scientific data

Check for updates

OPEN

DATA DESCRIPTOR

An Integrated Database for Exploring Alternative Promoters in Animals

Feiyang Xue[1,6], Yuqin Yan[1,6], Weiwei Jin[1,6], Haotian Zhu[1], Yanbo Yang[1], Zhanhui Yu[1], Xuewen Xu[2]✉, Jing Gong[1,3]✉ & Xiaohui Niu[1,4,5]✉

Alternative promoter (AP) events, as a major pre-transcriptional mechanism, can initiate different transcription start sites to generate distinct mRNA isoforms and regulate their expression. At present, hundreds of thousands of APs have been identified across human tissues, and a considerable number of APs have been demonstrated to be associated with complex traits and diseases. Recent researches have also proven important effects of APs on animals. However, the landscape of APs in animals has not been fully recognized. In this study, 102,349 AP profiles from 23,077 samples across 12 species were systematically characterized. We further identified tissue-specific APs and investigated trait-related promoters among various species. In addition, we analyzed the associations between APs and enhancer RNAs (eRNA)/transcription factors (TF) as a means of identifying potential regulatory factors. Integrating these findings, we finally developed Animal-APdb, a database for the searching, browsing, and downloading of information related to Animal APs. Animal-APdb is expected to serve as a valuable resource for exploring the functions and mechanisms of APs in animals.

Background & Summary

Promoters, as cis-regulatory elements located upstream of genes' transcription start sites (TSSs), are fundamental in gene regulation[1]. Over half of all human genes possess multiple promoters, referred to as alternative promoters (APs)[2]. Therefore, AP events, as a major pre-transcriptional mechanism, contribute to the generation of various 5' untranslated regions and first exons[3], thereby enriching the diversity of mRNA and protein isoforms. Additionally, some studies have demonstrated that the selection of APs can differ across various tissues, developmental stages[2,4], and the process of cellular differentiation[5]. For instance, the selection of APs in CCND1 can change during the development of retinal cells[6]. Furthermore, increasing evidence also shows that AP events may lead to a range of diseases, especially cancers[2]. For example, the use of a specific AP in acetyl-CoA synthetase 2 (ACSS2) generates *ACSS2-S2*, which is associated with amplified ribosome biogenesis in hepatocellular carcinoma (HCC)[7]. In pan-cancer studies, AP events were also found to display cancer-specific regulation, and AP usage was significantly associated with patient survival outcomes[8].

Besides humans, APs also play a vital role in other eukaryotic animals. For instance, it has been observed that the different isoforms, because of AP events in Rbfox1 within the mouse brain, serve distinct functions during cortical development[9]. Furthermore, the study conducted by Damir *et al.* on cis-regulatory elements in zebrafish revealed that signal transduction-associated genes with APs exhibit vertebrate conservation[10]. Recently, Alfonso-Gonzalez *et al.* also found that in Drosophila heads, 3′ end site choice is globally influenced by AP events[11]. Moreover, AP events in KATNAL1 have been proven to be associated with the reproductive traits of male bulls[12]. Overall, in animals, AP events are also essential in pre-transcriptional regulation, possess important biological functions and are associated with some important traits.

[1]Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, 430070, China. [2]Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Education & College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, 430070, China. [3]College of Biomedicine and Health, Huazhong Agricultural University, Wuhan, 430070, China. [4]Key Laboratory of Smart Farming for Agricultural Animals, Ministry of Agriculture and Rural Affairs, Wuhan, 430070, China. [5]Hubei Engineering Technology Research Center of Agricultural Big Data, Wuhan, 430070, China. [6]These authors contributed equally: Feiyang Xue, Yuqin Yan, Weiwei Jin. ✉e-mail: xuewen_xu@mail.hzau.edu.cn; gong.jing@mail.hzau.edu.cn; niuxiaoh@mail.hzau.edu.cn

Species	Source	No. of samples	No. of tissues
Gallus gallus (Chicken)	NCBI SRA	656	15
Bos taurus (Cow)	NCBI SRA[39–42]	794	28
Canis familiaris (Dog)	NCBI SRA[43,44]	289	34
Xenopus tropicalis (Frog)	NCBI SRA[45–47]	284	1
Drosophila melanogaster (Fruitfly)	NCBI SRA[48–50], EBI	774	3
Homo sapiens (Human)	GTEx V8	16 563	48
Mus musculus (Mouse)	NCBI SRA[51–55]	1 235	37
Sus scrofa (Pig)	NCBI SRA[56–64]	808	24
Rattus norvegicus (Rat)	NCBI SRA[65,66]	901	16
Macaca mulatta (Rhesus)	NCBI SRA[67–70]	257	13
Caenorhabditis elegans (Worm)	NCBI SRA[71–73]	317	1
Danio rerio (Zebrafish)	NCBI SRA[53,74,75]	199	7

Table 1. Samples summary in Animal-APdb. Summary: The table provides an overview of the sample distribution across 12 species in Animal-APdb, including the total number of samples and the number of distinct tissues represented. And the detailed information of the Bioproject IDs utilized in Animal-APdb is provided in Supplementary Table 1.

Regarding the potential regulators of AP events, it has been shown that AP events could be regulated by cis-acting elements and trans-acting factors. Among them, enhancers, as important cis-acting elements, can form a loop structure with the target promoter and are involved in the recruitment of TFs and cofactors, thus regulating AP events[13,14]. Additionally, TFs, as important trans-acting factors, can recognize TF motifs in the flanking regions of TSSs and activate or inhibit transcription initiation[15,16]. Furthermore, DNA methylation, as an important epigenetic modification, is enriched in the promoter region and affects the selection of APs[17]. For example, in the human mammary gland, the overexpression of the TF *Ets-1* activates the AP events of the lactoferrin gene[18].

