41–54.
38. Rich A, Nordheim A, Wang AH: **The chemistry and biology of left-handed Z-DNA.** *Annu Rev Plant Physiol Plant Mol Biol* 1984, **53**:791–846.
39. Bayele HK, Peyssonnaud C, Giatromanolaki A, Arrais-Silva WW, Mohamed HS, Collins H, Giorgio S, Koukourakis M, Johnson RS, Blackwell JM, Nizet V, Srai

- SK: HIF-1 regulates heritable variation and allele expression phenotypes of the macrophage immune response gene SLC11A1 from a Z-DNA forming microsatellite. *Blood* 2007, **110**(8):3039–3048.
40. Bacolla A, Larson JE, Collins JR, Li J, Milosavljevic A, Stenson PD, Cooper DN, Wells RD: **Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties.** *Genome Res* 2008, **18**(10):1545–1553.
41. Martin P, Makepeace K, Hill SA, Hood DW, Moxon ER: **Microsatellite instability regulates transcription factor binding and gene expression.** *Proc Natl Acad Sci U S A* 2005, **102**(10):3800–3804.
42. Zhou W, Chen Z, Hu W, Shen M, Zhang X, Li C, Wen Z, Wu X, Hu Y, Zhang X, Duan X, Han X, Tao Z: **Association of short tandem repeat polymorphism in the promoter of prostate cancer antigen 3 gene with the risk of prostate cancer.** *PLoS One* 2011, **6**(5):e20378.
43. Primmer CR, Raudsepp T, Chowdhary BP, Moller AP, Ellegren H: **Low frequency of microsatellites in the avian genome.** *Genome Res* 1997, **7**(5):471–482.
44. Mayer C, Leese F, Tollrian R: **Genome-wide analysis of tandem repeats in daphnia pulex—a comparative approach.** *BMC Genomics* 2010, **11**:277.
45. Kruglyak S, Durrett RT, Schug MD, Aquadro CF: **Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations.** *Proc Natl Acad Sci U S A* 1998, **95**(18):10774–10778.
46. McQueen HA, Fantes J, Cross SH, Clark VH, Archibald AL, Bird AP: **CpG islands of chicken are concentrated on microchromosomes.** *Nat Genet* 1996, **12**(3):321–324.
47. Hughes AL, Hughes MK: **Small genomes for better flyers.** *Nature* 1995, **377**(6548):391.
48. Vardhanabhuti S, Wang J, Hannehalli S: **Position and distance specificity are important determinants of cis-regulatory motifs in addition to evolutionary conservation.** *Nucleic Acids Res* 2007, **35**(10):3203–3213.
49. Moses AM, Chiang DY, Kellis M, Lander ES, Eisen MB: **Position specific variation in the rate of evolution in transcription factor binding sites.** *BMC Evol Biol* 2003, **3**:19.
50. Brandström M, Ellegren H: **Genome-wide analysis of microsatellite polymorphism in chicken circumventing the ascertainment bias.** *Genome Res* 2008, **18**(6):881–887.
51. Beilina A, Tassone F, Schwartz PH, Sahota P, Hagerman PJ: **Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element.** *Hum Mol Genet* 2004, **13**(5):543–549.
52. Tomita N, Fujita R, Kurihara D, Shindo H, Wells RD, Shimizu M: **Effects of triplet repeat sequences on nucleosome positioning and gene expression in yeast minichromosomes.** *Nucleic Acids Res Suppl* 2002, **2**:231–232.
53. Bansal M, Kumar A, Yella VR: **Role of DNA sequence based structural features of promoters in transcription initiation and gene expression.** *Curr Opin Struct Biol* 2014, **25**:77–85.
54. Brukner I, Sanchez R, Suck D, Pongor S: **Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides.** *EMBO J* 1995, **14**(8):1812–1818.
55. Anderson JD, Widom J: **Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites.** *Mol Cell Biol* 2001, **21**(11):3830–3839.
56. Shimizu M, Mori T, Sakurai T, Shindo H: **Destabilization of nucleosomes by an unusual DNA conformation adopted by poly(dA) small middle dotpoly(dT) tracts in vivo.** *EMBO J* 2000, **19**(13):3358–3365.
57. Finkenzeller G, Sparacio A, Technau A, Marme D, Siemeister G: **Sp1 recognition sites in the proximal promoter of the human vascular endothelial growth factor gene are essential for platelet-derived growth factor-induced gene expression.** *Oncogene* 1997, **15**(6):669–676.
58. Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM: **Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome.** *Genome Res* 2006, **16**(1):1–10.
59. Deaton AM, Bird A: **CpG islands and the regulation of transcription.** *Genes Dev* 2011, **25**(10):1010–1022.
60. Chan ET, Quon GT, Chua G, Babak T, Trochesset M, Zirngibl RA, Aubin J, Ratcliffe MJ, Wilde A, Brudno M, Morris QD, Hughes TR: **Conservation of core gene expression in vertebrate tissues.** *J Biol* 2009, **8**(3):33.
61. Kumar P, Yadav VK, Baral A, Kumar P, Saha D, Chowdhury S: **Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide.** *Nucleic Acids Res* 2011, **39**(18):8005–8016.
62. 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: **Ensembl variation resources.** *BMC Genomics* 2010, **11**:293.
63. **The UCSC genome browser.** <http://genome.ucsc.edu>.
64. International Chicken Genome Sequencing Consortium (ICGSC): **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.** *Nature* 2004, **432**(7018):695–716.
65. Medvedeva YA, Fridman MV, Oparina NJ, Malko DB, Ermakova EO, Kulakovskiy IV, Heinzel A, Makeev VJ: **Intergenic, gene terminal, and intragenic CpG islands in the human genome.** *BMC Genomics* 2010, **11**:48.
66. **The European bioinformatics institute.** <http://www.ebi.ac.uk>.
67. Gardiner-Garden M, Frommer M: **CpG islands in vertebrate genomes.** *J Mol Biol* 1987, **196**(2):261–282.
68. Martins WS, Lucas DCS, Neves KFS, Bertoli DJ: **WebSat—a web software for microsatellite marker development.** *Bioinformatics* 2009, **3**(6):282–283.
69. **Phobos 3.3.11.** [http://www.ruhr-uni-bochum.de/spezzoo/cm/cm\\_phobos.htm](http://www.ruhr-uni-bochum.de/spezzoo/cm/cm_phobos.htm).
70. Kraemer L, Beszteri B, Gabler-Schwarz S, Held C, Leese F, Mayer C, Pohlmann K, Frickenhaus S: **STAMP: Extensions to the STADEN sequence analysis package for high throughput interactive microsatellite marker design.** *BMC Bioinformatics* 2009, **10**:41.
71. Kikin O, D'Antonio L, Bagga PS: **QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences.** *Nucleic Acids Res* 2006, **34**(Web Server issue):W676–W682.
72. da Huang W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**(1):44–57.
73. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The gene ontology consortium.** *Nat Genet* 2000, **25**(1):25–29.

doi:10.1186/1471-2164-15-900

**Cite this article as:** Abe and Gemmell: **Abundance, arrangement, and function of sequence motifs in the chicken promoters.** *BMC Genomics* 2014 **15**:900.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

