The Fourth International Conference on Computational Systems Biology (ISB2010)
Suzhou, China, September 9–11, 2010

Genome-wide Analysis of the Transcription Factor Binding Preference of Human Bidirectional Promoters and Functional Annotation of the Related Gene Pairs

Bingchuan Liu[†1,2], Jiajia Chen[†1,3], Bairong Shen[1,*]

1 Center for Systems Biology, Soochow University, Suzhou, 215006, China
2 School of life Science and Technology, Tongji University, Shanghai, China
3 School of Chemistry and Biological Engineering, Suzhou University of Science and Technology, 215009, China

Abstract Bidirectional gene pairs have received considerable attention for their prevalence in vertebrate genomes. However, their biological relevance and exact regulatory mechanism remain less understood. To study the inner properties of this gene organization and the difference between bi- and unidirectional genes, we conducted a genome-wide investigation in terms of the promoter sequence analysis, functional association and regulation motif discovery. 1210 bidirectional gene pairs were identified based on the GRCh37 human genome assembly data. CpG islands were detected in 98.42% bidirectional promoters by intergenetic promoter analysis. Functional enrichment analysis revealed that bidirectional genes tend to be associated with housekeeping functions in metabolism pathways and nuclear processes, and pair members tend to be involved in the same biological function. Furthermore, a distinct collection of putative transcription factor binding sites that preferentially occurs in bi-directional promoters were determined by overrepresentation analysis.

Keywords bidirectional promoter; unidirectional gene; CpG island; functional enrichment; Transcription Factor Biding Site(TFBS)

1 Introduction

According to the orientation and status of the 5’ end, the adjacently located genes can be arranged in convergent, divergent, tandem, anti-sense or interleaving configuration[1]. Among these categories, the divergent gene arrangement is a common architectural feature of the human genome, accounting for about 10% of all genes[2]. Bi-directional gene pairs is defined as two genes arranged in a head-to-head (adjacent 5’ ends) fashion on opposite strands of DNA with less than 1,000 bp between their transcription start sites(TSS)[1]. Accordingly, the entire intervening region between the two TSS of the gene pair is designated as a putative

†Equal contributors; *Corresponding author: bairong.shen@suda.edu.cn

bi-directional promoter. A gene is termed as uni-directional if no oppositely oriented TSS is found within 10kb upstream of the given TSS, or if a similarly oriented TSS is found at least 1kb upstream. Thus the entire 1kb of 5' flanking DNA is considered as the unidirectional promoter.

Examples of bi-directional genes including LRRC49/THAP10[3], SURF-1/SURF-2[4], COL4A1/COL4A2[5], PCD10/SERPINI1[6] and HAND2/DEIN[7] have been identified in human through individual experiments. Whereas most of these bi-directional gene pairs have been found in the process of studying a single gene, a genome-wide analysis of bi- and unidirectional genes on latest human genome assembly, especially in terms of functional association, is still insufficient.

A considerable number of bi-directional gene pairs were found to be conserved among mammalian species[8, 9]. Since evolutionary conservation usually indicates functional implications, we proposed that the conservation of head-to-head gene organization is under selection to fulfill a specific functional role. Nevertheless, evidence supporting the function and physiologic consequences of this gene organization is currently insufficient.

The expression data obtained from biotechnologies such as SAGE and microarray indicated a correlated expression profile between bi-directional genes[10-12]. Based on the assumption that 'co-expression implies co-regulation', the requirement for co-regulation of functionally related genes appears to underlie observed coexpression. However, it is still under discussion whether the coexpression of two genes evolved merely as a consequence of their physical proximity or if function dictated their co-regulation. There are several examples of bi-directional gene pairs that are related by function, e.g. in DNA repair[1, 2], aging[13], de novo purine synthesis[14] and carcinogenesis[4]. Despite this observation, a systematic study on the degree of internal co-function of the bi-directional genes has not been carried out to date.