To date, several technologies can be utilized to identify promoters with the development of high-throughput sequencing technology, such as cap analysis of gene expression (CAGE-seq)[19], rapid amplification of 5' complementary DNA ends (5' RACE) and RNA annotation and mapping of promoters for analysis of gene expression (RAMPAGE)[20]. These approaches involve elaborate experimental procedures and are not as routinely used as RNA-seq. In contrast, RNA-seq data for diverse organisms, tissues, and cell types are relatively easy to produce and are plentifully available in public repositories. While detecting alternative promoters with RNA-seq data has lower sensitivity compared to other techniques, the availability of relatively abundant data and cost-effectiveness make it a viable approach to investigate AP events at the genome-wide level across multiple tissues and various animal species using RNA sequencing. Hence, several algorithms have been developed to identify alternative promoters with RNA-seq data, such as SEASTAR[21], *proActiv*[8] and mountClimber[22].

Considering the significance of APs, numerous AP events have been detected in multiple human tissues, and relevant datasets have been constructed. For example, Demircioğlu *et al.* estimated promoter activity using RNA-seq data from 18,468 cancer and normal samples and found that AP events show obvious tissue-specific regulation and association with patients' prognosis[8]. The Eukaryotic Promoter Database (EPD) has collected experimentally validated promoters for model organisms and also includes some alternative promoters. However, EPD does not focus on the APs and only includes limited APs[23]. Hence, the landscape of alternative promoters in animals other than humans has not been fully explored, and thus far, no database provides information on potential regulators of APs for animals.

Moreover, considering the dataset with 6,674 human normal samples included in Demircioğlu's study was GTEx v7. The updated GTEx dataset with many more samples was also included in our study. Therefore, in this study, we systematically characterized the AP profiles in 23,077 samples from 12 animal species, including human, by analyzing RNA-seq data sourced from publicly available databases. These species include chicken (*Gallus gallus*), cow (*Bos taurus*), dog (*Canis familiaris*), frog (*Xenopus tropicalis*), fruitfly (*Drosophila melanogaster*), human (Homo sapiens), mouse (*Mus musculus*), pig (*Sus scrofa*), rat (*Rattus norvegicus*), rhesus (*Macaca mulatta*), worm (*Caenorhabditis elegans*), and zebrafish (*Danio rerio*). Then, we analyzed the associations between alternative promoters and different animal traits, such as age and sex, to identify potential trait-related AP events. Moreover, putative AP regulators, including TFs and eRNAs, were identified. Finally, we developed Animal-APdb, a database for browsing, searching, and downloading animal AP-related information.

Methods

Collection and processing of data and identification of AP events. The aligned RNA-seq data of human normal tissues were downloaded from the GTEx[24] (version: 8) (Table 1). Moreover, we downloaded the RNA-seq data from normal tissue samples of other animals by accessing the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) of the National Center for Biotechnology Information (NCBI) and EMBL's European Bioinformatics Institute (EBI)[25–27] (Table 1). Detailed sample information, including tissue type, age, sex, and developmental stage, was also downloaded and manually curated. The raw SRA files of RNA-seq data were processed as follows: firstly, they were converted into FASTQ format, and subjected to quality control using FastQC (version: v0.11.8). Subsequently, data cleaning was performed using Trim Galore, followed by alignment

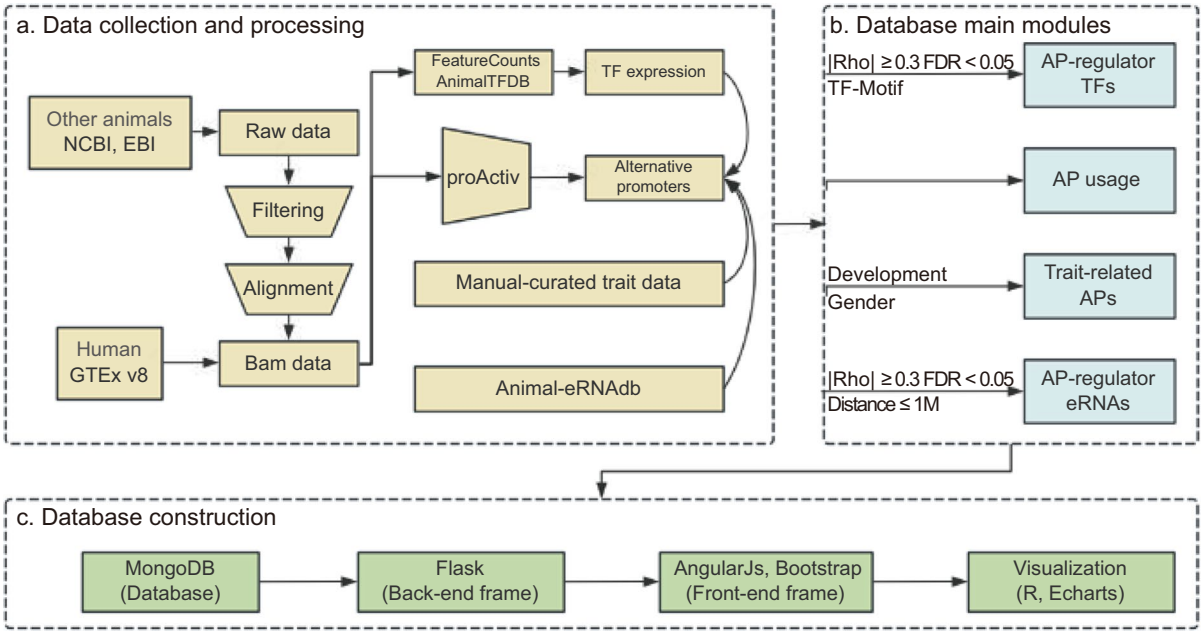

Fig. 1 Flow charts of Animal-APdb. (**a**) Data collection and processing of Animal-APdb. (**b**) Main modules of Animal-APdb. (**c**) Database construction of Animal-APdb.

to the respective reference genome with HISAT2[28]. In addition, we calculated the gene-level read counts with FeatureCounts and employed transcripts per million (TPM) normalization for gene expression (Fig. 1a).

