Insect Biochemistry and Molecular Biology 110 (2019) 10–18

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

Synergism of open chromatin regions involved in regulating genes in *Bombyx mori*

Quan Zhang[a,b], Tingcai Cheng[b,c,*], Yueting Sun[b], Yi Wang[a,c], Tieshan Feng[b], Xiaohong Li[b], Lihaoyu Liu[b], Zhiqing Li[a,c], Chun Liu[b,c], Qingyou Xia[a,c,**], Huawei He[a,c,***]

[a] *Biological Science Research Center, Southwest University, Chongqing 400715, China*
[b] *State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400715, China*
[c] *Chongqing Key Laboratory of Sericultural Science, Chongqing Engineering and Technology Research Center for Novel Silk Materials, Southwest University, Chongqing 400715, China*

ARTICLE INFO

Keywords:
Silkworm
Formaldehyde-assisted isolation of regulatory elements-sequencing
Transcriptional regulation
Open chromatin regions
Synergism

ABSTRACT

The dynamic variability of transcription factors (TFs) and their binding sites makes it challenging to conduct genome-wide transcription regulation research. The silkworm *Bombyx mori*, which produces silk, is one of the most valuable model insects in the order Lepidoptera. The "opening" and "closing" of chromatin in different silk yield strains is associated with changes in silk production, making this insect a good model for studying the transcriptional regulation of genes. However, few studies have examined the open chromatin regions (OCRs) of silkworms, and studying OCR synergism and their function in silk production remains challenging. Here, we performed formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate OCRs from the silk glands of fifth-instar larvae of the DaZao and D872 strains. In total, 128,908 high confidence OCRs were identified and approximately 80% of OCRs were located in non-coding regions. OCRs upregulated adjacent genes and showed signal-dependent vulnerability to single-nucleotide polymorphisms. Mid- and low-signal OCRs were more likely to have single-nucleotide polymorphisms (SNP). Further, OCRs interacted with each other within a distance of 5 kb. We named the OCR interaction complex as the "cluster of related regions" (COREs). The functions of the CORE and its harbored OCRs showed some differences. Additionally, COREs enriched many silk protein synthesis-associated genes, some of which were upregulated. This study identified numerous high confidence regulation sites and synergistic regulatory modes of OCRs that affect adjacent genes. These results provide insight into silkworm transcriptional regulation and improve our understanding of *cis*-element cooperation.

1. Introduction

The silkworm *Bombyx mori* has been domesticated for more than five thousand years. Strains with higher silk production have been bred through decades of sericulture. Previous studies of genomic variation and gene expression have attempted to explore the molecular mechanism of silkworm domestication and traits affecting silk yield (Yu et al., 2011). Further, genomic resequencing of domesticated and wild silkworm identified approximately 16 million single-nucleotide polymorphisms (SNPs) and 1041 candidate genomic regions of selective signals harboring 354 protein-coding genes that were good candidates for domestication genes. Two regulatory factors, SGF-1 and Sage, regulate transcription of the sericin and fibroin genes and also influence the silk yield (Xia et al., 2009). A recent study of silkworm evolution found that the nitrogen metabolism pathway was the most significantly enriched pathway in the set of domestication-associated genes. Further, haplotypes of a cytoplasmic aspartate aminotransferase were nearly fixed within improved strains and diverged from local strains with relatively poor cocoons, suggesting an association between GOT1 and silk yield regulation (Xiang et al., 2018). Several comparative transcriptomic and quantitative proteomic approaches have screened a large number of differentially expressed genes and proteins associated

Abbreviations: FAIRE, formaldehyde-assisted isolation of regulatory elements; OCR, open chromatin regions; COREs, cluster of related regions; DZ, DaZao; TSS, transcription start site; TES, transcription end site; Ubx, ultrabithorax; CPH43, cuticular protein hypothetical 43; CI, carboxypeptidase inhibitor; SNP, single-nucleotide polymorphisms; Deaf1, deformed epidermal autoregulatory; Sna, snail; Vis, vismay; Achi, achintya; Hth, homothorax; Ara, araucan

* Corresponding author. State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400715, China.
** Corresponding author. Biological Science Research Center, Southwest University, Chongqing 400715, China.
*** Corresponding author. Biological Science Research Center, Southwest University, Chongqing 400715, China.
E-mail addresses: chengtc@swu.edu.cn (T. Cheng), xiaqy@swu.edu.cn (Q. Xia), hehuawei@swu.edu.cn (H. He).

https://doi.org/10.1016/j.ibmb.2019.04.014
Received 18 December 2018; Received in revised form 18 March 2019; Accepted 13 April 2019
Available online 18 April 2019

with silk production (Fang et al., 2015; Li et al., 2016; Wang et al., 2014, 2016)1504. These studies indicated a relationship between silk yield and genotype. However, the gene regulation mechanism in genomic regulatory regions among different silk yield strains remains unclear.

The transcriptional regulation of genes coding for silk proteins is very complex and involves many regulatory elements and regulatory models. Generally, spatial and temporal expression is controlled by related hormones and homeodomain factors. For example, 20-hydroxyecdysone (20E) primary response gene *E93* modulates 20E signaling to promote larval-pupal metamorphosis (Liu et al., 2015). Ultrabithorax (Ubx) and abdominal-A specify the abdominal appendage in a dosage-dependent manner in silkworm (Tong et al., 2017). Additionally, transcription factors (TFs), such as SGF-1 (Mach et al., 1995), SGF-2 (Ohno et al., 2013), and SGF-3 (Matsuno et al., 1990), can regulate the expression of silk protein genes. Fibroin heavy chain is a major component of fibroin. The promoter of *Fib-H* contains five functional regions (A–E) and the intron and exon reduplicate regions of *Fib-H* contain numerous *cis*-elements contributing to *Fib-H* regulation. Hox transcription factor Antp and basic helix-loop-helix TFs Bmsage and Bmdimm have been reported to be involved in regulating *Fib-H* and other silk protein genes (Tsubota et al., 2016; Zhao et al., 2014, 2015). Antp was also reported to function in the development of thoracic legs and segmentation (Chen et al., 2013). These studies indicate that regulation of the silk protein gene is complicated and associated with various biological processes, including segmentation and hormones. While the studies described above were valuable for advancing silkworm gene regulation research, they all focused on TFs and their binding sites. However, the regulatory regions of genes dynamically change (Thakurela et al., 2015) and can appear in promoter-proximal, promoter-distal, or gene-body regions (Schoenfelder et al., 2015). Conventional methods may not detect TFs and their regulatory sites.

To evaluate TFs and motifs in detail, we applied bioinformatic approaches for transcriptional regulation analysis. Because of the decreasing costs of next-generation sequencing and availability of methods for capturing specific genomic regions, such as chromatin immunoprecipitation (ChIP)-seq (Mardis, 2007), formaldehyde-assisted isolation of regulatory elements (FAIRE)-seq (Gaulton et al., 2010), and DNase1-seq (He et al., 2014), genome-wide studies of genomic regions of interest have become possible. FAIRE-seq is convenient method for isolating and analyzing open chromatin regions (OCRs), which play a role as *cis*-element predictors (Naval-Sanchez et al., 2015). OCRs have been widely evaluated in genome-wide transcriptional regulation research. For instance, FAIRE can isolate active *cis*-active elements for human chromatin, including the active promoter, transcription start site (TSS), and cell type-specific regulatory sites (Giresi et al., 2007). FAIRE has also been used to analyze stage-specific regulation in drosophila to compare changes in OCRs and gene signals during different stages to produce regulation models (McKay and Lieb, 2013). In silkworm, our team was the first to isolate the OCRs; specifically, we identified more than 56 types of motifs, including development-associated TFs (sna, ara, etc.) and hormone-responsive TFs (Kr-h1, eip74EF, etc.), many of which had not been identified previously in silkworms. We developed new regulatory models and detected numerous new binding sites of known TFs (Zhang et al., 2017). Thus, FAIRE-seq is a simple, economical, and efficient method for conducting transcriptional regulation research.

