



OPEN ACCESS

EDITED BY

Surendra Pratap Singh,
Chhatrapati Shahu Ji Maharaj University, India

REVIEWED BY

Fengli Zhao,
China National Rice Research Institute, China
Shatrujeet Pandey,
National Botanical Research Institute (CSIR),
India

*CORRESPONDENCE

Junkang Rong
✉ junkangrong@126.com

[†]These authors have contributed equally to
this work

RECEIVED 01 April 2025

ACCEPTED 30 May 2025

PUBLISHED 02 July 2025

CITATION

Xie Y, Yang L, Zhao Z, Ding M, Cao Y, Hu X
and Rong J (2025) Transcriptomic exploration
yields novel perspectives on the regulatory
network underlying trichome initiation in
Gossypium arboreum hypocotyl.
Front. Plant Sci. 16:1604186.
doi: 10.3389/fpls.2025.1604186

COPYRIGHT

© 2025 Xie, Yang, Zhao, Ding, Cao, Hu and
Rong. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums
is permitted, provided the original author(s)
and the copyright owner(s) are credited and
that the original publication in this journal is
cited, in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Transcriptomic exploration yields novel perspectives on the regulatory network underlying trichome initiation in *Gossypium arboreum* hypocotyl

Yuxing Xie[†], Luying Yang[†], Zewei Zhao, Mingquan Ding,
Yuefen Cao, Xin Hu and Junkang Rong*

College of Advanced Agricultural Sciences, Zhejiang A&F University, Linan, Hangzhou, Zhejiang, China

Trichomes play a crucial role in plant stress tolerance and serve as an excellent model for studying epidermal cell differentiation. To elucidate the molecular mechanisms underlying trichome development in cotton stems, we investigated two *Gossypium arboreum* mutants that exhibit abnormal trichome patterns during hypocotyl growth. Based on morphological characteristics, we classified four developmental stages: preinitiation, initiation, elongation, and maturation. Comparative transcriptome profiling of epidermal cells across these stages identified differentially expressed genes (DEGs) through maSigPro analysis, which revealed that these DEGs were primarily associated with pathways involved in cell wall metabolism. Additionally, integrated weighted gene co-expression network analysis (WGCNA) and Cytoscape analyses identified 20 core regulatory genes from a total of 59 candidates linked to epidermal development. Utilizing three machine learning algorithms (SVM-RFE, Boruta, and LASSO), we consistently prioritized five key regulators: *Ga02G1392 (TBR)*, *Ga03G0474 (OMR1)*, *Ga12G2860 (ACO1)*, *Ga11G2117 (BBX19)*, and *Ga12G2864 (CUE)*. RT-qPCR validation confirmed their stage-specific expression patterns, which were consistent with the RNA-Seq data. Our study establishes a comprehensive framework for research on cotton trichomes and identifies critical genetic components governing epidermal hair development, thereby providing new insights for the molecular breeding of stress-resistant cotton varieties.

KEYWORDS

Gossypium arboreum, trichome initiation development, transcriptome analysis, WGCNA, machine learning

1 Introduction

Trichomes, which are surface hairs that arise from the protrusion of plant epidermal cells, can be functionally categorized into glandular and non-glandular types. Glandular hairs form a chemical barrier by secreting secondary metabolites, while non-glandular hairs enhance plant adaptability to extreme temperatures and biological stress through physical barriers (Ding et al., 2020; Feng et al., 2019). Investigating the molecular regulatory mechanisms underlying the differentiation and development of plant epidermal cells represents a critical area of biological research, with the establishment of an optimal research model serving as a foundational strategy for unraveling these complex biological processes. Since the 1990s, epidermal hairs of *Arabidopsis thaliana* have emerged as a quintessential model system for studying plant cell differentiation and morphogenesis, owing to their unique morphological features and well-defined developmental pathways. Significant progress has been made in understanding aspects such as cell fate determination, cell cycle regulation, cell polarity establishment, and cell shape modulation (Hulskamp, 2004).

In *Arabidopsis*, the morphological changes of epidermal hair cells can be categorized into three stages based on their orderly spatial distribution: the initiation of nuclear replication, the establishment of the polarity axis, and the development of branching morphology (Bouyer et al., 2001). By leveraging the established trichome system, researchers identified that the GLABRA1 (GL1) - GLABRA3 (GL3) - TRANSPARENT TESTA GLABRA1 (TTG1) complex forms a feedback regulatory loop by activating downstream R3 MYB inhibitory factors, which precisely regulates the spatiotemporal specificity of epidermal hair initiation (Schwab et al., 2000). Simultaneously, the C2H2 zinc finger protein GIS plays a crucial role in determining epidermal hair fate by modulating the activity of the MYB-bHLH-WDR complex, thereby unveiling a novel mechanism of synergistic regulation that integrates epigenetic modifications and transcriptional networks (Zhang et al., 2018).