In total, 23,077 samples across 227 tissues of 12 species were included in Animal-APdb, ranging from 199 samples in zebrafish to 16,563 samples in human (Table 1) and from one tissue in frogs to 48 tissues in human.

Based on the collected RNA-seq data, the R package *proActiv*[8] was utilized to identify possible APs in each sample and quantify promoter activity (Fig. 1a). Briefly, *proActiv* is an algorithm that estimates promoter activity based on RNA short-read sequencing data by mapping and quantifying first intron junctions of the genome. *ProActiv* has shown high performance in promoter activity estimates[29,30], as well as higher consistency with H3K4me3 histone data compared with other methods[8].

Specifically, for a promoter in a sample , using *proActiv*, we obtained each promoter's absolute activity and relative usage , as the ratio of its individual activity to the cumulative activity of the same gene's promoters:

$$U_{p,s} = \frac{A_{p,s}}{\sum_{p \in P'} A_{p,s}}$$

Here, and are the usage and absolute activity of promoter of sample, respectively, and denotes the set of promoters belonging to the same gene. Compared with absolute activity, promoter usage can better represent the frequency of the selection of the specific AP, and to some extent, promoter usage helps minimize the batch effects. Hence, we mainly applied promoter usage in this study.

Identification of tissue-specific AP events. In this study, we identified tissue-specific APs with Demircioğlu's method[8]. Tissue-specific alternative promoters were identified by applying a tissue-specific linear model, where each sample was tested for absolute promoter activity and relative usage. A promoter was considered tissue-specific if it met a Benjamini-Hochberg adjusted p-value threshold ($\leq$0.05) for both absolute activity and relative usage, with specific fold-change requirements to distinguish promoter activity from gene expression differences. These criteria ensured that tissue-specific promoter activity was significant, with at least a 2-fold change in activity between the target tissue and others, and minimal changes in overall gene expression.

Identification of trait-related AP events. The trait data of human which contains sex, height, weight and age was collected from GTEx. And trait data of other animals which contains sex, height, weight and development stage information for each animal sample in Animal-APdb was retrieved from SRA. We analyzed the association between the usage of individual AP and each trait across diverse tissues.

(1) For the trait of sex, the 'Mann–Whitney U test' was utilized to compare the difference in AP usage between the male and female groups. To establish statistical significance, we set the criteria at |fold change (FC)| $\geq$ 1.5 and a false discovery rate (FDR) $<$ 0.05.
(2) For the trait of developmental stage, in human samples, the Spearman's correlation would be applied to evaluate the association between AP usage and the age of the samples. We consider the correlation

Species	No. of trait-related APs	Regulators (eRNAs)		Regulators (TFs)	
		eRNAs	eRNA-related APs	TFs	TF-related APs
Gallus gallus (Chicken)	1 701	13 971	3 839	405	4 667
Xenopus tropicalis (Frog)	758	224	311	178	1 489
Drosophila melanogaster (Fruitfly)	634	480	599	88	2 153
Mus musculus (Mouse)	6 687	31 671	9 774	610	12 815
Rattus norvegicus (Rat)	1 223	7 291	1 555	534	1 883
Macaca mulatta (Rhesus)	1 229	3 441	1 906	493	7 007
Caenorhabditis elegans (Worm)	67	380	304	54	408
Danio rerio (Zebrafish)	5	6 396	1 525	246	2 154
Bos taurus (Cow)	41	—	—	454	3 600
Canis familiaris (Dog)	82	—	—	464	5 481
Homo sapiens (Human)	855	—	—	572	29 412
Sus scrofa (Pig)	58	—	—	475	4 126

Table 2. Data summary of Related Aps. -: There is no eRNA data for cow, dog, human, and pig in animal-eRNAdb Summary: The table provides a comprehensive overview of APs related to specific traits, along with their regulatory elements (eRNAs and TFs) across 12 species. It highlights the diversity and scale of trait-associated APs and their regulators, with some species lacking eRNA data.

with $|Rho| \geq 0.3$ and $FDR < 0.05$ as statistically significant. For other animal samples, all tissue samples were categorized into two categories: tissues with both embryo and postnatal samples, and the tissues with either embryo or postnatal samples exclusively. With regard to tissues with only embryo or postnatal samples, the Spearman's correlation would be applied, using developmental index as a numerical variable, to evaluate the association between AP usage and the developmental index. Besides, if the development index was a dichotomous variable, the significance level of difference in AP usage between two groups would be evaluated with the 'Mann–Whitney U test'. As for tissues with both embryo and postnatal samples, firstly, we utilized the 'Mann–Whitney U test' to detect the APs whose usage is significantly different between the embryo and postnatal groups. Secondly, the same methods as above were utilized to identify development-related APs in embryo and postnatal samples, respectively (Fig. 1b).

Identification of eRNAs related to AP events. Here, we used enhancer RNA (eRNA) data, a kind of non-coding RNA molecule transcribed from the loci of enhancers and whose expression can characterize the activity of the corresponding enhancer[31], to calculate the associations between enhancer activities and AP events. We downloaded the locus and expression data of eRNAs from Animal-eRNAdb (http://gong_lab.hzau.edu.cn/Animal-eRNAdb/)[32]. Putative enhancer RNAs (eRNAs), presumed to regulate (AP) events, located within 1 Mb of the target AP, and their expressions showed significant associations with the target AP usage. (Spearman's correlation coefficient $|Rho| \geq 0.3$ and $FDR < 0.05$) (Fig. 1b).