In this study, we performed FAIRE-seq of the fifth-instar 3-day larvae silk glands, and OCRs were isolated from two silkworm strains, DaZao (DZ) and D872. D872-specific OCRs significantly upregulated adjacent genes. Further, SNPs were identified in the OCRs but had little influence on OCR signaling. We identified 2157 COREs, which indicated synergy among the OCRs. The COREs regulated the silk protein synthetic pathway-associated genes *Sericin-1b* and *Fib-H*. These results provide insight into the mechanism of high silk production in D872 and improve the understanding of transcriptional regulation among strains, which is valuable for strain improvement.

2. Methods

2.1. Sample preparation

The local strain DaZao (normal silk yield) and the D872 strain (high silk yield) were chosen from the colony maintained at the State Key Laboratory of Silkworm Genome Biology. The left silk gland of each larva was used for FAIRE-seq, with each larva serving as a biological replicate. RNA-seq was carried out using the right silk glands.

2.2. FAIRE-seq

FAIRE-seq was conducted as described previously (Giresi et al., 2007). Briefly, silk gland tissue was ground into powder in liquid nitrogen and suspended in 1 M phosphate-buffered saline. The powder was then treated with formaldehyde (37%) to induce cross-links in the DNA. After quenching the activity of the formaldehyde via 125 mM glycine, the cell and tissue lysate was sonicated to achieve an average DNA fragment size of approximately 300–400 base pairs (bp) (7 × 30 pulses of 1 s duration, followed by 3 s rest, at 21% amplitude). The ratio of the concentration of FAIRE DNA (which included only OCRs) to that of the FAIRE control DNA sample (whole genomic DNA) was used as the threshold. Only samples with a threshold lower than 0.05 were used for sequencing library construction.

Sequencing libraries were prepared according to the manufacturer's protocols provided with the TruSeq NanoDNA Library Preparation kit (Illumina, San Diego, CA, USA). The FAIRE-seq library was initiated using 100–450 ng of FAIRE DNA. Two rounds of purification were performed using Agencourt AMPure XP beads (Agencourt Biosciences Co., Beverly, MA, USA). The DNA samples were amplified using 18 polymerase chain reaction cycles. The amplified DNA was size-selected to 200–500 bp and sequenced using the Hiseq2000 system (Illumina). We found that a minimum of 1×10^6 aligned reads provided robust sequencing.

2.3. RNA-seq

Total RNA was isolated using the SV Total RNA Isolate System (Z3100; Promega, Madison, WI, USA) according to the manufacturer's instructions. All RNAs were screened using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) to ensure that the sample quality was sufficiently good for RNA-seq library preparation. The rRNA was cleaned using RNAClean XP beads (Illumina) and subjected to reverse transcription to obtain cDNAs. After adenylating the 3′ ends and ligating the adapters, the library was enriched by polymerase chain reaction. The average read size of the library was 260 bp, and more than 1×10^6 aligned reads were obtained. RNA-seq data were analyzed using standard methods. The quality of the raw and processed reads was evaluated using FastQC (Version 0.11.1). PolyA tails were filtered by fqtrim (Version 0.93). Low-quality reads were removed with Trimmomatic. Clean reads were aligned to the silkworm reference genome (KAIKObase) with TopHat (Version 2.0.12.). Differentially expressed genes were detected using RSEM (Version 1.2.29) and Cuffdiff (Trapnell et al., 2013).

2.4. Open chromatin regions and clusters of related regions

Open chromatin regions were generated by MACS2 with the parameter -p 5, and we confirmed that the number of OCRs did not decrease when the value of -p was increased. The DZ- and D872-specific OCRs were defined by MAnorm (Shao et al., 2012) with -r 125. The distribution of OCRs was annotated by ChIPseeker (Yu et al., 2015). The relationship between the locations of OCRs and genes was evaluated by deepTools (Ramirez et al., 2014). The coordinates of OCRs were not all

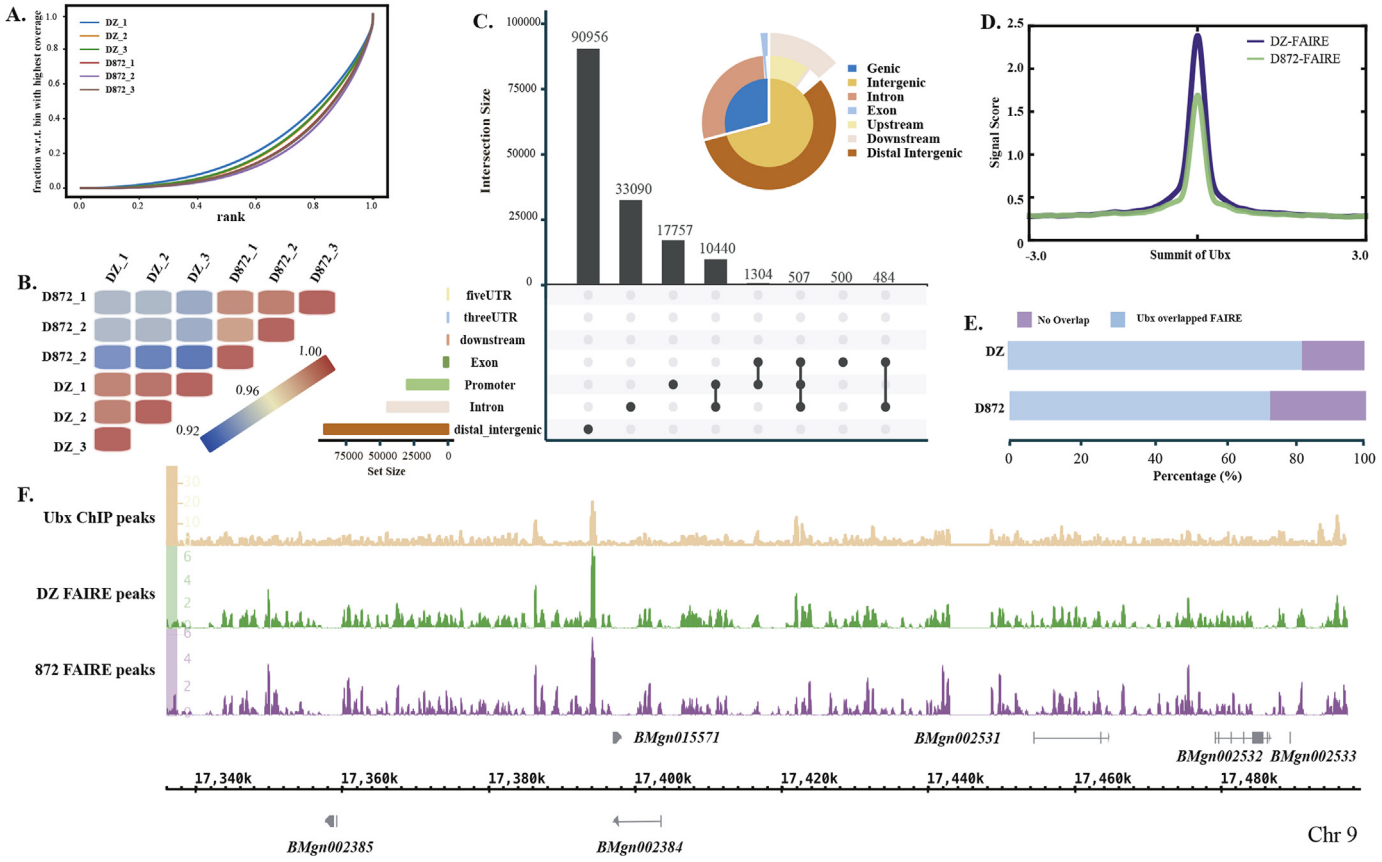

Fig. 1. Quality control of FAIRE-seq data. A. Coverage of reads in 500-bp genomic bins. B. Pearson correlation coefficient of samples. Red and blue represent high and low correlations, respectively. C. Genomic distribution of OCRs. D. Percentage of Ubx overlapped with OCRs in DZ and D872. E. OCRs peak signal enrichment profile in the summit of Ubx binding sites. F. Visualization of OCRs and Ubx sites by integrated genome browser. An approximately 150-kb region of chromosome 9 is displayed. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the same between different samples. Thus, we could not directly plot a heatmap for the genes. First, we normalized the OCRs of different samples to the same coordinate and the signal of each normalized OCR was compared among samples. Finally, the OCRs were divided into four clusters based on their locations in the genome, including Cluster 1 in the gene regions from the TSS to the transcription end site (TES), Cluster 2 downstream the 2-kb region of TESs, Cluster 3 upstream of the 2-kb region of TSSs, and Cluster 4 in intergenic regions.