More recent studies on a number of bi-directional gene pairs have shown that most bidirectional promoters lack TATA boxes and are enriched in G+C content and CpG islands[1, 2, 12]. This characteristic feature led us to hypothesize that divergent genes will be transcribed with a unique set of regulatory signals. Currently our understanding of transcription regulation relies greatly on experimental identification of prospective regulatory regions. However, relatively few studies have addressed main regulatory elements in bi-directional promoters specifically. Therefore, it seems necessary to reevaluate the underlying mechanisms and biological relevance of bidirectional promoters systematically.

In the present study, we have undertaken a genome-wide survey of gene organization in the human genome. To reveal functions collectively performed by such bi-directional genes, we mapped them to the Gene Ontology (GO) and GeneGO pathways. We also explored the functional similarity and the difference between the genes on the plus and minus strand. We devoted our effort into exploring the preference of transcription factor(TF) binding on the bi- and unidirectional promoters and statistically identified a set of over-represented transcription factor binding sites(TFBS) in bidirectional promoters, the research scheme is shown in figure 1.

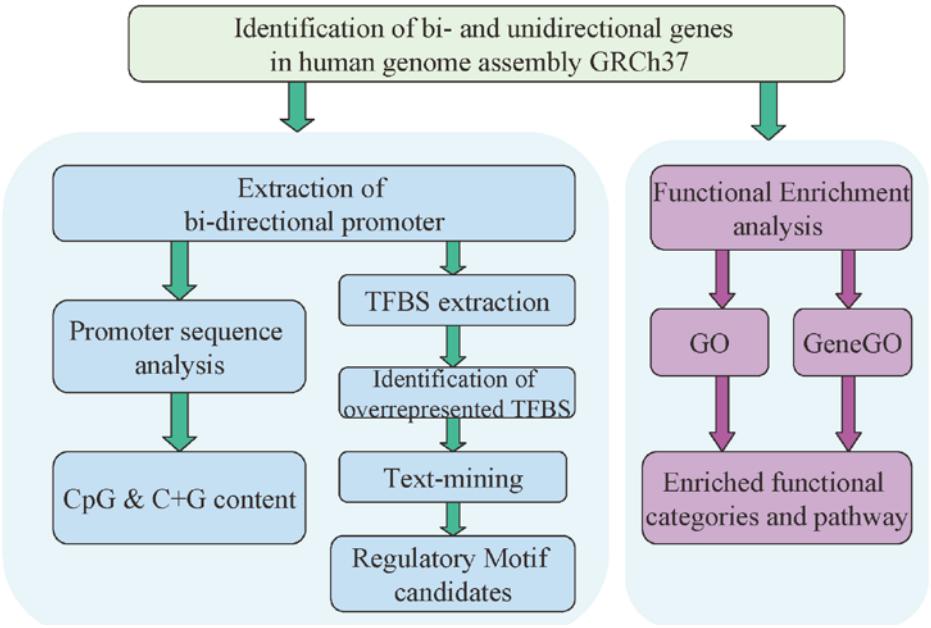

Figure 1 Diagrammatic Representation of the Research Scheme

2 Materials and Methods

2.1 Identification of bidirectional and unidirectional genes in human genome

Human genome assembly GRCh37 was downloaded from Genome Reference Consortium (ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/vertebrates_ma mmals/Homo_sapiens/GRCh37/). The gene annotation (NCBI Build36) was retrieved from the NCBI Entrez Gene ftp site (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/). The transcript mapping information was downloaded from hg19 RefGene table from UCSC Genome Browser (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/). A total of 45,408 genes (excluding mitochondrial genome) and 31,357 transcripts were collected and filtered for redundancy. This resulted in 44,293 non-redundant items of RefSeqs transcripts. Genes without clear mRNA information (NR, XR, XM) were filtered to ensure the exact transcription of all the genes. The 28520 mRNAs were collapsed into 21757 unique and non-overlapping clusters. Discrimination of bidirectional gene pairs and unidirectional genes was performed by a perl script according to the definition by Trinklein et al [2]. Redundant gene pair entries that share the same IG sequence were removed.