We identified a total of 19,813AP events related to 63,854 eRNAs (ranging from 304 AP events related to 380 eRNAs in worms to 9,774 AP events related to 31,671 eRNAs in mice). More detailed information is presented in Table 2.

Identification of TFs related to AP events. TFs can recognize their corresponding motifs in the flanking region of the TSS and activate or inhibit transcription initiation. To obtain TFs related to AP events, annotations of TFs were retrieved from AnimalTFDB (http://bioinfo.life.hust.edu.cn/AnimalTFDB4/#/)[33], and the known TF motifs were collected from JASPAR (https://jaspar.genereg.net/)[34]. Combined with gene expression data, we identified candidate TFs related to AP events according to two major criteria: 1) TF expression had significant associations with AP usage and 2) TF might bind the flanking region of the TSS (from 2,000 bp upstream to 500 bp downstream of the TSS). Specifically, firstly, average TPM of TF expression > 5 in each tissue and TF expression had significant association with AP usage (Spearman's correlation coefficient $|Rho| \geq 0.3$ and $FDR < 0.05$); secondly, two methods were adopted in this study to validate whether specific TF could bind to the flanking region of the TSS. One method was using FIMO[35] to scan TFBS motifs in the vicinity of each AP. Another method was adopting uniformly processed ChIP-seq data of specific TFs to overlap with the flanking region of the TSS. A total of 9,675 uniformly processed ChIP-seq data from 32 tissues of 6 species were collected from ChIP-Atlas[36]. Finally, the results were combined into the database.

Database framework. All data mentioned above were stored in the MongoDB database (version 3.6.8). The Animal-APdb website was built based on the Flask (version 1.0.3) framework with AngularJS (version 1.6.1) and Bootstrap, hosted on the Apache 2 webserver (version 2.4.18). In addition, ECharts and R are employed for database visualization. Animal-APdb is freely available online without registration or login for access (Fig. 1c).

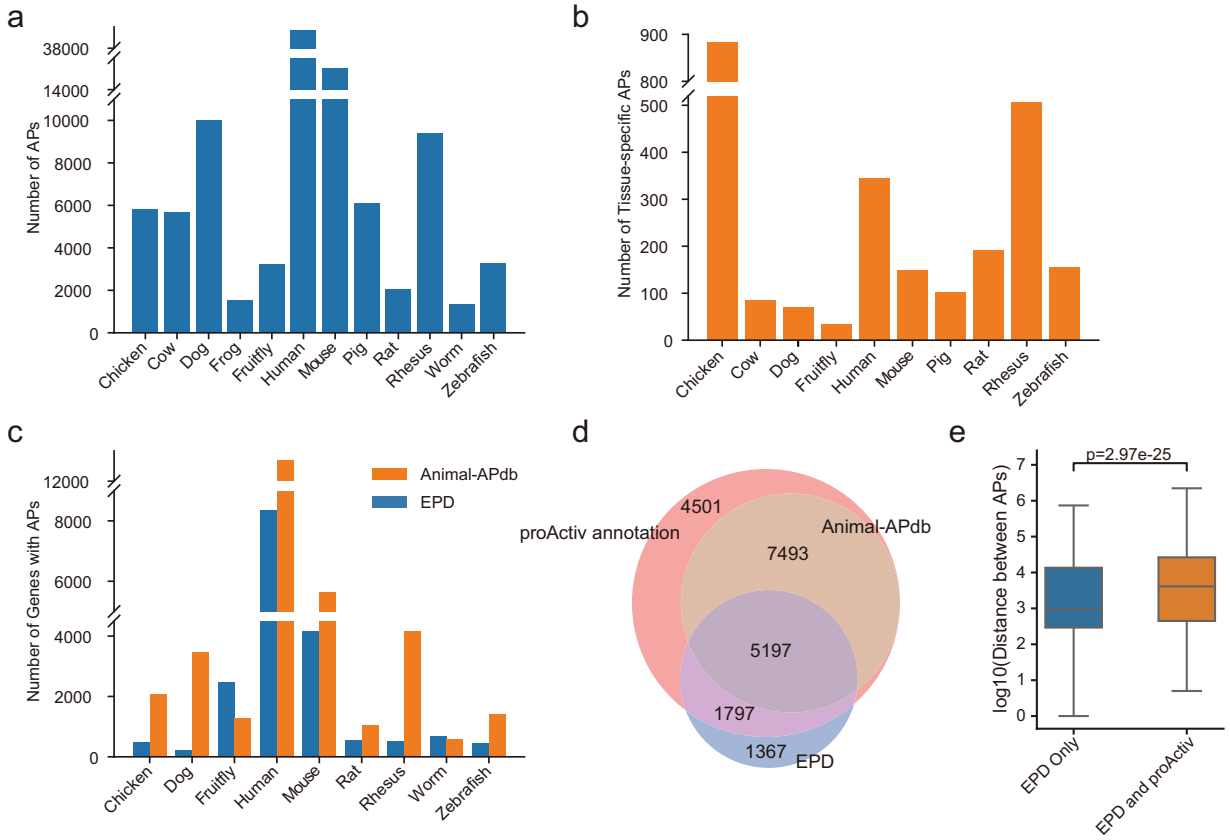

Fig. 2 Data summary and technical validation of Animal-APdb. (**a**) The number of APs identified for each species in Animal-APdb. (**b**) The number of tissue-specific APs identified for each species in Animal-APdb. (**c**) The total number of AP genes annotated in Animal-APdb compared to those annotated in EPD. (**d**) Comparison of human AP genes annotated by EPD, *proActiv*, and Animal-APdb. (**e**) Distribution of distances between APs for genes annotated exclusively in EPD and those annotated in both EPD and *proActiv*.