OCR signals were normalized to the Z-score, and high-, mid-, and low-signal OCRs were measured by the Z-score ($Z = \frac{\frac{1}{n}\sum_{i=1}^{n} xi - E[X]}{\sigma(X)/\sqrt{n}}$). The threshold of high-, low-, and mid-signal OCRs was empirical. Motif analysis was carried out using JASPAR (http://jaspar.genereg.net/) using the insect background database. The relative profile score was greater than 95%. Any motif with a score > 9 was defined as a confidence motif, and the interaction among TFs was generated by using STRING (https://string-db.org/). The Ubx ChIP-seq data were download form GSE71847 (Prasad et al., 2016).

We divided the genome into windows of varying sizes (x-axis) and performed a χ^2 analysis to determine if the number of windows with 0, 1, or > 1 FAIRE sites differed from randomly distributed sites (Pearson, 1900). The highest significance was observed in ~5-kb windows. Random peaks were generated by bedtools (Quinlan, 2014) and bedops (Neph et al., 2012). If two DZ- or D872-specific OCRs were 5 kb away from each other, they were considered as involved in a CORE. The CORE gene was enriched in the 2-kb upstream and downstream genes. The P value was generated by the Wilcoxon test with R package ggpubr (http://www.sthda.com/english/rpkgs/ggpubr/).

2.5. SNP calling

D872 resequencing data were downloaded from SRA009208 (Xia et al., 2009). The clean data were mapped to the reference genome (DZ genome from KAIKObase). SNPs were identified by Samtools (Li et al., 2009) and GATK (Pluss et al., 2017). The D872 genome sequence was generated by applying D872 vcf format variants to the genome fasta file using VCFtools (Danecek et al., 2011). The OCR sequences of D872 all used the D872 vcf consensus sequence. SNPs were annotated by SnpEff (Cingolani et al., 2012).

2.6. Dual luciferase assay

OCR sequences were inserted upstream of the *Luc* + of pGL3-Basic vector (Promega). The reverse complement of the sequence was used if the sequence was located on the negative strand (PCR mix, Novoprotein, Shanghai, China). Ligated products were transformed into *Trelief*™ 5α Chemically Competent Cells (TsingKe Co., Ltd., Beijing, China). Each clone was verified by Sanger sequencing. The correct clones were transfected into a duplicate silkworm *BME* cell line (silkworm embryo-derived cell). Cells were co-transfected with a TK vector to control for transfection efficiency. Fluorescence was detected with the SYNERGY-H1 microplate reader (BioTek, Winooski, VT, USA) after Dual Luciferase Assay reagent (Promega) was added (48 h after transfection). Every assay included three replicates and was repeated at least twice. Firefly luciferase activity was normalized to TK luciferase activity. A two-sided *t*-test was used to compare luciferase activity between HSP70 promoter-containing pGL3 vectors with and without OCRs.

3. Results

3.1. Data generation and quality control

FAIRE-seq was performed in the larval silk glands using triple biological repeats of both DZ and D872. In total, 96,232,723 raw reads were generated by the Illumina Hiseq2000 sequencer. Briefly, 96,212,611 high-quality reads remained, with an average trimming ratio of 99% (Table S1). The mapping ratio of qualified reads was 91% to the silkworm genome. For each sample, more than 60% of reads covered approximately 90% of non-overlapping 500-bp genomic sliding bins (Fig. 1A). The Pearson correlation coefficient was high among replicates and low among different samples (Fig. 1B, Fig. S1).

OCRs were reported to be the predictors of *cis*-elements, such as enhancers, promoters, and insulators, with the "open" and "closed" states of chromatin influencing adjacent genes (Davie et al., 2015). The signal values and distribution of OCRs and genes often indicates a regulatory relationship (Bajic et al., 2018). For subsequent analysis, we merged the peaks from the biological replicates. Only peaks associated with -log10 (q-value) ≥5 and fold-change values ≥ 5 were considered as OCRs. In total, 305,718 OCRs were generated by the parameter *-p* 5 of MACS2, covering approximately 10% of the silkworm genome (Table S2). Approximately 80% of the OCRs were in intergenic regions and noncoding regions (Fig. 1C). In D872, 4,510 specific OCRs were identified and covered 0.288% of the genome. DZ contained 7,337 specific OCRs and covered 0.466% of the genome (Table S3, Fig. S2). We also compared these FAIRE peaks with published Ubx ChIP-seq peaks (Prasad et al., 2016) (Fig. 1D–F). The peak of the FAIRE peaks was highly enriched in the location of the Ubx ChIP site peak, and more than 70% of Ubx peaks overlapped with FAIRE peaks. These data suggest that FAIRE peaks are reliable.

3.2. Function of open chromatin regions to adjacent genes

We first defined OCR intersected genes as OCR enriched genes and genes 10 kb away from OCR genes as noOCR enriched genes. We found that OCR adjacent genes were significantly upregulated compared to noOCR adjacent genes (P_{D872} = 3.4e-07, P_{DZ} = 1.6e-05) (Fig. 2A). The expression level of all OCR adjacent genes significantly differed between DZ and D872 (P_{D872} = 2e-16, P_{DZ} = 1.8e-06). However, only D872-specific OCR adjacent genes were differentially expressed between DZ and D872 (P = 3e-05) (Fig. 2B). This suggests that OCRs regulate adjacent genes, and D872-specific OCRs are more functional in determining the specific traits of D872.

OCRs can be found anywhere in the genome and have variable functions in the organism. Previous studies reported that FAIRE is sensitive to the enhancer, but not to the promoter (McKay and Lieb, 2013). Therefore, we divided the OCRs into four clusters depending on their distances from genes. Most OCRs were located in the gene body and 2 kb up/downstream of genes. However, both upregulated and downregulated genes appeared adjacent to all four regions (Fig. 2C), indicating that OCR directionally regulate genes. Here we used two OCRs in Cluster 1 as examples. In the first intron of *CPH43*, a D872-specific OCR (CPH43_OCR) was found, and the expression of *CPH43* in D872 was lower than in DZ (Fig. 2D), suggesting that CPH43_OCR downregulates the expression of *CPH43* in D872. We followed the order of OCR, promoter of *CPH43,* and linked them upstream of *Luc* + of the pGL3-Basic vector (Fig. 2F). The results showed that CPH43_OCR inhibited promoter activity (Fig. 2G). Further, the DZ-upregulated genes were stimulated by DZ-specific OCRs. Carboxypeptidase inhibitor (CI) catalyzes amino acid hydrolysis from the C-terminus of the peptide chain. We identified a DZ-specific OCR (CI_OCR) in the second intron of *CI* (BMgn016259), and the expression of *CI* in DZ was higher than that in D872. The firefly fluorescence signal of the CI_OCR-linked vector was significantly higher than that of the non-CI_OCR vector (Fig. 2G), indicating that CI_OCR is involved in positively regulating *CI*. Thus, OCRs can regulate genes in different directions.

3.3. Single nucleotide polymorphisms affect the function of open chromatin regions

The functions of OCRs are variable. Previous studies using model organisms have shown that SNPs function in many biological processes, including transcriptional regulation (Noto et al., 2018; Zou et al., 2017). In this study, 1,493,558 SNPs were found in D872 (DZ genome was the reference genome). A mutation appeared at every 277 bp throughout the D872 genome (Table S4). Less than 7% of SNPs were found in exons. More than 50% of mutations were in non-coding regions, and 37.4% of SNPs were in OCRs of D872 (Fig. 3). We examined whether SNPs are associated with gene regulation differences between DZ and D872. The expression of genes adjacent to SNPs harboring OCRs significantly differed between DZ and D872 (P < 2e-16). Thus, we inferred that SNPs harboring OCRs affected gene transcriptional regulation. To identify the main SNP functional regions, we investigated the signal strength of OCRs and its association with the level of interaction between TF and DNA. We normalized the signal strength to the Z-score and recorded the distribution of the high signals (Z-score ≥ 1.05), mid signals (−0.814 ≤ Z-score < 1.05), low signals (Z-score < −0.814), and complex signals (Z-score = high, mid, and low complex) of SNPs harboring OCRs. We found that genes adjacent to mid- and low-signal OCRs were significantly different between DZ and D872 (P_{mid} < 2e-16, P_{low} = 2.07e-07). However, the high- and complex-signal OCR adjacent genes were similar (P_{high} = 0.13, $P_{complex}$ = 0.28). This suggests that the effects of SNP in mid- and low-signal OCRs were higher than those of high- and complex-signal OCRs.

