Acta Pharmaceutica Sinica B 2024;14(9):4179−4182

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

LETTER TO THE EDITOR

Decoding the chromatin accessibility in *Andrographis paniculata* genome, a case study of genome-wide investigation of the *cis*-regulatory elements in medicinal plants

KEY WORDS

Andrographis paniculata;
Chromatin accessibility;
Medicinal plant;
Epigenetics;
ENCODE;
cis-regulatory elements

To the Editor:

Medicinal plants have the ability to synthesize a large number of pharmaceutical metabolites, the production and regulation of which are influenced by both intrinsic signals (*e.g.,* the spatial and temporal stage) and extrinsic environmental conditions (*e.g.,* temperature, light availability, and water availability)[1]. According to the Encyclopedia of DNA Elements (ENCODE) project[2], accessible chromatin regions (ACRs) served essential roles in maintaining the genome architecture and gene regulation in both mammals and plants. Nowadays, identifying ACRs within medicinal plant genomes remains a challenge, as genomic sequences have been determined for numerous species yet the characterization of such functional elements lags. In this study, we selected *Andrographis paniculata* (APA)[3] as a model system to map ACRs genome-wide by the assay for transposase-accessible chromatin with sequencing (ATAC-seq) method. We aimed to uncover the genomic features of ACRs as well as their association with the expression of genes involved in secondary metabolite biosynthesis (*e.g.,* andrographolide) within medicinal plants (Supporting Information Fig. S1).

Two ATAC-seq replicates were generated, yielding approximately 106−125 million (M) filtered reads (Fig. 1A; Supporting Information Fig. S2; Table S1). After mapping to the APA genome, a total of 30,532 enriched peaks were identified (Fig. 1A; Supporting Information Table S2), and found to be enriched at transcription start sites (TSSs) (Fig. 1B). ATAC-seq signals were positively correlated with gene expression levels (Fig. 1B; Supporting Information Fig. S3; Table S3). Genomic annotation revealed that these ACRs were mainly located in the exonic (35%), distal intergenic (34%) and promoter (14%) regions (Fig. 1C).

To further characterize the histone modification landscape of ACRs, H3K27me3, H3K4me3 and H3K27ac ChIP-seq were conducted[4] (Supporting Information Figs. S4, S5A and S5B; Table S4). As previously reported, H3K27me3 was an epigenetic marker for gene repression, while H3K4me3 and H3K27ac were associated with gene activation (Fig. S5A). By the K-mean algorithm, ACRs could be clustered into five clusters based on the histone modification signals (Fig. 1D; Supporting Information Table S5). Most ACRs exhibited minimal H3K27me3 modification. ACRs within clusters 1, 2, and 3 were modified by H3K4me3 and H3K27ac in different patterns. In cluster 1, H3K4me3 and H3K27ac signals were distributed across the ACRs. Within cluster 2, H3K4me3 and H3K27ac modifications were enriched in the right flanking regions of ACRs, while the left flanking regions of ACRs in cluster 3 exhibited a profusion of modified signals. ACRs belonging to clusters 4 and 5 were with little histone modifications. These data suggested a crucial role of H3K4me3 and H3K27ac in modulating chromatin accessibility. In addition, genes associated with distinctly modified ACRs exhibited divergent expression levels (Fig. 1E and Fig. S4C). Genes associated with H3K4me3- and H3K27ac-modified ACRs expressed higher