In contrast to the single-cell model, the morphological development of multicellular epidermal hairs, such as those found in cucumber and tomato, entails more intricate processes of cell division and differentiation. In tomato (*Lycopersicon esculentum*), the *Hairless-2* (*HL-2*) gene interacts with the HD-zip IV transcription factor *SIHDZIV8* to modulate the morphology of multicellular epidermal hairs by regulating actin cytoskeleton remodeling and cellulose synthesis (Xie et al., 2020). Similarly, in cucumber (*Cucumis sativus*), the molecular basis of functional differentiation in epidermal hairs has been elucidated; specifically, *CSHOMEBOX3* (*CsHOX3*) plays a role in the elongation of non-glandular hairs, while *CsbHLH1* is implicated in the formation of apical secretory structures within glandular hairs (Dong et al., 2022). Moreover, the receptor-like kinase *CsTM* (*CSA1G056960*) has been shown to establish a direct connection between the morphological development of the surface coat and aphid resistance by integrating mechanical signals and redox homeostasis (Yang et al., 2025).

The single-cell development model established in cotton (Qin et al., 2022) provides a comprehensive analysis of the dynamic regulatory network influencing fiber cell development from primordium initiation (2–5 days post-anthesis, DPA) and polar elongation (3–20 DPA) to secondary wall thickening (16–40 DPA) (Feng et al., 2018). The MBW complex plays a crucial role in this process, with various transcription factors, including the R2R3-type MYB transcription factors *GhMYB109* (Wang L, et al., 2021), *GhMYB25* (Wu et al., 2006), and *GhMYB25-like* (Walford et al., 2011; Wan et al., 2016), as well as the HD-Zip family transcription factors *GhHDI* (Walford et al., 2012) and *GhHOX3* (Shan et al., 2014), positively regulating cotton fiber initiation. Moreover, the NAC-MYB-CESAs module is instrumental in secondary wall synthesis by modulating cellulose synthase activity (Taylor-Teeples et al., 2015; Cao et al., 2020). Additionally, plant hormones have a significant impact on the development of surface hairs, as demonstrated by the overexpression of *GhPIN3*, which enhances fiber-specific auxin (AUX) accumulation (Ma et al., 2019; Zhang et al., 2017b), thereby facilitating fiber initiation and development (Li et al., 2007; Zhang et al., 2017a). Flavanone 3-hydroxylase (F3H)-related products (Naoumkina et al., 2015; Tan et al., 2013) and other flavonoids also play critical roles in fiber development by influencing the distribution, activity, and content of auxin carriers, as well as altering the levels of reactive oxygen species (ROS) and the activity of associated proteins. This intricate interplay ultimately affects the progression of fiber development (Edward, 2013; Silva-Navas et al., 2016; Wang et al., 2015). Collectively, these studies highlight that the formation of unicellular trichomes differs from that of multicellular trichomes and is governed by a diverse array of genes.

Numerous studies have demonstrated that RNA transcriptional regulation plays a critical role in plant growth and development. RNA sequencing technology has become widely utilized in crop research to analyze gene expression patterns and elucidate the molecular mechanisms underlying various traits (Aleem et al., 2021; Li et al., 2017; Shankar et al., 2016). While research on the initiation and development of cotton trichomes has primarily concentrated on cotton fibers (Qin et al., 2022), significant morphological differences exist between cotton stem hairs and cotton fibers, suggesting that the developmental mechanisms of trichomes may differ (Hu et al., 2019). Although some studies have reported quantitative trait locus (QTL) mapping and have cloned several key genes that regulate cotton epidermal hairs (Ding et al., 2015; Wan et al., 2007), the underlying molecular mechanisms remain poorly understood.

In this study, we selected two mutants of *Gossypium arboreum*: SMA-4, a *GaHDI* loss-of-function mutant, and DPL972, a *GaGIR1* overexpression mutant, as our research subjects. We developed a system for investigating epidermal hair initiation and development at the cellular level during hypocotyl growth, which, in combination with transcriptome analysis of epidermal cells, led us to identify five key genes. We examined the effects of these genes on trichome growth and development, as well as the regulatory networks involved in epidermal hair initiation and development. Our findings provide a theoretical foundation for a deeper

understanding of the molecular mechanisms that govern the complex processes underlying cotton epidermal hair initiation and development. Moreover, our research offers valuable insights for future investigations into single-cell trichome development.