2.2 Extraction of bi-directional promoter region

The intergenic regions between bidirectional genes' TSS were taken as bi-directional promoters. For unidirectional genes the region of 1000 bp upstream of

the TSS was extracted as promoter. Promoter regions were extracted from the chromosome fasta files of the latest GRCh37 version genome assembly datasets. (ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/vertebrates_mammals/Homo_sapiens/GRCh37/Primary_Assembly/assembled_chromosomes/FASTA/).

2.3 Analysis of Promoter Sequences

The intergenic sequences of bidirectional genes were extended in both sides symmetrically into 1000 bp to meet the definition of a CpG island length. CpG island finder script[15] was run with two types of parameter criteria, %GC>=50, Obs/Exp>=0.60, length 500 and %GC >=55, Obs/Exp>=0.60, length 500 respectively. CpG frequency within both the bi-directional and non bi-directional promoters was calculated.

2.4 Evaluation of Functional Enrichment

We utilize Gene Ontology (GO) categories (http://www.geneontology.org/) and a commercial software MetaCore-GeneGO Pathway Maps (http://www.genego.com/metacore.php) to group functionally related genes and to contrast the functional distribution of bidirectional genes to the average distribution in the whole genome. The analysis of overrepresented GO terms for bi-directional genes was performed by GOEAST[16]. Statistical enrichment of a category was quantified using the Hypergeometric test method. Yekutieli multi-test adjustment method was applied to correct for multiple testing.

Genes were then mapped to GeneGO database by MetaCore™ tools to infer pathways preferentially targeted by bidirectional genes. In MetaCore™, the statistical significance of the enriched pathways was indicated by a p-value yielded from the Fisher's exact test. The False discovery rate (FDR) was applied to correct for multiple testing.

2.5 Discovery of over-represented transcription factor binding sites

Putative TFBS in promoter regions were searched for matches to the position-weight matrix(PWM) in the JASPAR[17, 18] and TRANSFAC[19] database. Predetermined PWMs for 73 and 87 vertebrate TFBSs were extracted from TRANSFAC(public version 7.0) and JASPAR PSSM respectively. Alignment of PWMs on genomic sequence was performed with COTRASIF[20] (http://biomed.org.ua/COTRASIF/). TFBSs within bi-directional promoters were categorized as over-represented, shared or under-represented at 2-fold threshold. A total of 18840 unidirectional promoters were used to give a contrast of bi-directional genes.

3 Results

3.1 Identification of bidirectional and unidirectional genes

We identified 1210 bidirectional gene pairs based on the curated transcript cluster NMs and NRs, accounting for 11.67% of all the genes owning RNAs. The number

was slightly larger than previous report[1, 2] as a result of updated gene annotations. Our work focus on the pure mRNA gene pairs and a large part of non-coding RNA, transcribed RNA and miscRNA are excluded from further analysis. If only transcripts with conclusive mRNA were reserved, 878 bidirectional gene pairs(9.31%) were discovered upon the removal of pairs consisting of NMs and NRs. Redundant gene pair entries that share the same IG sequence were removed to yield 822 bidirectional gene pairs for the analyses.

3.2 CpG islands are preferentially located in bidirectional promoters

There have been two contradictory observations on the CpG island frequency in bi-directional promoters. Adachi and Lieber[1] considered the presence of a CpG island to be a common feature of bidirectional promoters. In contrast, Takai et al. [12] reported that CpG islands are not preferentially associated with bi-directional promoters. They attributed the discrepancy to the different criteria adopted to define a CpG island. Therefore, in order to rationalize these controversial observations, we performed genome-wide computational analysis of the bi-directional promoters on the basis of two different definition systems. According to traditional definition by Gardiner-Garden[21], CpG islands were detected in 809 bi-directional promoters, representing 98.42% of a total of 822. A lower percentage of 61.07% was recorded for uni-directional promoters. Based on more strict criteria [22] (DNA fragment no less than 500bp with GC-content >= 55% and Obs/Exp value >=0.60), CpG-islands were present in 86.37% of bidirectional promoters compared to 28.48% of uni-directional promoters. In addition, we analyzed pure IG sequence to remove the difference caused by the extended IG region. Invariably the frequency of CpG island in bi-directional promoters is higher than those in uni-directional ones. As shown in Figure 2, the CpG density in bidirectional promoters (histogram in top left) is significantly higher than that in unidirectional promoter (histogram in top right) in all comparisons. Consistent with a significant enrichment of CpG-islands, bidirectional promoters feature a high C+G content (histogram in bottom left and right). This suggests an intrinsic difference in nucleotide composition of bi-directional promoters compared to unidirectional backgrounds.