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2024.06.012

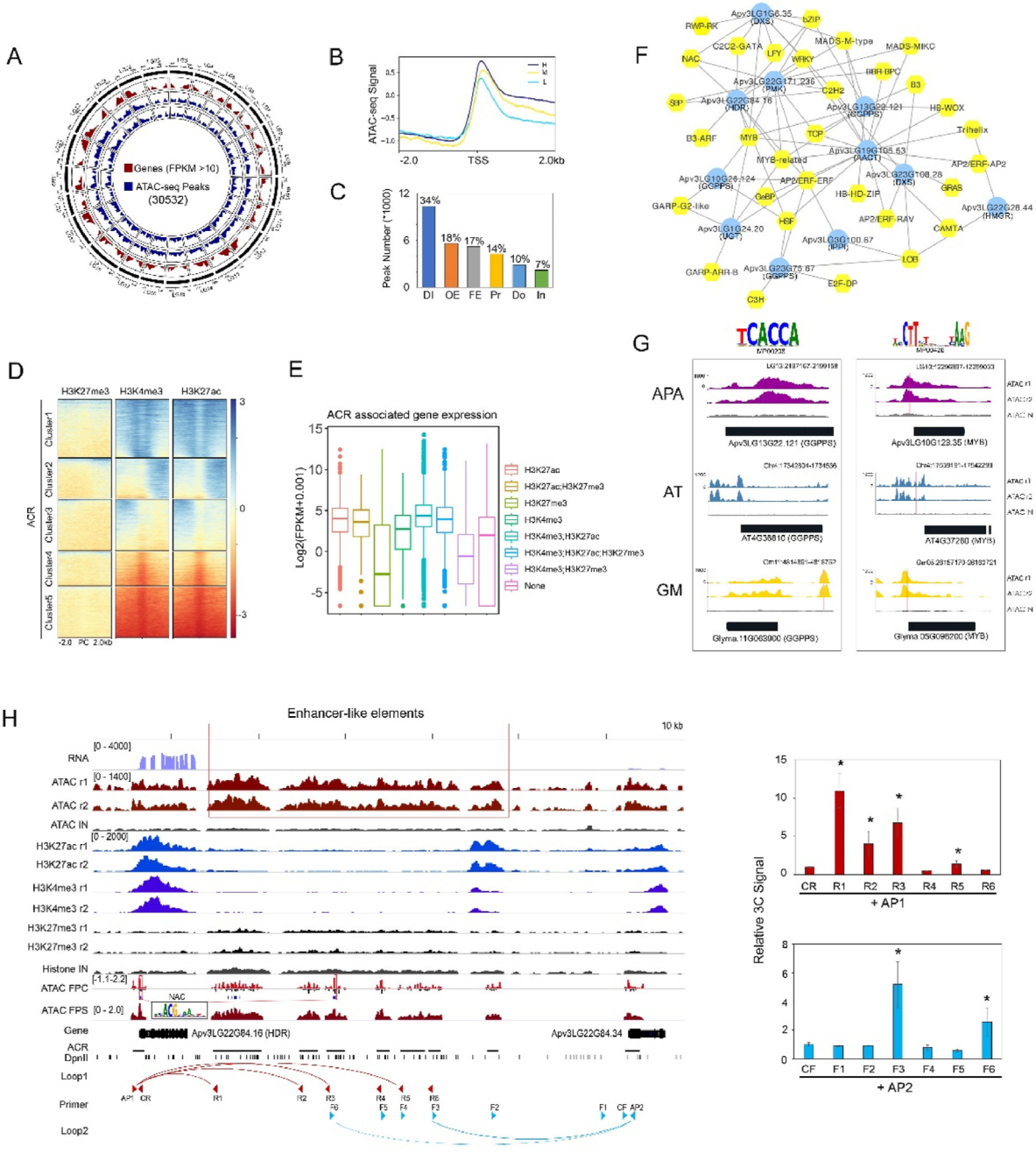

Figure 1 Genomic features of ACRs in APA genome. (A) A visualization of ATAC-seq data. Red circle indicates the distribution of highly expressed genes (FPKM > 10). Two blue circles indicate the ATAC-seq enriched peaks for two biological replicates. LG1−LG25 indicates the 25 gap-free chromosomes. FPKM, Fragments Per Kilobase of transcript per Million mapped reads. (B) Enrichment of ATAC-seq signal at transcription start sites (TSSs) and a positive correlation with the gene expression levels were observed. H, M, L indicate the high (FPKM > 10), middle (10 ≥ FPKM > 1) and low (1 ≥ FPKM) expressed genes, respectively. (C) Distribution of ACRs in the APA genome. DI, Distal intergenic region; OE, other exons excluding the first exon; FE, first exon; Pr, Promoter region; Do, Downstream to transcription termination site (TTS) region; In, intron. (D) Heatmap analysis of ACRs, revealing five classifications (Cluster 1−5) based on H3K27me3, H3K4me3, and H3K27ac ChIP-seq signal intensities using the K-mean algorithm. PC: ACR peak center. (E) Expression levels of ACR-associated genes modified by distinct histones. (F) A potential TF regulatory network related to the andrographolide biosynthesis pathway, predicted by the ATAC footprint analysis. Yellow circles indicate the TFs, while the blue circles indicate the genes involved in andrographolide biosynthesis. (G) Conserved TF binding motifs identified in the ACRs associated homologous genes. The homologous genes of GGPPS and MYB along with their associated ACRs in genome of APA, Arabidopsis (AT) and soybean (GM), were used as the examples. Red lines indicate the location of conserved TF binding motifs, such as the MP00208 and MP00426, respectively. (H) (Left panel) An example of enhancer-promoter interaction is validated by

than those associated with other types of ACRs, suggesting that these dual-modified ACRs possess relatively elevated activity in regulating gene expression.