3.4. Synergy regulation of open chromatin regions

For OCRs, there were also some different synergy mechanisms that have been proven in humans (Spitz and Furlong, 2012). Here, we found that OCRs are not evenly distributed; instead, they are located in physically linked clusters throughout the genome. The silkworm OCRs were significantly different from random peaks separated through 5 kb genomic sliding windows (Fig. 4A). Therefore, we defined that COREs were observed with at least two specific OCRs with a separation distance of less than 5 kb. In total, we identified 640 and 1517 COREs in DZ and D872 (Table S5), respectively. The length of specific COREs ranged from 1,031 to 15,177 bp (average length 2,574 bp) in D872 and 782 to 55,604 bp (average length 2,136 bp) in DZ. The average length of the COREs was more than 9-fold of the OCRs, and the longest CORE contained 27 OCRs (Fig. 4B).

Clusters of regulatory elements are considered extremely important in *cis*-regulation (Gaulton et al., 2010). COREs were separated throughout the genome, and genes in COREs showed significantly different expression levels. We found that the COREs of D872 specifically upregulated D872 genes (P = 1.7e-09). However, DZ COREs downregulated DZ genes (P = 0.022) (Fig. 4C and D). To evaluate the function of CORE-harboring genes in both DZ and D872, we annotated these genes by KEGG pathway analysis (Fig. 4E). The KEGG results indicated that more than 41% of DZ CORE-harboring genes were mainly enriched in DNA replication, metabolic cycle, and cell cycle-associated pathways. Approximately 32% of D872 genes regulated by the COREs were mostly enriched in energy metabolism, transmembrane transport, and synthetic secretion-associated pathways, which are related to silk protein synthesis (Table S6). This suggests that the COREs function in silk protein synthesis-associated genes and may be involved in increasing the silk production of D872 by upregulating these genes.

Sericin is the main component of silk, and *Sericin-1b* is the main sericin-coding gene in silkworm (Xia et al., 2014). The expression level of *Sericin-1b* in D872 was approximately 2-fold higher than that in DZ. We detected eight specific OCRs (Ocr1–8) in *Sericin-1b* (Fig. 5A). The distance between Ocr3 and Ocr4 was more than 5 kb, and thus were

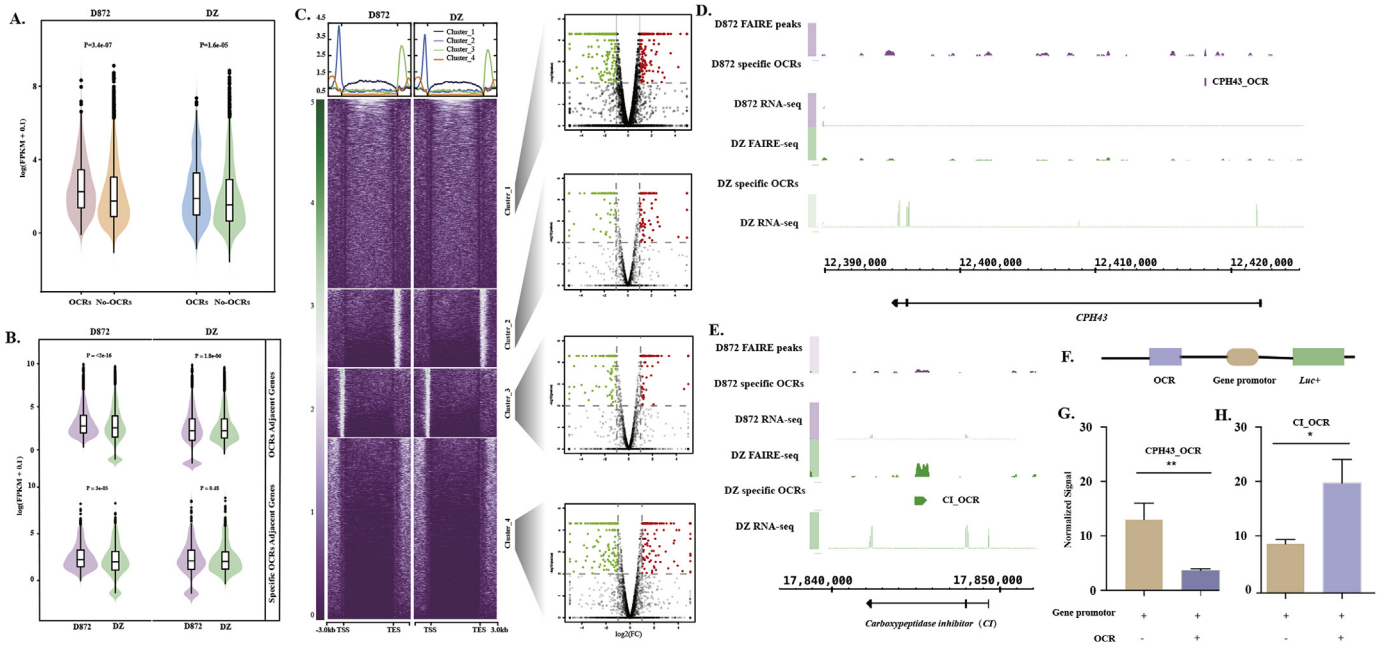

Fig. 2. Function of OCRs to adjacent genes. A. Comparisons of OCRs and specific OCRs adjacent gene expression levels. Top violin plot represents the expression level comparison of genes adjacent to all OCRs of D872 and DZ. Button violin plot shows specific OCRs of DZ and D872 adjacent genes expression level comparison. B. Gene expression level of OCR adjacent genes and noOCRs adjacent genes in DZ and D872. C. Cluster of OCRs by both DZ and D872 distribution. On the left is the cluster heatmap and profile. On the right are the differentially expressed genes adjacent to OCRs. Red and green points indicate upregulated and downregulated genes, respectively. D, E Peaks of CI and CPH43 genes and their specific OCRs. F. Arrangement of OCR and gene promoter in the pGL3 vector. G.H. Fluorescence report of OCR function in these two genes. ns: no significance, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