3.3 Functional Enrichment of Bidirectional Genes

3.3.1 Gene ontology associated with bidirectional promoter regulation

Genes regulated by bi-directional promoters were examined for functional classifications and associations. Among the 1,644 genes involved in the 822 human bi-directional gene pairs, 1,121, 1,219, and 1,256 genes were directly annotated by ‘biological process’, ‘molecular function’ and ‘cellular component’ subcategories in GO annotation system, respectively. We found several GO classes significantly overrepresented among bi-directional genes. Cellular, metabolic and biosynthetic processes emerged as the most significantly enriched functional class. GO items of cell cycle and its child nodes were also significantly presented. Cellular response to stress or stimulus and their related subclasses of damage response, break repair were also focused. To summarize, the most enriched GO categories correspond to the

known physiological roles of the cell, indicating that bi-directional genes are frequently involved in basic cellular metabolic processes. See supplementary TableS1.xls for the list of enriched GO terms.

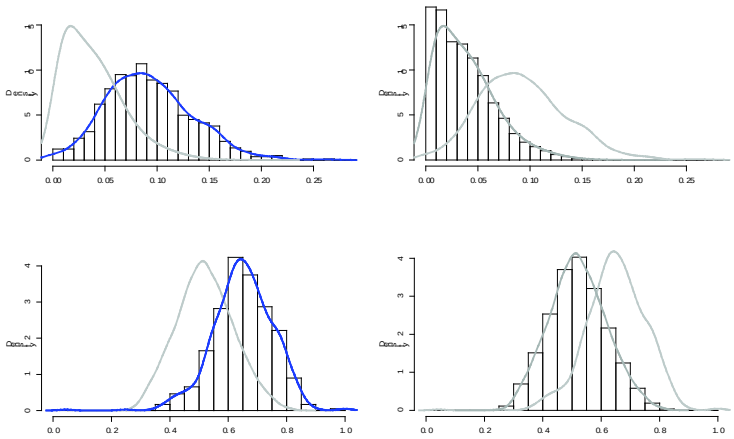

Figure2 - Density Distribution of CpG Islands between Bi-directional Genes and Unidirectional Genes (The figure is reproduced with permission from the rights owner Liu,B. [23]**)**

3.3.2 Functional similarities for annotated bidirectional gene pairs

Among 822 annotated bi-directional gene pairs, we found 385 pairs (46.84%) sharing at least one GO annotation. Such shared or related function supports the hypothesis that bi-directional genes are more likely to be functionally associated than unidirectional genes. We also provided separate estimates for each of the Gene Ontologies. We obtained 337 annotated pairs in subcategory "cellular component", 185 pairs in "molecular function" and 146 pairs in "biological process" respectively. It's observed that, in general, head-to-head gene products are more likely to perform coordinated roles in the same cellular component, compared to the other two subsystems.

Then we set out to find out the GO terms that represent coordinated functions of bi-directional gene pairs. In Biological Process, the GO terms related to metabolic process and its branch such as primary metabolic process, cellular process and biopolymer biosynthetic process topped the list of both gene pair members. Their child nodes were focused on RNA (mRNA, ncRNA) metabolic process, cellular (macromolecule or biopolymer) catabolic process, organelle organization, mitotic cell cycle etc. In molecular function, the GO terms involved in DNA-directed RNA polymerase activity, RNA methyltransferase activity, purine NTP-dependent helicase activity, NAD or NADH binding, NADH dehydrogenase (quinone) activity, etc. are significantly overrepresented as compared to others. In Cellular Component, we found that bi-directional genes tend to be tightly associated in the same class of organelle, organelle envelope, nucleus, nucleoplasm, nucleolus, membrane-bounded

or non-membrane-bounded organelle, etc. Interestingly, almost all the items shared by the two divergent genes are related to metabolism and energy transfer. We proposed that genes involved in functions including metabolism, are more likely to be organized in the head-to head configuration.