Data records

These datasets are available on Figshare[37], Zenodo[38], and the Animal-APdb download page (http://gong_lab.hzau.edu.cn/Animal_AP#!/download). Each module file for each species is provided in '.tsv' format. Files on AP usage offer detailed information about APs across multiple tissues for specific species. Trait-related AP files provide data on the correlation between APs and various traits across tissues. Regulator files include detailed information on eRNAs and TFs potentially involved in AP selection.

Technical Validation

All results mentioned above have been integrated into Animal-APdb. A summary of data entry can be found in Fig. 2 and Table 2.

Data summary of Animal-APdb. As shown in Fig. 2a, a total of 102,349 AP events in these species, ranging from 1,346 in worms to 38,849 in human at the species level. Many AP events' expressions vary a lot in multiple tissues, which corroborates previous research[2]. Notably, the number of AP events of each species related with the number of samples, genome complexity and the number of tissue types. Moreover, a total of 2,523 tissue-specific AP events were identified in species with two or more tissues, ranging from 34 in fruitfly to 884 in chicken (Fig. 2b).

A total of 13,340 trait-related AP events in all species (ranging from 5 in zebrafish to 6,687 in mouse) were identified. More detailed information is presented in Table 2.

We identified a total of 19,813 AP events related to 63,854 eRNAs in 8 species (ranging from 304 AP events related to 380 eRNAs in worm to 9,774 AP events related to 31,671 eRNAs in mouse). Moreover, a total of 75,195 AP events associated with 4,573 TFs in all 12 species (from 408 AP events associated with 54 TFs in worm to 29,412 AP events associated with 572 TFs in human). More detailed information is presented in Table 2.

Technical validation process of Animal-APdb. To ensure the quality and validity of the data in Animal-APdb, several rigorous steps were implemented during curation. First, the meta-information for all species was manually curated from the NCBI SRA database and GTEx to guarantee accuracy and reliability. To address potential batch effects between RNA-seq data from different BioProjects, BioProjects with insufficient data were excluded, thereby maintaining the integrity and consistency of the dataset. During RNA-seq processing,

stringent quality control measures were applied to remove samples with poor sequencing quality. Filtering and alignment procedures were meticulously carried out to retain only high-quality data for downstream analyses.

Second, the R package *proActiv* was employed to identify alternative promoters and estimate their activities. The reliability of *proActiv* in estimating promoter activities has been validated using H3K4me3 histone modification data, CAGE-seq data, and Iso-seq data[29]. To ensure biological relevance, promoters with low activity, which are unlikely to have significant functional implications, were excluded from certain tissues and species. These steps collectively contribute to a robust and high-quality dataset that underpins the Animal-APdb resource. The annotation quality of APs in Animal-APdb was validated by comparing it with experimentally verified promoters in the EPD database. For most species, Animal-APdb contains a much greater number of genes with APs compared to EPD (Fig. 2c). However, it is important to note that some discrepancies arise due to differences in the reference genome versions used by EPD and Animal-APdb, which could affect the results for certain species.

To further investigate the representation of EPD-annotated genes with APs in Animal-APdb, the case of humans was analyzed as instance (Fig. 2d). Among the 8,361 genes with APs annotated in EPD, 6,994 were also identified by the *proActiv*. This substantial overlap highlights the consistency between the two methods when applied to the same reference genome. However, 1,367 AP genes annotated in EPD were not detected by *proActiv*. This discrepancy arises because *proActiv* categorizes transcripts with identical or closely located TSSs as being regulated by the same promoter. Supporting this, the distances between APs for genes annotated by both EPD and *proActiv* were significantly greater than those for genes annotated only by EPD (Fig. 2e). 4,501 AP genes were excluded due to low promoter activity, reflecting the stringency of the activity-based filtering process. In contrast, EPD-validated AP genes were reduced by only 1,797 in Animal-APdb. These results highlight the efficiency and necessity of the activity-based filtering process.

Usage Notes

The Animal-APdb provides a user-friendly web interface. It contains four main modules: 'AP events', 'Trait', 'eRNA', and 'Transcription Factor' for data searching, browsing, and visualization. To maximize the utility of this resource, users can query genes of interest to identify the presence of alternative promoters in specific species and tissues. This capability enables further investigation into how APs influence associated traits and the factors regulating the selection of APs.

Additionally, the database facilitates advanced data mining by integrating information across multiple species. This integration allows researchers to explore the relationship between APs' usage and species evolution, shedding light on how promoter variation may have evolved in different species. Furthermore, the inclusion of multi-omics data enables the identification of regulatory factors that drive APs' usage in key genes across species which offer a powerful framework for dissecting gene regulatory networks.

Code availability

The source code of the data processing of Animal-APdb has been shared on GitHub (https://github.com/flysheeeep/Animal-APdb/).