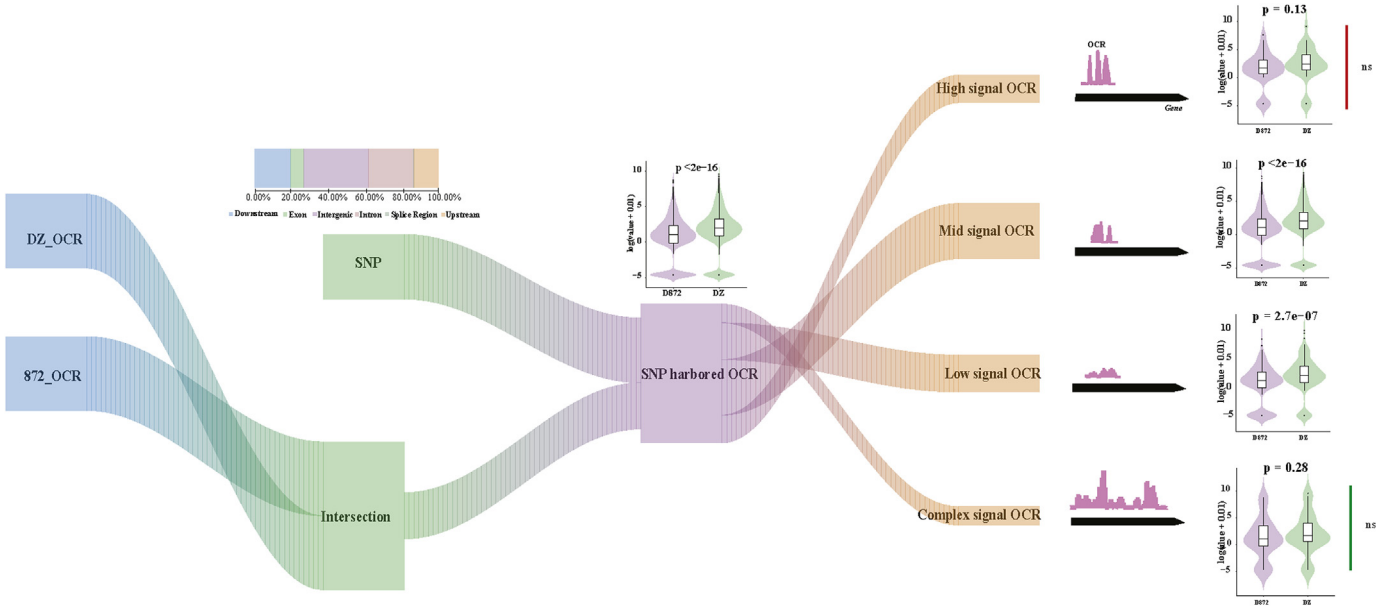

Fig. 3. Riverplot of SNP effect on OCRs. The "river" flowed from left to right. Blue regions represent whole OCRs of DZ and D872. The green regions labeled as "Intersection" indicate overlapped OCRs between DZ and D872. The green regions labeled as "SNP" represent the SNPs from D872. Bar plot represents the distribution of SNPs in D872. The purple regions labeled as "SNP harbored OCRs" represent SNPs harbored in overlapped OCRs between DZ and D872. "High," "mid," and "low" indicate the signal strength of OCRs. "Complex" indicates that more than one type of signal was enriched for a gene. The violin plot represents the expression level of genes adjacent to the right regions OCRs. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

split into two COREs (Core1, Core2). Here, we built three types of pGL3-Basic vectors (Fig. 5B). First, single OCRs and the HSP70 promoter were ligated upstream of *Luc*+, which was used to detect the function of single OCRs. Second, the CORE and HSP70 were ligated upstream of *Luc*+, which was used to detect the function of the signal CORE. Third, two COREs were ligated upstream of the HSP70 promoter in the order Core1 and Core2, after which the CORE string and HSP70 were ligated upstream of *Luc*+. Finally, their firefly luciferase signals were compared to those of pGL3 containing only the HSP70 promoter ligated upstream of *Luc*+. The dual luciferase assay showed that the

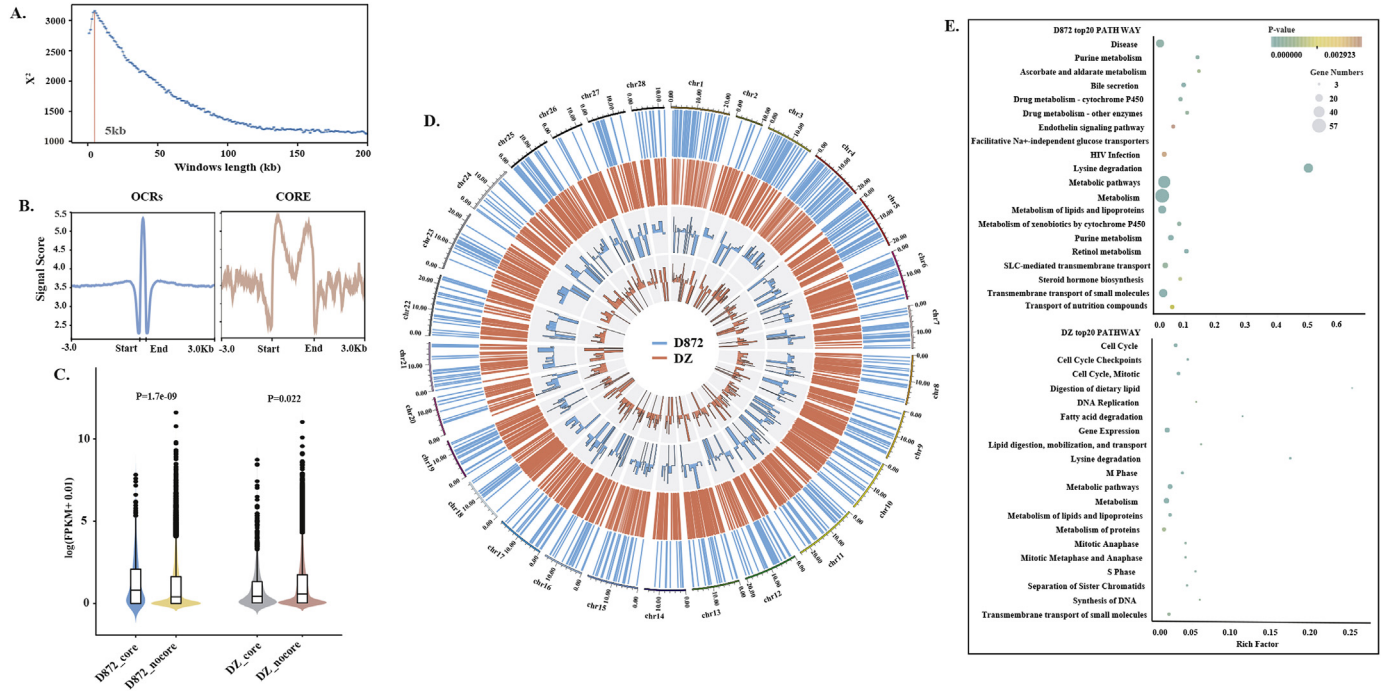

Fig. 4. Generation and functional analysis of COREs. A. Chi-square test results of OCRs and random OCRs frequency detected in different genomic windows. B. Signal profile between the start and end of OCRs and COREs. C. Gene expression comparison between the CORE-harbored genes and noCORE genes in both DZ and D872. D. CORE distribution and gene expression level of CORE-harbored genes in both DZ and D872. From outside to inside, they are: D872 COREs, DZ COREs, D872 CORE harbored genes expression level, and DZ COREs harbored genes expression level. E. P-values of top 20 pathways of DZ and D872. The top is D872, and the bottom is DZ.

single OCRs had a variety of regulatory functions. Most OCRs had negative regulation traits, except for Ocr2, Ocr4, and Ocr5. However, the function of OCRs changed much more when the single OCRs composed a CORE. Interestingly, neither Core1 nor Core2 showed significant positive regulation traits (Fig. 5C and D). The firefly fluorescence signal of Core1&Core2 was significantly higher (approximately 2-fold) than that of the control (Fig. 5E). The synergy of eight OCRs (Ocr1–8) showed a significant positive regulatory function, which may explain