2 Materials and methods

2.1 Plant materials

In this study, we utilized the Asian cotton varieties SMA-4, DPL972, and A2-47 as plant materials. SMA-4 is a *GaHDI* mutant that exhibits a loss of function (Ding et al., 2020), resulting in the absence of epidermal hairs on the stems, leaves, and fibers. DPL972 is a mutant characterized by high expression of *GaGIR1* (Wang X, et al., 2021), leading to short trichomes on the stem and long fibers on the seed, while lacking long trichomes and short fibers. In contrast, A2-47 is an Asian cotton variety that possesses both normal long and short trichomes on the stem, as well as long and short fibers on the seed.

2.2 Technical process

In this study, we observed the epidermal hairs on the hypocotyls of Asian cotton and analyzed the associated molecular mechanisms using RNA-seq and RT-PCR techniques, as outlined in (Supplementary Figure S1). Following phenotypic observations, we conducted principal component analysis (PCA) and differential expression analysis of genes (DEGs). MaSigPro time series analysis was utilized to identify differentially expressed genes among materials at various time points, while Weighted Gene Co-expression Network Analysis (WGCNA) was employed to identify key gene modules. The results from MaSigPro time series analysis were integrated with other relevant data using a Venn diagram. For feature selection, we applied Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis, Support Vector

Machine Recursive Feature Elimination (SVM-RFE) analysis, and Boruta analysis. Through this comprehensive series of analytical steps, we identified the DEGs associated with the growth and development of hypocotyl trichomes.

2.3 Scanning electron microscopy observation

To investigate the initial development of hypocotyl trichomes, cotton seeds were soaked in water for 8 to 12 hours and then placed on germination paper. After the radicle had emerged, samples were collected at 24, 60, 72, and 104 hours, subsequently referred to as T0, T1, T3, and T4. The seeds were examined under a stereomicroscope to document the early trichome development. The epidermal cells of the hypocotyl at these four developmental stages were carefully removed with tweezers (Figure 1) and immediately placed in liquid nitrogen to prevent RNA degradation. Additionally, the epidermal cells corresponding to these four stages were prepared for microscopy, and the morphology of the trichomes was analyzed using a scanning electron microscope (HITACHI S-530) (Hu et al., 2018).

2.4 RNA - seq data analysis

At T0, T1, T3, and T4, the epidermis of the growing portion of the epicuticular hairs on the hypocotyl was removed, RNA was extracted, and both mRNA and sRNA libraries were constructed following the guidelines provided in the Illumina Transcription Library Manual. Each sample consisted of three biological replicates, resulting in a total of 36 libraries. The libraries were sequenced using the Illumina Novaseq platform, employing double-ended sequencing technology.

Use Cutadapt and FASTQ (Chen et al., 2018) software for quality control of raw data. Hisat2 software (version 2.2.1) (Kim

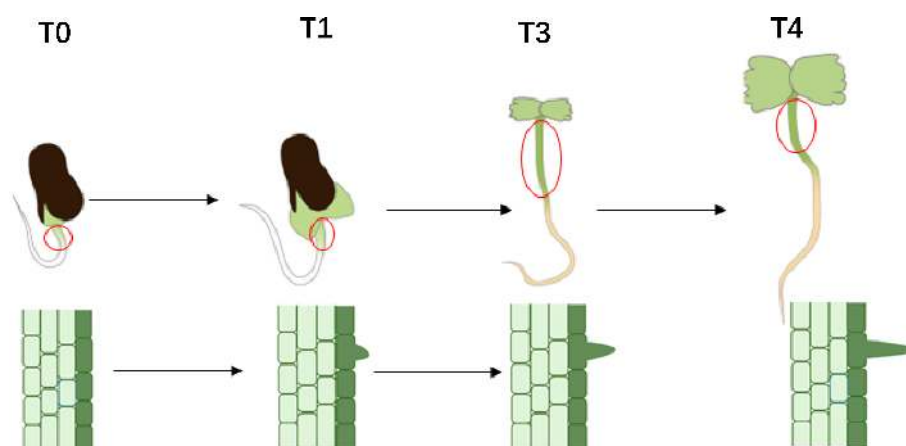


FIGURE 1
Trichome initiation and development pattern on the hypocotyl of *Gossypium arboreum*.

et al., 2015) was used to map clean reads to the reference gene of Asian cotton (*G. arboreum* _ CRI_ A2, version 1.0). SAMTOOLS (v.0.1.19; Danecek et al., 2021) was used to filter sequenced reads with mapping quality < 25. The rest used R language (version 4.2.2) to calculate the TPM value of the sample (Supplementary Table S1). Using DESeq2 (version 1.26.0) (Love et al., 2014) and $|\log_2(\text{fold change})| \geq 1$ and P-adjusted < 0.05 as screening conditions, DEGs were identified.