3.3.3 GeneGO pathway enrichment

On the base of p-values from MetaCore™, totally we found 45 pathways that are significantly enriched with the divergent genes out of the total 451 distinct pathways. According to the different classification criterion, the 45 pathways were assigned to 18 regulatory processes, 8 protein function, 4 disease maps and 15 metabolic maps. Extreme enrichment occurred for, in order of descending significance level, NHEJ mechanisms of DSBs repair, Oxidative phosphorylation, Nucleotide excision repair and GTP-XTP metabolism, Chromosome condensation in prometaphase, Role of Brca1 and Brca2 in DNA repair. Enriched pathways are further cluster into larger functional categories according to the GeneGO annotation. Regulatory processes/Cell cycle and Regulatory processes/DNA-damage ranked among the top enriched functional categories. Table 1 lists some most enriched categories ordered in decreasing level of significance.

Table 1 - Statistically enriched GeneGO Pathway categories

Pathway category	p-value
Regulatory processes/Cell cycle	5.35E-09
Regulatory processes/DNA-damage	2.07E-08
Metabolic maps/Metabolic maps (common pathways)/Energy metabolism	1.12E-06
Metabolic maps/Metabolic maps (common pathways)	1.03E-04
Metabolic maps/Metabolic maps (common pathways)/Nucleotide metabolism	6.96E-04
Metabolic maps/Metabolic maps (common pathways)/Vitamin and cofactor metabolism	5.71E-03

3.4 Bi-directional promoters are characterized by a distinct collection of putative transcription factor binding sites

We characterized the enrichment of known motifs from TRANSFAC and JASPAR in bidirectional promoters relative to background unidirectional promoters. Based on the Jaspar PSSM information, we categorized 43 transcription factors as overrepresented (AR, ARNT, BRCA1, CREB1, E2F1, ELF5, ELK1, ELK4, EN1, ESR1, ETS1, GABPA, Gfi, HINFP, HLF, HNF4A, MAFB, MAX, MYB, MYCN, Myf[N], MZF1(1-4), MZF1(5-13), NFKB1, NFYA, NHLH1, PAX2, PAX5, PAX6, PBX1, REL, REST, Roaz, SOX17, SOX9, SP1, SPI1, SPIB, SPZ1, TFAP2A, USF1, ZNF143, ZNF354C), 18 as shared (NKX2-5, NKX3-1, NKX3-2, NOBOX, NR2F1, NR3C1, PAX4, PDX1, PRRX2, RELA, RORA, RUNX1, SOX5, SRY, STAT1, T box, YY1, ZEB1), and 6 as underrepresented (TP53, TEAD1, TBP, SRF, RREB1, PPARG).

In the TRANSFAC database, 73 TFBSs such as ARNT, ATF1, ATF2, ATF6, BACH2, CCAAT box, CREB1, E2F[N], EGR1, EGR2, EGR3, EGR4, ELK1, EP300, ESR1, ETS1, GABPA, GBP[N], GC box, HMX3, HNF4A, HSF1, HSF2, IKZF1, IKZF2, JUN(v-), KLF12, LGALS4, LHX1, MAF(v-), MAX, MSC, MYB(v-), MYCN, MYOG, MZF1, NF1, NFE2, NFE2L1, NFIL3, NFKB1, OR5I1, P53, PATZ1, PAX3, PAX4, PAX5, PAX6, PPARA, REL, RELA, repressor of CAR1 expression, RFX1, RUNX1, SLC25A4, SOX9, SP1, SPZ1, SREBF1, STAT[x], STAT3, STAT4, STAT5A, STAT5B, TFAP2A, TFAP2C, TFAP4, XBP1, ZBTB6, ZIC1, ZIC2, ZIC3, ZNF143, find increased presence in bidirectional promoters.