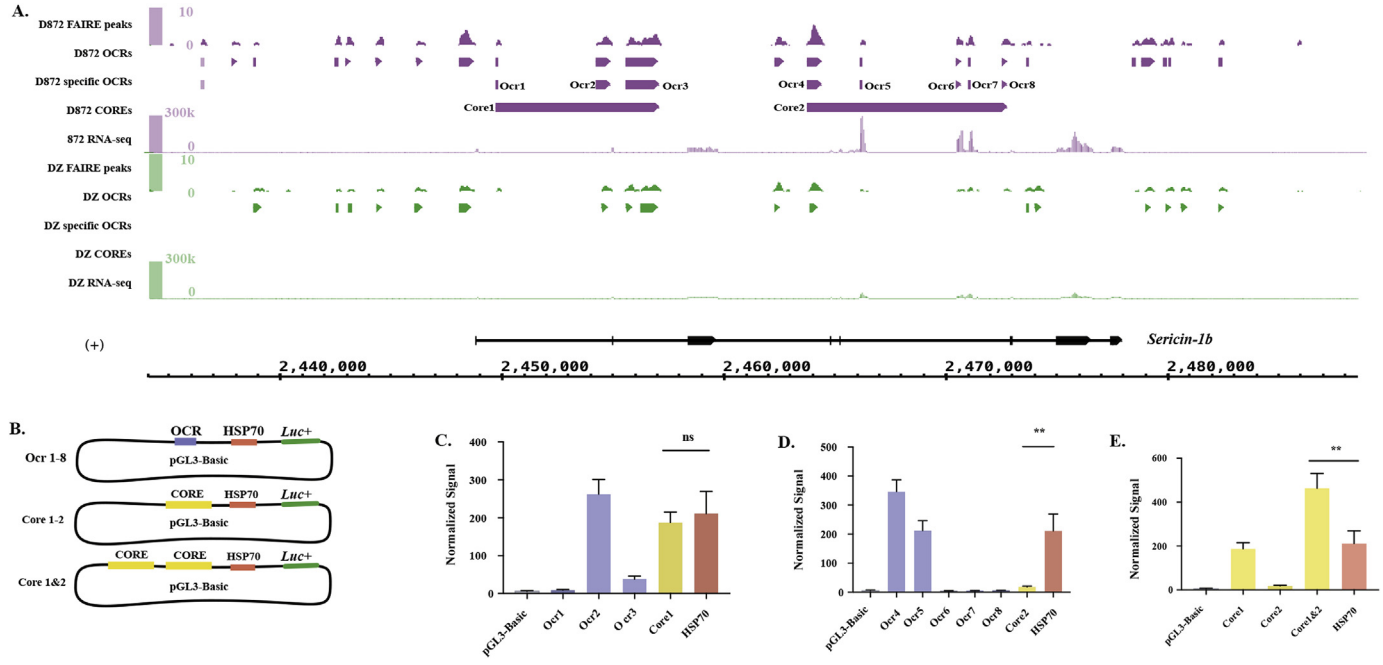

Fig. 5. Functional analysis of COREs harbored in *Sericin-1b*. A. Visualization of *Sericin-1b* expression peaks, OCR, specific OCRs, and COREs in the genome browser of both DZ and D872. B. Structure of vectors used to verify the function of COREs. Ocr1–8 contained single OCRs ligated upstream of the HSP70 promoter. Core1, Core2 contained single COREs ligated upstream of the HSP70 promoter. Core1&Core2 contained Core1 and Core2 ligated in a series upstream of HSP70. In the control, nothing was ligated upstream of the HSP70 pGL3-Basic vector. C. D. E. Dual luciferase assay results of Ocr1–Ocr8, Core1–2, Core1&2. H. Ocr1–Ocr3, Core1. I. Ocr4–Ocr8, Core2. J. Core1-2, Core2&Core2. ns: no significance, *: $p < 0.05$, **: $p < 0.01$, and ***: $p < 0.001$.

why *Sericin-1b* expression is higher in D872.

4. Discussion

Studies of transcriptional regulation among strains of silkworm are challenging because of the lack of silkworm TF annotations. This also makes it difficult to conduct genome-wide of transcription regulation studies. In this study, we investigated the transcriptional regulation model by comparing differences in OCR signals and distributions between the silkworm strains DZ and D872. We found several regulatory mechanisms in both DZ and D872, including those involving "open" or "closed" chromatin, mutations (SNPs), and OCR synergies (COREs). The expression of OCR adjacent genes differed from noOCR genes in both DZ and D872. D872-specific OCRs significantly upregulated their adjacent genes. OCRs were also found to function as clusters on genes; these clusters were named as COREs. The CORE genes showed higher expression than no-CORE genes in D872, but the expression in DZ was decreased. The function of CORE-harboring genes significantly differed between DZ and D872. Further, SNPs affected the regulation of genes by OCRs, which correlated with the OCR signals; high-signal OCRs were more stable than mid- or low-signal OCRs. These results demonstrated the dynamic balance of transcriptional regulation between silkworm strains and revealed new differences in regulatory models between DZ and D872.

The OCR profile is a snapshot of the chromatin state, which can reflect the real-time conditions of chromatin (Yang and Sawa, 2017). Additionally, OCRs are high-confidence loci of nucleosome depleted regions, which facilitate the binding of *trans*-factors (Natarajan et al., 2012). The *trans*-factors that dynamically bind or unbind to OCRs may regulate gene expression. For instance, *P450* encodes a cytochrome, which is involved in biocatalysis, hormone metabolism, and detoxification in silkworm (Daimon et al., 2012). *P450* was expressed in DZ, but nearly no expression was observed in D872 (Fig. S3A). A DZ-specific OCR was found in the first exon of *P450*. In this OCR, we detected six types of motifs (confidence score ≥95%), including Deaf1 (Reed et al., 2008), sna (Rembold et al., 2014), vis (Hyman et al., 2003), achi (Hyman et al., 2003), hth (Li-Kroeger et al., 2008), ara (GomezSkarmeta et al., 1996), and Ubx (Johnson et al., 1995) (Fig. S3B). Most were positive regulation TF-binding motifs, such as those for vis and achi. Vis and achi have similar protein structures; both can positively regulate upstream genes and were predicted to interact with each other (interaction score = 0.965) by the STRING database (Fig. S3C). Ubx has been reported to be involved in somite segmentation control and hormone response (Johnson et al., 1995). Thus, achi, hth, and vis may comprise a positive regulatory complex and function in the locus specified by Ubx. This may have resulted in *P450* upregulation in DZ.

Previous studies reported that high-signal OCRs were more conserved (McKay and Lieb, 2013). This may be consistent with the lack of significant differences between high-signal and mix-signal OCR adjacent genes observed between DZ and D872. Further, high-signal OCRs significantly regulated adjacent genes ($P_{high\text{-}low}$ = 0.013, $P_{high\text{-}mid}$ = 0.0061) (Fig. S4). For instance, OCR_1 (236 bp) was in the second exon of *Fib-H* (Fig. S5). Ten SNPs were detected in D872 OCR_1. D872 lost two eip74EF (Urness and Thummel, 1995) and two sna motifs because of four mutations (A-G, G-A) and one mutation (T-A) in OCR1. However, the dual fluorescence assay indicated that SNPs had a minimal effect on the function of OCR_1 (high-signal OCR) ($P_{DZ\_OCR\_1}$ VS $P_{D872\_OCR\_1}$ = 0.9898) but significantly upregulated the expression of target genes (P_{DZ_OCR1} = 0.0135, P_{D872_OCR1} = 0.0248). The cross-talk between the function and sequence of high-signal OCRs may be important in silkworm if the function of high-signal OCRs is greatly altered, which would also significantly change the gene expression. These changes may cause qualitative changes in silkworm. Thus, the stability of high-signal OCRs may contribute to the evolutionary conservation of silkworm.

Sericin-1b encodes a well-known key silk sericin in silkworm. This gene is approximately 24 kb and contains nine exons (Kragesteen et al., 2018; Went et al., 2018). The upstream and inside of *Sericin-1b* harbors many regulatory regions, which create many intertwined topological structures that are important for gene regulation. Here, we clustered the OCRs using an appropriate threshold (5 kB), which could objectively explain the synergistic mechanism of OCRs. In the *Sericin-1b*-enriched CORE, we predicted that topological structures formed among the OCRs to enable TF binding to OCRs to form a complex. The TF complex upregulated *Sericin-1b* expression. Further, the topological structure and its corresponding TF complexes changed dynamically in their spatial and temporal properties (Kragesteen et al., 2018). Therefore, the CORE is important for the growth and development and individual differences between silkworm strains. Additionally, *Sericin-1b* was enriched for eight OCRs, most of which were negative regulators. However, Core1 and Core2 showed positive regulatory effects. COREs play very important roles in maintaining the balance of material and energy metabolism in silkworm. Moreover, we found that COREs were ubiquitous in the silkworm genome. CORE-enriched genes were most abundant in silk protein synthetic pathways in D872. Additionally, the two most important silk protein genes, *Fib-H* and *Sericin-1b*, were both regulated by COREs, and their expression was significantly higher in D872. COREs may also explain the differences in silk production between DZ and D872. By comparing the regulation models of silk protein genes in the entire silk gland, although we found some specific COREs for regulating silk protein genes between the two strains, their strict transcriptional regulation based on spatial and temporal properties was not determined. Comparison of the spatial and temporal differences of the silk gland, including OCRs, may provide accurate insight into the transcriptional regulation of silk protein genes.

Disclosure statement

The authors have no conflict of interests to disclose.

Database linking

The datasets supporting the conclusions of this article are available under BioProject accession number PRJNA495893 for both RNA-seq and FAIRE-seq in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/).