Received: 1 July 2024; Accepted: 28 January 2025;
Published online: 07 February 2025

References

1. Ayoubi, T. A. & Van De Ven, W. J. Regulation of gene expression by alternative promoters. *FASEB J* **10**, 453–460 (1996).
2. Davuluri, R. V., Suzuki, Y., Sugano, S., Plass, C. & Huang, T. H. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet* **24**, 167–177, https://doi.org/10.1016/j.tig.2008.01.008 (2008).
3. Bieberstein, N. I., Carrillo Oesterreich, F., Straube, K. & Neugebauer, K. M. First exon length controls active chromatin signatures and transcription. *Cell Rep* **2**, 62–68, https://doi.org/10.1016/j.celrep.2012.05.019 (2012).
4. Schibler, U. & Sierra, F. Alternative promoters in developmental gene expression. *Annu Rev Genet* **21**, 237–257, https://doi.org/10.1146/annurev.ge.21.120187.001321 (1987).
5. Maqbool, M. A. *et al.* Alternative Enhancer Usage and Targeted Polycomb Marking Hallmark Promoter Choice during T Cell Differentiation. *Cell Rep* **32**, 108048, https://doi.org/10.1016/j.celrep.2020.108048 (2020).
6. Hu, Y. *et al.* Single-cell RNA cap and tail sequencing (scRCAT-seq) reveals subtype-specific isoforms differing in transcript demarcation. *Nat Commun* **11**, 5148, https://doi.org/10.1038/s41467-020-18976-7 (2020).
7. Wang, Y. H. *et al.* Alternative transcription start site selection in ACSS2 controls its nuclear localization and promotes ribosome biosynthesis in hepatocellular carcinoma. *Biochem Biophys Res Commun* **514**, 632–638, https://doi.org/10.1016/j.bbrc.2019.04.193 (2019).
8. Demircioglu, D. *et al.* A Pan-cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. *Cell* **178**, 1465–1477 e1417, https://doi.org/10.1016/j.cell.2019.08.018 (2019).
9. Casanovas, S. *et al.* Rbfox1 Is Expressed in the Mouse Brain in the Form of Multiple Transcript Variants and Contains Functional E Boxes in Its Alternative Promoters. *Front Mol Neurosci* **13**, 66, https://doi.org/10.3389/fnmol.2020.00066 (2020).
10. Baranasic, D. *et al.* Multiomic atlas with functional stratification and developmental dynamics of zebrafish cis-regulatory elements. *Nat Genet* **54**, 1037–1050, https://doi.org/10.1038/s41588-022-01089-w (2022).
11. Alfonso-Gonzalez, C. *et al.* Sites of transcription initiation drive mRNA isoform selection. *Cell* **186**, 2438–2455, https://doi.org/10.1016/j.cell.2023.04.012 (2023).
12. Zhang, X. *et al.* Association between an alternative promoter polymorphism and sperm deformity rate is due to modulation of the expression of KATNAL1 transcripts in Chinese Holstein bulls. *Anim Genet* **45**, 641–651, https://doi.org/10.1111/age.12182 (2014).
13. Wang, J., Zhang, S., Lu, H. & Xu, H. Differential regulation of alternative promoters emerges from unified kinetics of enhancer-promoter interaction. *Nat Commun* **13**, 2714, https://doi.org/10.1038/s41467-022-30315-6 (2022).
14. Hah, N., Murakami, S., Nagari, A., Danko, C. G. & Kraus, W. L. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res* **23**, 1210–1223, https://doi.org/10.1101/gr.152306.112 (2013).
15. Wang, Z. *et al.* An autoimmune pleiotropic SNP modulates IRF5 alternative promoter usage through ZBTB3-mediated chromatin looping. *Nat Commun* **14**, 1208, https://doi.org/10.1038/s41467-023-36897-z (2023).
16. Cheng, C. *et al.* Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Res* **22**, 1658–1667, https://doi.org/10.1101/gr.136838.111 (2012).