Although there is slight difference between the two databases, a large majority of the TFBSs overlap. We further investigated the experimental evidence supporting the roles of these transcriptional factors in regulating bi-directional genes. Table 2 lists the experimentally validated TFBS that occurred in bi-directional promoters.

Table 2 - The experimentally validated TFBS that occurred in bi-directional promoters

TF name	Fold Enrichment	Regulated gene pair	Reference
GABPA	7.069	Gapba/Atp5j PREPL-C2ORF34	[24] [25]
E2F1	6.893	TK/KF genes	[9]
NFY	5.255	Mrps12/Sarsm PREPL-C2ORF34 Mrps12/Sars2	[26] [25] [27]
SP1	3.398	OSGEP/APEX Gapba/Atp5j	[28] [24]
CCAAT box	2.687	DEIN/HAND2 HSF-1/Bop1 E14/ATM BRCA1/NBR2 GPAT/AIRC OSGEP/APEX mOsgep/mApex	[7] [29] [30] [31] [32] [33] [34]
NF1	2.591	Pxmp2/PoleI	[35]

Some of the reported physiological functions are consistent with our functional enrichment analysis. For example, previous work[24] has demonstrated that GABPA regulates genes involved in a variety of cellular processes including adipocyte differentiation, mitochondrial respiration, and neuromuscular signaling, corresponding to enriched GO terms of cell cycle, cellular and metabolic processes and their child nodes. E2F1 are observed to regulate cell growth during the G0/G1-S phase transition, and over-expression of E2F1 induces apoptosis and DNA synthesis in quiescent fibroblasts [36]. These are in agreement with the significantly enriched GeneGo pathways such as Regulatory processes/Cell cycle and Regulatory

processes/DNA-damage.

Interestingly, the overrepresented recognition sequence for MYC, ELK1, NF-Y, SP1, ATF, GABPA, SREBP-1, NF-E2, STAT5A, NF-1 as well as SOX-9 rank among the most conserved motifs found in human promoters[37].

Given the enrichment of these motifs in bi-directional promoters and its strong evolutionary conservation across mammalian promoters, we assume that the predicted TFBSs located within bi-directional promoters are more likely to be functional in co-regulation than other TFBSs. Interestingly it would appear that TF within the same family tend to have similar binding preference. A TFBS is either overrepresented or underrepresented in parallel with other family members. These observations suggest a common mode of expression across the family members of transcription factors.

4 Discussion

In this study, 11.6 % of the human genes were shown to be arranged in a head-to-head fashion, and this proportion is slightly larger than the previous report of 11%[2]. We attribute the inconsistency to the update of TSS coordinates during the accumulation of EST and mRNA evidence. We provided a solid evidence for the previous observation[1] that bidirectional promoters had a significant enrichment of CpG-islands as well as a high GC content. Since CpG island is usually the target of regulation by methylation, it may induce changes in chromatin structure that can confer either positive or negative effects on transcription. Misregulation of bidirectional promoters elicited by mutation or hypermethylation will simultaneously silence genes on both sides. Loss of their vital biological function well explains the role of bidirectional genes in the development of human diseases such as aging[13], brain disease[6] and oncogenesis [3].

Our study provides a comprehensive functional evaluation of bi-directional genes. Bi-directional genes are significantly enriched in housekeeping functions such as metabolism pathways and nuclear processes. Further analyses revealed that the significant functional categories are more likely to be shared by bi-directional gene pair members. This indicated that the bi-directional genes are strongly biased toward functional similarities and coordinated regulation. We postulate that for bi-directional genes involved in basic biological processes, coordinated regulation ensures their synchronized action and thus minimizes transcriptional error. In contrast, genes with less coordinated regulation may be involved in pathways that are more flexible in responding to environmental changes.

We compared the TFBSs between bi- and uni-directional promoters according to their rate of occurrence. We discovered several transcription factors that preferentially regulate bi-directional promoters. Some of the TFBSs matched well with experimentally determined ones and several novel binding motifs were also identified. These bi-directional gene associated motifs may be envisaged as the best candidates for functional regulatory elements. In addition, the motif search result could help identify novel genes, which is linked to a known gene via a bidirectional promoter. And these genes probably perform important conserved functions.