Acknowledgements

Q.Z. and T.C. conceived and designed the study. Q.Z., Y.S., X.L., and L.L. performed the biologic experiments. Q.Z., Y.W., T.F., Z.L., and C.L. analyzed the NGS dataset. Q.Z. and T.C. wrote the manuscript. Z.L., H.H., and Q.X. revised the manuscript. This work was supported by the National Natural Science Foundation of China (Grant No. 31872296, 31572465, 31530071), Chongqing Research Program of Basic Research and Frontier Technology (Grant No. CSTC2018JCYJAX0396, CSTC2015JCYJBX0035), and Chongqing Postgraduate Research and Innovation Project (Grant No. CYB18103).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ibmb.2019.04.014.

References

Bajic, M., Maher, K.A., Deal, R.B., 2018. Identification of open chromatin regions in plant genomes using ATAC-seq. Methods Mol. Biol. 1675, 183–201.
Chen, P., Tong, X.L., Li, D.D., Fu, M.Y., He, S.Z., Hu, H., Xiang, Z.H., Lu, C., Dai, F.Y., 2013. Antennapedia is involved in the development of thoracic legs and segmentation in the silkworm, Bombyx mori. Heredity 111, 182–188.
Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X.,

Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6, 80–92.
Daimon, T., Kozaki, T., Niwa, R., Kobayashi, I., Furuta, K., Namiki, T., Uchino, K., Banno, Y., Katsuma, S., Tamura, T., Mita, K., Sezutsu, H., Nakayama, M., Itoyama, K., Shimada, T., Shinoda, T., 2012. Precocious metamorphosis in the juvenile hormone-deficient mutant of the silkworm, Bombyx mori. PLoS Genet. 8 e1002486.
Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R. E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., Genomes Project Analysis, G., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156-2158.
Davie, K., Jacobs, J., Atkins, M., Potier, D., Christiaens, V., Halder, G., Aerts, S., 2015. Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. PLoS Genet. 11 e1004994.
Fang, S.M., Hu, B.L., Zhou, Q.Z., Yu, Q.Y., Zhang, Z., 2015. Comparative analysis of the silk gland transcriptomes between the domestic and wild silkworms. BMC Genomics 16, 60.
Gaulton, K.J., Nammo, T., Pasquali, L., Simon, J.M., Giresi, P.G., Fogarty, M.P., Panhuis, T.M., Mieczkowski, P., Secchi, A., Bosco, D., Berney, T., Montanya, E., Mohlke, K.L., Lieb, J.D., Ferrer, J., 2010. A map of open chromatin in human pancreatic islets. Nat. Genet. 42, 255–259.
Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., Lieb, J.D., 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res. 17, 877–885.
GomezSkarmeta, J.L., Diez del Corral, R., delaCalleMustienes, E., FerresMarco, D., Modolell, J., 1996. Araucan and caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. Cell 85, 95–105.
He, H.H., Meyer, C.A., Hu, S.S., Chen, M.W., Zang, C., Liu, Y., Rao, P.K., Fei, T., Xu, H., Long, H., Liu, X.S., Brown, M., 2014. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. Nat. Methods 11, 73–78.
Hyman, C.A., Bartholin, L., Newfeld, S.J., Wotton, D., 2003. Drosophila TGIF proteins are transcriptional activators. Mol. Cell Biol. 23, 9262–9274.
Johnson, F.B., Parker, E., Krasnow, M.A., 1995. Extradenticle protein is a selective co-factor for the Drosophila homeotics - role of the homeodomain and Ypwm amino-acid motif in the interaction. P. Natl. Acad. Sci. USA 92, 739–743.
Kragesteen, B.K., Spielmann, M., Paliou, C., Heinrich, V., Schopflin, R., Esposito, A., Annunziatella, C., Bianco, S., Chiariello, A.M., Jerkovic, I., Harabula, I., Guckelberger, P., Pechstein, M., Wittler, L., Chan, W.L., Franke, M., Lupianez, D.G., Kraft, K., Timmermann, B., Vingron, M., Visel, A., Nicodemi, M., Mundlos, S., Andrey, G., 2018. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat. Genet. 50, 1463–1473.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Genome Project Data Processing, S., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078-2079.
Li, J., Qin, S., Yu, H., Zhang, J., Liu, N., Yu, Y., Hou, C., Li, M., 2016. Comparative transcriptome analysis reveals different silk yields of two silkworm strains. PLoS One 11 e0155329.
Li-Kroeger, D., Witt, L.M., Grimes, H.L., Cook, T.A., Gebelein, B., 2008. Hox and senseless antagonism functions as a molecular switch to regulate EGF secretion in the Drosophila PNS. Dev. Cell 15, 298–308.
Liu, X., Dai, F., Guo, E., Li, K., Ma, L., Tian, L., Cao, Y., Zhang, G., Palli, S.R., Li, S., 2015. 20-Hydroxyecdysone (20E) primary response gene E93 modulates 20E signaling to promote Bombyx larval-pupal metamorphosis. J. Biol. Chem. 290, 27370–27383.
Mach, V., Takiya, S., Ohno, K., Handa, H., Imai, T., Suzuki, Y., 1995. Silk gland factor-1 involved in the regulation of Bombyx sericin-1 gene contains fork head motif. J. Biol. Chem. 270, 9340–9346.
Mardis, E.R., 2007. ChIP-seq: welcome to the new frontier. Nat. Methods 4, 613–614.
Matsuno, K., Takiya, S., Hui, C.C., Suzuki, T., Fukuta, M., Ueno, K., Suzuki, Y., 1990. Transcriptional stimulation via SC site of Bombyx sericin-1 gene through an interaction with a DNA binding protein SGF-3. Nucleic Acids Res. 18, 1853–1858.
McKay, D.J., Lieb, J.D., 2013. A common set of DNA regulatory elements shapes Drosophila appendages. Dev. Cell 27, 306–318.
Natarajan, A., Yardimci, G.G., Sheffield, N.C., Crawford, G.E., Ohler, U., 2012. Predicting cell-type-specific gene expression from regions of open chromatin. Genome Res. 22, 1711–1722.
Naval-Sanchez, M., Potier, D., Hulselmans, G., Christiaens, V., Aerts, S., 2015. Identification of lineage-specific cis-regulatory modules associated with variation in transcription factor binding and chromatin activity using ornstein-uhlenbeck models. Mol. Biol. Evol. 32, 2441–2455.
Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., Sandstrom, R., Humbert, R., Stamatoyannopoulos, J.A., 2012. BEDOPS: high-performance genomic feature operations. Bioinformatics 28, 1919–1920.
Noto, J.M., Chopra, A., Loh, J.T., Romero-Gallo, J., Piazuelo, M.B., Watson, M., Leary, S., Beckett, A.C., Wilson, K.T., Cover, T.L., Mallal, S., Israel, D.A., Peek, R.M., 2018. Pan-genomic analyses identify key Helicobacter pylori pathogenic loci modified by carcinogenic host microenvironments. Gut 67, 1793–1804.
Ohno, K., Sawada, J., Takiya, S., Kimoto, M., Matsumoto, A., Tsubota, T., Uchino, K., Hui, C.C., Sezutsu, H., Handa, H., Suzuki, Y., 2013. Silk gland factor-2, involved in fibroin gene transcription, consists of LIM homeodomain, LIM-interacting, and single-stranded DNA-binding proteins. J. Biol. Chem. 288, 31581–31591.
Pearson, K., 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos. Mag. A 50, 157–175.
Pluss, M., Kopps, A.M., Keller, I., Meienberg, J., Caspar, S.M., Dubacher, N., Bruggmann, R., Vogel, M., Matyas, G., 2017. Need for speed in accurate whole-genome data analysis: GENALICE MAP challenges BWA/GATK more than PEMapper/PECaller and Isaac. Proc. Natl. Acad. Sci. U. S. A. 114, E8320–E8322.
Prasad, N., Tarikere, S., Khanale, D., Habib, F., Shashidhara, L.S., 2016. A comparative genomic analysis of targets of Hox protein Ultrabithorax amongst distant insect species. Sci. Rep. 6, 27885.
Quinlan, A.R., 2014. BEDTools: the Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinf. 47 11 12 11-34.
Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., Manke, T., 2014. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–W191.
Reed, D.E., Huang, X.M., Wohlschlegel, J.A., Levine, M.S., Senger, K., 2008. DEAF-1 regulates immunity gene expression in Drosophila. Proc. Natl. Acad. Sci. U. S. A. 105, 8351–8356.
Rembold, M., Ciglar, L., Yanez-Cuna, J.O., Zinzen, R.P., Girardot, C., Jain, A., Welte, M.A., Stark, A., Leptin, M., Furlong, E.E.M., 2014. A conserved role for Snail as a potentiator of active transcription. Genes Dev. 28, 167–181.
Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W., Dimitrova, E., Dimond, A., Edelman, L.B., Elderkin, S., Tabbada, K., Darbo, E., Andrews, S., Herman, B., Higgs, A., LeProust, E., Osborne, C.S., Mitchell, J.A., Luscombe, N.M., Fraser, P., 2015. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res. 25, 582–597.
Shao, Z., Zhang, Y., Yuan, G.C., Orkin, S.H., Waxman, D.J., 2012. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. Genome Biol. 13, R16.
Spitz, F., Furlong, E.E., 2012. Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. 13, 613–626.
Thakurela, S., Sahu, S.K., Garding, A., Tiwari, V.K., 2015. Dynamics and function of distal regulatory elements during neurogenesis and neuroplasticity. Genome Res. 25, 1309–1324.
Tong, X.L., Fu, M.Y., Chen, P., Chen, L., Xiang, Z.H., Lu, C., Dai, F.Y., 2017. Ultrabithorax and abdominal-A specify the abdominal appendage in a dosage-dependent manner in silkworm, Bombyx mori. Heredity 118, 578–584.
Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., Pachter, L., 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31, 46–53.
Tsubota, T., Tomita, S., Uchino, K., Kimoto, M., Takiya, S., Kajiwara, H., Yamazaki, T., Sezutsu, H., 2016. A hox gene, antennapedia, regulates expression of multiple major silk protein genes in the silkworm Bombyx mori. J. Biol. Chem. 291, 7087–7096.
Urness, L.D., Thummel, C.S., 1995. Molecular analysis of a steroid-induced regulatory hierarchy: the Drosophila E74A protein directly regulates L71-6 transcription. EMBO J. 14, 6239–6246.
Wang, S.H., You, Z.Y., Feng, M., Che, J.Q., Zhang, Y.Y., Qian, Q.J., Komatsu, S., Zhong, B.X., 2016. Analyses of the molecular mechanisms associated with silk production in silkworm by iTRAQ-based proteomics and RNA-sequencing-based transcriptomics. J. Proteome Res. 15, 15–28.
Wang, S.H., You, Z.Y., Ye, L.P., Che, J.Q., Qian, Q.J., Nanjo, Y.H., Komatsu, S., Zhong, B.X., 2014. Quantitative proteomic and transcriptomic analyses of molecular mechanisms associated with low silk production in silkworm Bombyx mori. J. Proteome Res. 13, 735–751.
Went, M., Sud, A., Forsti, A., Halvarsson, B.M., Weinhold, N., Kimber, S., van Duin, M., Thorleifsson, G., Holroyd, A., Johnson, D.C., Li, N., Orlando, G., Law, P.J., Ali, M., Chen, B., Mitchell, J.S., Gudbjartsson, D.F., Kuiper, R., Stephens, O.W., Bertsch, U., Broderick, P., Campo, C., Bandapalli, O.R., Einsele, H., Gregory, W.A., Gullberg, U., Hillengass, J., Hoffmann, P., Jackson, G.H., Jockel, K.H., Johnsson, E., Kristinsson, S.Y., Mellqvist, U.H., Nahi, H., Easton, D., Pharoah, P., Dunning, A., Peto, J., Canzian, F., Swerdlow, A., Eeles, R.A., Kote-Jarai, Z., Muir, K., Pashayan, N., Nickel, J., Nothen, M.M., Rafnar, T., Ross, F.M., da Silva Filho, M.I., Thomsen, H., Turesson, I., Vangsted, A., Andersen, N.F., Waage, A., Walker, B.A., Wihlborg, A.K., Broyl, A., Davies, F.E., Thorsteinsdottir, U., Langer, C., Hansson, M., Goldschmidt, H., Kaiser, M., Sonneveld, P., Stefansson, K., Morgan, G.J., Hemminki, K., Nilsson, B., Houlston, R.S., consortium, P., 2018. Identification of multiple risk loci and regulatory mechanisms influencing susceptibility to multiple myeloma. Nat. Commun. 9, 3707.
Xia, Q., Guo, Y., Zhang, Z., Li, D., Xuan, Z., Li, Z., Dai, F., Li, Y., Cheng, D., Li, R., Cheng, T., Jiang, T., Becquet, C., Xu, X., Liu, C., Zha, X., Fan, W., Lin, Y., Shen, Y., Jiang, L., Jensen, J., Hellmann, I., Tang, S., Zhao, P., Xu, H., Yu, C., Zhang, G., Li, J., Cao, J., Liu, S., He, N., Zhou, Y., Liu, H., Zhao, J., Ye, C., Du, Z., Pan, G., Zhao, A., Shao, H., Zeng, W., Wu, P., Li, C., Pan, M., Li, J., Yin, X., Li, D., Wang, J., Zheng, H., Wang, W., Zhang, X., Li, S., Yang, H., Lu, C., Nielsen, R., Zhou, Z., Wang, J., Xiang, Z., Wang, J., 2009. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (Bombyx). Science 326, 433–436.
Xia, Q., Li, S., Feng, Q., 2014. Advances in silkworm studies accelerated by the genome sequencing of Bombyx mori. Annu. Rev. Entomol. 59, 513–536.
Xiang, H., Liu, X., Li, M., Zhu, Y., Wang, L., Cui, Y., Liu, L., Fang, G., Qian, H., Xu, A., Wang, W., Zhan, S., 2018. The evolutionary road from wild moth to domestic silkworm. Nat. Ecol. Evol. 2, 1268–1279.
Yang, K., Sawa, A., 2017. Open sesame: open chromatin regions shed light onto non-coding risk variants. Cell Stem Cell 21, 285–287.
Yu, G., Wang, L.G., He, Q.Y., 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383.
Yu, H.S., Shen, Y.H., Yuan, G.X., Hu, Y.G., Xu, H.E., Xiang, Z.H., Zhang, Z., 2011. Evidence of selection at melanin synthesis pathway loci during silkworm domestication. Mol. Biol. Evol. 28, 1785–1799.
Zhang, Q., Cheng, T., Jin, S., Guo, Y., Wu, Y., Liu, D., Xu, X., Sun, Y., Li, Z., He, H., Xia, Q., 2017. Genome-wide open chromatin regions and their effects on the regulation of silk