17. de Mendoza, A. *et al*. Large-scale manipulation of promoter DNA methylation reveals context-specific transcriptional responses and stability. *Genome Biol* **23**, 163, https://doi.org/10.1186/s13059-022-02728-5 (2022).
18. Liu, D., Wang, X., Zhang, Z. & Teng, C. T. An intronic alternative promoter of the human lactoferrin gene is activated by Ets. *Biochem Biophys Res Commun* **301**, 472–479, https://doi.org/10.1016/s0006-291x(02)03077-2 (2003).
19. Shiraki, T. *et al*. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci USA* **100**, 15776–15781, https://doi.org/10.1073/pnas.2136655100 (2003).
20. Batut, P. & Gingeras, T. R. RAMPAGE: promoter activity profiling by paired-end sequencing of 5'-complete cDNAs. *Curr Protoc Mol Biol* **104**, Unit 25B 11, https://doi.org/10.1002/0471142727.mb25b11s104 (2013).
21. Qin, Z., Stoilov, P., Zhang, X. & Xing, Y. SEASTAR: systematic evaluation of alternative transcription start sites in RNA. *Nucleic Acids Res* **46**, e45, https://doi.org/10.1093/nar/gky053 (2018).
22. Cass, A. A. & Xiao, X. mountainClimber Identifies Alternative Transcription Start and Polyadenylation Sites in RNA-Seq. *Cell Syst* **9**, 393–400 e396, https://doi.org/10.1016/j.cels.2019.07.011 (2019).
23. Dreos, R., Ambrosini, G., Perier, R. C. & Bucher, P. The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res* **43**, D92–D96, https://doi.org/10.1093/nar/gku1111 (2015).
24. Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580–585, https://doi.org/10.1038/ng.2653 (2013).
25. Kodama, Y., Shumway, M., Leinonen, R. & International Nucleotide Sequence Database, C. The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res* **40**, D54–56, https://doi.org/10.1093/nar/gkr854 (2012).
26. Sayers, E. W. *et al*. Database resources of the national center for biotechnology information. *Nucleic Acids Res* **50**, D20–D26, https://doi.org/10.1093/nar/gkab1112 (2022).
27. Thakur, M. *et al*. EMBL's European Bioinformatics Institute (EMBL-EBI) in 2022. *Nucleic Acids Res* **51**, D9–D17, https://doi.org/10.1093/nar/gkac1098 (2023).
28. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357–360, https://doi.org/10.1038/nmeth.3317 (2015).
29. Huang, K. K. *et al*. Long-read transcriptome sequencing reveals abundant promoter diversity in distinct molecular subtypes of gastric cancer. *Genome Biol* **22**, 44, https://doi.org/10.1186/s13059-021-02261-x (2021).
30. Sundar, R. *et al*. Epigenetic promoter alterations in GI tumour immune-editing and resistance to immune checkpoint inhibition. *Gut* **71**, 1277–1288, https://doi.org/10.1136/gutjnl-2021-324420 (2022).
31. Sartorelli, V. & Lauberth, S. M. Enhancer RNAs are an important regulatory layer of the epigenome. *Nat Struct Mol Biol* **27**, 521–528, https://doi.org/10.1038/s41594-020-0446-0 (2020).
32. Jin, W. *et al*. Animal-eRNAdb: a comprehensive animal enhancer RNA database. *Nucleic Acids Res* **50**, D46–D53, https://doi.org/10.1093/nar/gkab832 (2022).
33. Shen, W. K. *et al*. AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. *Nucleic Acids Res* **51**, D39–D45, https://doi.org/10.1093/nar/gkac907 (2023).
34. Castro-Mondragon, J. A. *et al*. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **50**, D165–D173, https://doi.org/10.1093/nar/gkab1113 (2022).
35. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018, https://doi.org/10.1093/bioinformatics/btr064 (2011).
36. Zou, Z., Ohta, T., Miura, F. & Oki, S. ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and Bisulfite-seq data. *Nucleic Acids Res* **50**, W175–W182, https://doi.org/10.1093/nar/gkac199 (2022).
37. Xue, F. Animal-APdb: a comprehensive animal alternative promoter database. *figshre* https://doi.org/10.6084/m9.figshare.26130373.v2
38. Xue, F. Animal-APdb: a comprehensive animal alternative promoter database [Data set]. *Zenodo* https://doi.org/10.5281/zenodo.14054379 (2024).
39. de Las Heras-Saldana, S. *et al*. Combining information from genome-wide association and multi-tissue gene expression studies to elucidate factors underlying genetic variation for residual feed intake in Australian Angus cattle. *BMC Genomics* **20**, 939, https://doi.org/10.1186/s12864-019-6270-4 (2019).
40. Liang, G. *et al*. Transcriptome analysis reveals regional and temporal differences in mucosal immune system development in the small intestine of neonatal calves. *BMC Genomics* **17**, 602, https://doi.org/10.1186/s12864-016-2957-y (2016).
41. Malmuthuge, N., Liang, G. & Guan, L. L. Regulation of rumen development in neonatal ruminants through microbial metagenomes and host transcriptomes. *Genome Biol* **20**, 172, https://doi.org/10.1186/s13059-019-1786-0 (2019).
42. Seo, M. *et al*. Comprehensive identification of sexually dimorphic genes in diverse cattle tissues using RNA-seq. *BMC Genomics* **17**, 81, https://doi.org/10.1186/s12864-016-2400-4 (2016).
43. Naqvi, S. *et al*. Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* **365**, https://doi.org/10.1126/science.aaw7317 (2019).
44. Meyers-Wallen, V. N. *et al*. XX Disorder of Sex Development is associated with an insertion on chromosome 9 and downregulation of RSPO1 in dogs (Canis lupus familiaris). *PLoS One* **12**, e0186331, https://doi.org/10.1371/journal.pone.0186331 (2017).
45. Owens, N. D. L. *et al*. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Rep* **14**, 632–647, https://doi.org/10.1016/j.celrep.2015.12.050 (2016).
46. Collart, C. *et al*. High-resolution analysis of gene activity during the Xenopus mid-blastula transition. *Development* **141**, 1927–1939, https://doi.org/10.1242/dev.102012 (2014).
47. Tan, M. H. *et al*. RNA sequencing reveals a diverse and dynamic repertoire of the Xenopus tropicalis transcriptome over development. *Genome Res* **23**, 201–216, https://doi.org/10.1101/gr.141424.112 (2013).
48. Lin, Y., Chen, Z. X., Oliver, B. & Harbison, S. T. Microenvironmental Gene Expression Plasticity Among Individual Drosophila melanogaster. *G3 (Bethesda* **6**, 4197–4210, https://doi.org/10.1534/g3.116.035444 (2016).
49. Lin, Y. *et al*. Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual Drosophila melanogaster. *BMC Genomics* **17**, 28, https://doi.org/10.1186/s12864-015-2353-z (2016).
50. Mahadevaraju, S. *et al*. Dynamic sex chromosome expression in Drosophila male germ cells. *Nat Commun* **12**, 892, https://doi.org/10.1038/s41467-021-20897-y (2021).
51. Weger, B. D. *et al*. The Mouse Microbiome Is Required for Sex-Specific Diurnal Rhythms of Gene Expression and Metabolism. *Cell Metab* **29**, 362–382 e368, https://doi.org/10.1016/j.cmet.2018.09.023 (2019).
52. Terry, E. E. *et al*. Transcriptional profiling reveals extraordinary diversity among skeletal muscle tissues. *Elife* **7**, https://doi.org/10.7554/eLife.34613 (2018).
53. Aramillo Irizar, P. *et al*. Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nat Commun* **9**, 327, https://doi.org/10.1038/s41467-017-02395-2 (2018).
54. Crowley, J. J. *et al*. Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat Genet* **47**, 353–360, https://doi.org/10.1038/ng.3222 (2015).
55. Arpat, N. H., De Matos, A. B. & Gatfield, M. D. MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale. *Elife* **3**, e02510, https://doi.org/10.7554/eLife.02510 (2014).
56. Chen, M. *et al*. Comprehensive Profiles of mRNAs and miRNAs Reveal Molecular Characteristics of Multiple Organ Physiologies and Development in Pigs. *Front Genet* **10**, 756, https://doi.org/10.3389/fgene.2019.00756 (2019).
57. Keel, B. N. *et al*. Using SNP Weights Derived From Gene Expression Modules to Improve GWAS Power for Feed Efficiency in Pigs. *Front Genet* **10**, 1339, https://doi.org/10.3389/fgene.2019.01339 (2019).