Transcription factor (TF) binding motifs (TFBMs) are usually enriched at ACRs. To determine TFBMs binding preferences, ATAC-seq footprint[5] analysis was employed (Supporting Information Fig. S6A). In total, 366 TFBMs corresponding to 776 TFs in 43 families were used for footprint analysis, and a total of 190,490 binding events were identified in the APA genome (Supporting Information Table S6), facilitating a comprehensive understanding of how TF binding to ACRs on a genome-wide scale. For example, 412 and 515 binding sites were identified for MP00290 (CCGTAC-like element) and MP00630 (GGATA-like element), respectively (Fig. S6B). Furthermore, based on the footprint results, a potential regulatory network focused on the andrographolide biosynthesis-related genes (Fig. 1F; Fig. S1; Supporting Information Table S7) was constructed. For instance, Apv3LG19G105.53 (an AACT gene) was showed to be regulated by multiple TFs, such as GRAS, Trihelix and TCP (Fig. S6C and D), providing insights into how TF regulates the andrographolide biosynthesis. In addition, compared to *Arabidopsis*[6] and soybean (*Glycine max*)[7], conserved TFBMs (*e.g.*, MP00208 and MP00426) within ACRs associated with homologous genes (GGPPS and MYB) were identified (Fig. 1G; Supporting Information Table S8), potentially supporting a conserved role for TFBMs in the regulating gene expression.

Enhancer-like elements (ELEs), essential *cis*-regulatory elements in gene activation, can be predicted from ATAC-seq data. However, the studies about the involvement of ELEs in secondary metabolism remain limited. In the APA genome, we first predicted a total of 9649 ELEs (Supporting Information Table S9). Most of them were 1 kb in width (Supporting Information Fig. S7A), with some spanning several kilobases of genomic region, such as ACR29879 and ACR24327 (Supporting Information Fig. S8), potentially acting as super enhancers. Moreover, most ELEs located within approximately 25 kb of the transcription start site (TSS) (Fig. S7B). In addition, unlike the proximal ACRs, most ELEs (74.7%) lacked histone modifications (*e.g.,* H3K27me3, H3K27ac and H3K4me3) (Fig. S7C). Among the modified ELEs, about 86 % exhibited the H3K27me3 modification and the remainder did not (Fig. S7C), suggesting that H3K27me3 is the predominant histone modification for these modified ELEs. However, the gene expression of H3K27me3-modified ELEs was significantly lower than those without H3K27me3 modifications (Fig. S7D), indicating that H3K27me3 modification may compromise ELE activity. Given that ELEs can mediate the formation of chromatin loops to regulate gene expression[8], we selected a candidate gene Apv3LG22G84.16, an HDR gene involved in the andrographolide biosynthesis (Fig. S1) which associated multiple ELEs in the downstream regions spanning several kilobases for further examination via 3C-qPCR assay. As the results showed, chromatin interactions such as AP1-R1, AP1-R2, AP1-R3, and AP1-R5, supported these ELEs might regulate the HDR gene expression in a chromatin interaction manner. Interestingly, we also observed that these ELEs associated with another gene, Apv3LG22G84.34 (a PDV protein) proximal to the HDR gene, but with significantly lower expression (FPKM value ~7.2) compared to the HDR gene (FPKM value ~65). Subsequently, we performed the 3C-qPCR using the AP2 as well as the F1−F6 primers (Fig. 1H). The results indicated that only AP2-F6 and AP2-F3 interactions associated with Apv3LG22G84.34 were identified (Fig. 1H). Correspondingly, the 3C-qPCR signal was comparatively weaker than that of the HDR gene, supporting the notion that loop strength may influence downstream gene activation.

Together, this study provides a genomic landscape of ACR features in APA, thereby advancing our further understanding of the functions of such important non-coding regulatory elements in medicinal plants.

Data availability

The RNA-seq raw data in this study have been submitted to NCBI under the accession: SRX12305817, SRX12305811 and SRX12305810. The ATAC-seq and histone ChIP-seq are submitted to National center for bioinformation database (https://ngdc.cncb.ac.cn) under the accession number PRJCA026190.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (82260745), Jiangxi Provincial Natural Science Foundation (20232BAB216120, China) and Jiangxi Province Major Discipline Academic and Technical Leaders Training Program−Leading Talents Project (20225BCJ22018, China).