protein genes in Bombyx mori. Sci. Rep. 7, 12919.
Zhao, X.M., Liu, C., Jiang, L.J., Li, Q.Y., Zhou, M.T., Cheng, T.C., Mita, K., Xia, Q.Y., 2015. A juvenile hormone transcription factor Bmdimm-fibroin H chain pathway is involved in the synthesis of silk protein in silkworm, Bombyx mori. J. Biol. Chem. 290, 972–986.
Zhao, X.M., Liu, C., Li, Q.Y., Hu, W.B., Zhou, M.T., Nie, H.Y., Zhang, Y.X., Peng, Z.C., Zhao, P., Xia, Q.Y., 2014. Basic helix-loop-helix transcription factor Bmsage is involved in regulation of fibroin H-chain gene via interaction with SGF1 in Bombyx mori. PLoS One 9 e94091.
Zou, S., Teixeira, A.M., Kostadima, M., Astle, W.J., Radhakrishnan, A., Simon, L.M., Truman, L., Fang, J.S., Hwa, J., Zhang, P.X., van der Harst, P., Bray, P.F., Ouwehand, W.H., Frontini, M., Krause, D.S., 2017. SNP in human ARHGEF3 promoter is associated with DNase hypersensitivity, transcript level and platelet function, and Arhgef3 KO mice have increased mean platelet volume. PLoS One 12 e0178095.