5 Conclusion

In this work, we conducted a systematic investigation of bi-directional gene organization focusing on sequence features, functional association and regulatory motif discovery. We confirmed known properties of bi-directional gene organization and also provided significant new observations. We found that bidirectional gene pairs show a higher probability to be functionally associated, formulating hypotheses that the requirement for co-regulation of functionally related genes is a possible cause for the observed co-expression of bidirectional genes. We also proposed that a special set of motifs in the bidirectional promoters play a role in transcriptional regulation of bi-directional genes. Our data also provide the putative target putative regulatory motifs for experimental studies to investigate how the expression of bi-directional genes pairs is regulated.

Acknowledges

This research work has been supported in part by the National 973 Program of China (No. 2007CB947002) and the National Nature Science Foundation of China (20872107)

References

[1] Adachi NandLieber MR. Bidirectional gene organization: a common architectural feature of the human genome. Cell, 2002, 109(7): 807-809.
[2] Trinklein ND, Aldred SF, Hartman SJ et al. An abundance of bidirectional promoters in the human genome. Genome Res, 2004, 14(1): 62-66.
[3] De Souza Santos E, De Bessa SA, Netto MM et al. Silencing of LRRC49 and THAP10 genes by bidirectional promoter hypermethylation is a frequent event in breast cancer. Int J Oncol, 2008, 33(1): 25-31.
[4] Lennard ACandFried M. The bidirectional promoter of the divergently transcribed mouse Surf-1 and Surf-2 genes. Mol Cell Biol, 1991, 11(3): 1281-1294.
[5] Pollner R, Schmidt C, Fischer G et al. Cooperative and competitive interactions of regulatory elements are involved in the control of divergent transcription of human Col4A1 and Col4A2 genes. FEBS Lett, 1997, 405(1): 31-36.
[6] Chen PY, Chang WS, Chou RH et al. Two non-homologous brain diseases-related genes, SERPINI1 and PDCD10, are tightly linked by an asymmetric bidirectional promoter in an evolutionarily conserved manner. BMC Mol Biol, 2007, 8: 2.
[7] Voth H, Oberthuer A, Simon T et al. Co-regulated expression of HAND2 and DEIN by a bidirectional promoter with asymmetrical activity in neuroblastoma. BMC Mol Biol, 2009, 10: 28.
[8] Lennard A, Gaston KandFried M. The Surf-1 and Surf-2 genes and their essential bidirectional promoter elements are conserved between mouse and human. DNA Cell Biol, 1994, 13(11): 1117-1126.
[9] Schuettengruber B, Doetzlhofer A, Kroboth K et al. Alternate activation of two divergently transcribed mouse genes from a bidirectional promoter is linked to changes in histone modification. J Biol Chem, 2003, 278(3): 1784-1793.
[10] Williams EJandBowles DJ. Coexpression of neighboring genes in the genome of Arabidopsis thaliana. Genome Res, 2004, 14(6): 1060-1067.
[11] Kruglyak SandTang H. Regulation of adjacent yeast genes. Trends Genet, 2000, 16(3):