Author contributions

Mingkun Huang: Formal analysis, Data curation, Conceptualization. Yufang Hu: Formal analysis, Data curation. Ling Zhang: Formal analysis, Data curation. Hua Yang: Formal analysis. Chen Feng: Validation. Chunhong Jiang: Validation, Writing − review & editing. Ning Xie: Validation, Writing − review & editing. Difa Liu: Validation, Writing − review & editing. Shilin Chen: Writing − review & editing, Conceptualization. Jihua Wang: Writing − review & editing, Writing − original draft. Wei Sun: Writing − review & editing, Writing − original draft, Conceptualization.

Conflicts of interest

The authors declared no conflict of interest.

3C-qPCR. Genes, including Apv3LG22G84.16 and Apv3LG22G84.34 were associated with several enhancer like elements (indicated by the red box) in the distal intergenic regions. AP1 and AP2 located close to the genes were used as the anchor primers. Chromatin loops like AP1-R1, AP1-R2, AP1-R3 and AP1-R5 as well as AP2-F6 and AP2-F3 were detected by 3C-qPCR. (Right panel) The relative 3C-qPCR signal in AP1-R1, AP1-R2, AP1-R3 and AP1-R5 as well as AP2-F6 and AP2-F3. Primers, AP1-CR and AP2-CF served as the internal controls. Genomic DNA was used as the control template. The relative DNA interaction signal was further normalized to the amplification signal of AP1-CR or AP2-CF. Asterisks (*) indicate significant 3C signals compared with the internal control ($P < 0.05$, by Student's t test).

Appendix A. Supporting information

Supporting information to this article can be found online at https://doi.org/10.1016/j.apsb.2024.06.012.

References

1. Sun W, Xu Z, Song C, Chen S. Herbgenomics: decipher molecular genetics of medicinal plants. *Innov* 2022;**3**:100322.
2. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57−74.
3. Sun W, Leng L, Yin Q, Xu M, Huang M, Xu Z, et al. The genome of the medicinal plant Andrographis paniculata provides insight into the biosynthesis of the bioactive diterpenoid neoandrographolide. *Plant J* 2019;**97**:841−57.
4. Zhang L, Yung WS, Hu Y, Wang L, Sun W, Huang M. Establishment of a convenient ChIP-seq protocol for identification of the histone modification regions in the medicinal plant *Andrographis paniculata*. *Med Plant Biol* 2023;**2**:6.
5. Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun* 2020;**11**:4267.
6. Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ. The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat Plants* 2019;**5**:1250−9.
7. Huang M, Zhang L, Zhou L, Yung WS, Wang Z, Xiao Z, et al. Identification of the accessible chromatin regions in six tissues in the soybean. *Genomics* 2022;**114**:110364.
8. Huang M, Zhang L, Yung WS, Hu Y, Wang Z, Li MW, et al. Molecular evidence for enhancer-promoter interactions in light responses of soybean seedlings. *Plant Physiol* 2023;**193**:2287−91.

Mingkun Huang[a,b,*], Yufang Hu[a], Ling Zhang[a], Hua Yang[a], Chen Feng[a], Chunhong Jiang[f], Ning Xie[f], Difa Liu[f], Shilin Chen[b,d], Jihua Wang[e,*], Wei Sun[b,c,*]

[a]*Jiangxi Provincial Key Laboratory of Ex Situ Plant Conservation and Utilization, Lushan Botanical Garden, Chinese Academy of Sciences, Jiujiang 332900, China*

[b]*Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, China Academy of Chinese Medical Sciences, Institute of Chinese Materia Medica, Beijing 100070, China*

[c]*State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China*

[d]*Institute of Herbgenomics, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China*

[e]*Guangdong Provincial Key Laboratory of Crops Genetics and Improvement, Crop Research Institute, Guangdong Academy of Agriculture Sciences, Guangzhou 510640, China*

[f]*State Key Laboratory of Innovative Natural Medicine and TCM Injections, Jiangxi Qingfeng Pharmaceutical Co. Ltd., Ganzhou 341008, China*

*Corresponding authors.
E-mail addresses: huangmk@lsbg.cn (Mingkun Huang), wangjihua@gdaas.cn (Jihua Wang), wsun@icmm.ac.cn (Wei Sun)

Received 24 April 2024
Received in revised form 31 May 2024
Accepted 13 June 2024