58. Li, M. *et al*. Comprehensive variation discovery and recovery of missing sequence in the pig genome using multiple de novo assemblies. *Genome Res* **27**, 865–874, https://doi.org/10.1101/gr.207456.116 (2017).
59. Li, Y. *et al*. Genome-wide differential expression of genes and small RNAs in testis of two different porcine breeds and at two different ages. *Sci Rep* **6**, 26852, https://doi.org/10.1038/srep26852 (2016).
60. Liu, Y. *et al*. Trait correlated expression combined with eQTL and ASE analyses identified novel candidate genes affecting intramuscular fat. *BMC Genomics* **22**, 805, https://doi.org/10.1186/s12864-021-08141-9 (2021).
61. Perez-Montarelo, D. *et al*. Identification of genes regulating growth and fatness traits in pig through hypothalamic transcriptome analysis. *Physiol Genomics* **46**, 195–206, https://doi.org/10.1152/physiolgenomics.00151.2013 (2014).
62. Veno, M. T. *et al*. Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. *Genome Biol* **16**, 245, https://doi.org/10.1186/s13059-015-0801-3 (2015).
63. Zambonelli, P., Gaffo, E., Zappaterra, M., Bortoluzzi, S. & Davoli, R. Transcriptional profiling of subcutaneous adipose tissue in Italian Large White pigs divergent for backfat thickness. *Anim Genet* **47**, 306–323, https://doi.org/10.1111/age.12413 (2016).
64. Zhang, Y. *et al*. Genome-wide identification of RNA editing in seven porcine tissues by matched DNA and RNA high-throughput sequencing. *J Anim Sci Biotechnol* **10**, 24, https://doi.org/10.1186/s40104-019-0326-9 (2019).
65. Yu, Y. *et al*. A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nat Commun* **5**, 3230, https://doi.org/10.1038/ncomms4230 (2014).
66. Yu, Y. *et al*. Comprehensive RNA-Seq transcriptomic profiling across 11 organs, 4 ages, and 2 sexes of Fischer 344 rats. *Sci Data* **1**, 140013, https://doi.org/10.1038/sdata.2014.13 (2014).
67. Bozek, K. *et al*. Exceptional evolutionary divergence of human muscle and brain metabolomes parallels human cognitive and physical uniqueness. *PLoS Biol* **12**, e1001871, https://doi.org/10.1371/journal.pbio.1001871 (2014).
68. Cross, R. W. *et al*. Comparative Transcriptomics in Ebola Makona-Infected Ferrets, Nonhuman Primates, and Humans. *J Infect Dis* **218**, S486–S495, https://doi.org/10.1093/infdis/jiy455 (2018).
69. Ramaswamy, S. *et al*. The testicular transcriptome associated with spermatogonia differentiation initiated by gonadotrophin stimulation in the juvenile rhesus monkey (Macaca mulatta). *Hum Reprod* **32**, 2088–2100, https://doi.org/10.1093/humrep/dex270 (2017).
70. Rhoads, T. W. *et al*. Caloric Restriction Engages Hepatic RNA Processing Mechanisms in Rhesus Monkeys. *Cell Metab* **27**, 677–688 e675, https://doi.org/10.1016/j.cmet.2018.01.014 (2018).
71. Hendriks, G. J., Gaidatzis, D., Aeschimann, F. & Grosshans, H. Extensive oscillatory gene expression during C. elegans larval development. *Mol Cell* **53**, 380–392, https://doi.org/10.1016/j.molcel.2013.12.013 (2014).
72. Janes, J. *et al*. Chromatin accessibility dynamics across C. elegans development and ageing. *Elife* **7**, https://doi.org/10.7554/eLife.37344 (2018).
73. Hastings, J. *et al*. Multi-Omics and Genome-Scale Modeling Reveal a Metabolic Shift During C. elegans Aging. *Front Mol Biosci* **6**, 2, https://doi.org/10.3389/fmolb.2019.00002 (2019).
74. Johnstone, T. G., Bazzini, A. A. & Giraldez, A. J. Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J* **35**, 706–723, https://doi.org/10.15252/embj.201592759 (2016).
75. Bazzini, A. A. *et al*. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J* **33**, 981–993, https://doi.org/10.1002/embj.201488411 (2014).

Acknowledgements

The work was supported by ST2030-Major Projects (2023ZD0404702 to Xiaohui Niu), the National Natural Science Foundation of China (31970644 to Jing Gong), the Natural Science Foundation of Hubei Province (2021CFB404 to Xiaohui Niu), Fundamental Research Funds for the Central Universities (2662024XXPY002 to GJ), and Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (11041810351 to Jing Gong, 2662022XXYJ008 to Xiaohui Niu).

Author contributions

Xiaohui Niu, Jing Gong and Xuewen Xu designed the project and provided critical advice on the research. Feiyang Xue and Weiwei Jin performed data curation and data processing and database construction. Haotian Zhu, Yanbo Yang and Zhanhui Yu analyzed data for the work. Feiyang Xue drafted the manuscript. Yuqin Yan supplement data analysis and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-04548-1.

Correspondence and requests for materials should be addressed to X.X., J.G. or X.N.