109-111.
[12] Takai DandJones PA. Origins of bidirectional promoters: computational analyses of intergenic distance in the human genome. Mol Biol Evol, 2004, 21(3): 463-467.
[13] Bellizzi D, Dato S, Cavalcante P et al. Characterization of a bidirectional promoter shared between two human genes related to aging: SIRT3 and PSMD13. Genomics, 2007, 89(1): 143-150.
[14] Gavalas AandZalkin H. Analysis of the chicken GPAT/AIRC bidirectional promoter for de novo purine nucleotide synthesis. J Biol Chem, 1995, 270(5): 2403-2410.
[15] Sujuan Y, Asaithambi AandLiu Y. CpGIF: an algorithm for the identification of CpG islands. Bioinformation, 2008, 2(8): 335-338.
[16] Burbelo PD, Martin GRandYamada Y. Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer. Proc Natl Acad Sci U S A, 1988, 85(24): 9679-9682.
[17] Sandelin A, Alkema W, Engstrom P et al. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res, 2004, 32(Database issue): D91-94.
[18] Vlieghe D, Sandelin A, De Bleser PJ et al. A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. Nucleic Acids Res, 2006, 34(Database issue): D95-97.
[19] Matys V, Kel-Margoulis OV, Fricke E et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res, 2006, 34(Database issue): D108-110.
[20] Tokovenko B, Golda R, Protas O et al. COTRASIF: conservation-aided transcription-factor-binding site finder. Nucleic Acids Res, 2009, 37(7): e49.
[21] Gardiner-Garden MandFrommer M. CpG islands in vertebrate genomes. J Mol Biol, 1987, 196(2): 261-282.
[22] Takai DandJones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A, 2002, 99(6): 3740-3745.
[23] Liu B, Chen JandShen B. Computationally Comparative Analysis of CpG islands in Human Bidirectional and Unidirectional Promoters. IEEE Proc, 2010, (in press).
[24] Patton J, Block S, Coombs C et al. Identification of functional elements in the murine Gabp alpha/ATP synthase coupling factor 6 bi-directional promoter. Gene, 2006, 369: 35-44.
[25] Huang CC and Chang WS. Cooperation between NRF-2 and YY-1 transcription factors is essential for triggering the expression of the PREPL-C2ORF34 bidirectional gene pair. BMC Mol Biol, 2009, 10: 67.
[26] Zanotto E, Hakkinen A, Teku G et al. NF-Y influences directionality of transcription from the bidirectional Mrps12/Sarsm promoter in both mouse and human cells. Biochim Biophys Acta, 2009, 1789(5): 432-442.
[27] Lin JM, Collins PJ, Trinklein ND et al. Transcription factor binding and modified histones in human bidirectional promoters. Genome Res, 2007, 17(6): 818-827.
[28] Ikeda S, Ayabe H, Mori K et al. Identification of the functional elements in the bidirectional promoter of the mouse O-sialoglycoprotein endopeptidase and APEX nuclease genes. Biochem Biophys Res Commun, 2002, 296(4): 785-791.
[29] Zhang Y, Koushik S, Dai R et al. Structural organization and promoter analysis of murine heat shock transcription factor-1 gene. J Biol Chem, 1998, 273(49): 32514-32521.
[30] Byrd PJ, Cooper PR, Stankovic T et al. A gene transcribed from the bidirectional ATM promoter coding for a serine rich protein: amino acid sequence, structure and expression studies. Hum Mol Genet, 1996, 5(11): 1785-1791.
[31] Xu CF, Brown MA, Nicolai H et al. Isolation and characterisation of the NBR2 gene

which lies head to head with the human BRCA1 gene. Hum Mol Genet, 1997, 6(7): 1057-1062.
[32] Gavalas A, Dixon JE, Brayton KA et al. Coexpression of two closely linked avian genes for purine nucleotide synthesis from a bidirectional promoter. Mol Cell Biol, 1993, 13(8): 4784-4792.
[33] Seki Y, Ikeda S, Kiyohara H et al. Sequencing analysis of a putative human O-sialoglycoprotein endopeptidase gene (OSGEP) and analysis of a bidirectional promoter between the OSGEP and APEX genes. Gene, 2002, 285(1-2): 101-108.
[34] Ikeda S, Mochizuki A, Sarker AH et al. Identification of functional elements in the bidirectional promoter of the mouse Nthl1 and Tsc2 genes. Biochem Biophys Res Commun, 2000, 273(3): 1063-1068.
[35] Otte DM, Schwaab UandLuers GH. The Pxmp2 and PoleI genes are linked by a bidirectional promoter in an evolutionary conserved fashion. Gene, 2003, 313: 119-126.
[36] Kowalik TF, DeGregori J, Leone G et al. E2F1-specific induction of apoptosis and p53 accumulation, which is blocked by Mdm2. Cell Growth Differ, 1998, 9(2): 113-118.
[37] Xie X, Lu J, Kulbokas EJ et al. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature, 2005, 434(7031): 338-345.