2.5 WGCNA

Co-expression networks were constructed using WGCNA (version 1.70) (Langfelder and Horvath, 2008) for all genes in 36 samples. The power was 8, minModuleSize was 30, mergeCutHeight was 0.25, and the rest of the modules were constructed using the default settings of the automatic network construction function blockwiseModules. Among the identified modules, the gray modules representing genes with classification failure are considered invalid. Modules with $|cox| > 0.3$ and P value < 0.05 were highly associated with the phenotype. Finally, the module genes were intersected with the differential analysis taken to construct the interaction network of intersected genes and GaHD1, GaGIR1 using Cytoscape software (version 3.9.0) (Otasek et al., 2019).

2.6 Machine learning

SVM-RFE analysis, Boruta's algorithm and LASSO regression analysis are statistical methods for feature selection and regression. See reference for a detailed description (Desai et al., 2008; Friedman et al., 2010; Kursu and Rudnicki, 2010). The intersecting genes of WGCNA and temporal analysis were algorithmically analyzed by constructing a model featuring velvet phenotypes at different stages of the hypocotyl to achieve automated feature selection to obtain more precise genes related to the initiation of velvet development in cotton.

2.7 Quantitative real-time PCR verification of candidate genes

The hypocotyl epidermal cells of A2-47, SMA-4, and DPL972 at T0, T1, T3, and T4 during the initial development of hypocotyls were taken as samples. The TB Green[®] Premix Ex TaqTM II (Tli RNase Plus) kit (Vazyme, R223-01) was utilized to transcribe 1 µg of RNA from each sample into complementary DNA (cDNA). Specific primers (Supplementary Table S2) were designed utilizing the <https://biodb.swu.edu.cn/qprimerdb/website>. The amplification program consisted of an initial denaturation step at 95°C for 30 seconds, followed by 39 cycles of denaturation at 95°C for 5 seconds and annealing at 60°C for 30 seconds. The relative expression levels of the genes were determined using the $2^{-\Delta\Delta C_t}$ method, as described by (Livak and Schmittgen, 2001).

3 Results

3.1 Initial development of epidermal hairs during hypocotyl elongation in *Gossypium arboreum* L.

To investigate the initial development of trichomes on the hypocotyl and elucidate the timing of morphological differences in trichomes, seeds were examined under a stereomicroscope following the exposure of the radicle. Observations revealed that 24 hours after radicle exposure, the hypocotyls of all three materials exhibited only round, shield-shaped cell processes at their tips, likely representing glandular hairs. After 60 hours, sharp cell processes were observed in both A2-47 and DPL972, indicating the presence of non-glandular hairs. By 72 hours, the majority of hair cells continued to elongate, and by 104 hours, the lengths of the trichomes stabilized (Figure 2). Accordingly, we categorized the time points of 24, 60, 72, and 104 hours post-radicle exposure into four developmental stages: pre-initiation, initiation, elongation, and fixed length, respectively.

At four developmental stages, the epidermis of the hypocotyls was removed for scanning electron microscopy observation, and the results were consistent with those observed under a stereomicroscope: at 24 hours, no protrusions appeared in the epidermal cells; at 60 hours, protrusions began to emerge; and the cells gradually elongated until they reached their final length between 72 and 104 hours (Figure 3). Furthermore, clustered hairs that differed from the branched clustered hairs of *Arabidopsis* were observed on the hypocotyl surfaces of DPL972 and A2-47, with the roots of these clusters closely attached to each other (Supplementary Figure S2). In this context, we speculate that the developmental pattern of these hairs may involve the simultaneous elongation of two or more closely adjacent epidermal hair cells, as well as a single cell producing two or more cells at the base through cell proliferation (such as asymmetric division or cell rearrangement following lateral division), which then further develop and arrange to form non-single root hairs.

3.2 DEGs analysis of epidermal hair cells in hypocotyls of *Gossypium arboreum* at different developmental stages based on PCA analysis

In this study, we constructed 36 cDNA libraries to sequence the epidermal hairs of the hypocotyls of three Asian cotton cultivars across four developmental stages. We obtained a total of 447.6 G of clean sequencing data, with an average data size of approximately 12.4 G per sample. The quality score Q20 exceeded 95.84%, while the Q30 quality score surpassed 90.00% (Supplementary Tables S1, S3), indicating a high-quality data output. Principal component analysis (PCA) results demonstrated that the biological replicate samples of the three Asian cotton cultivars exhibited strong clustering consistency, suggesting that the differences between

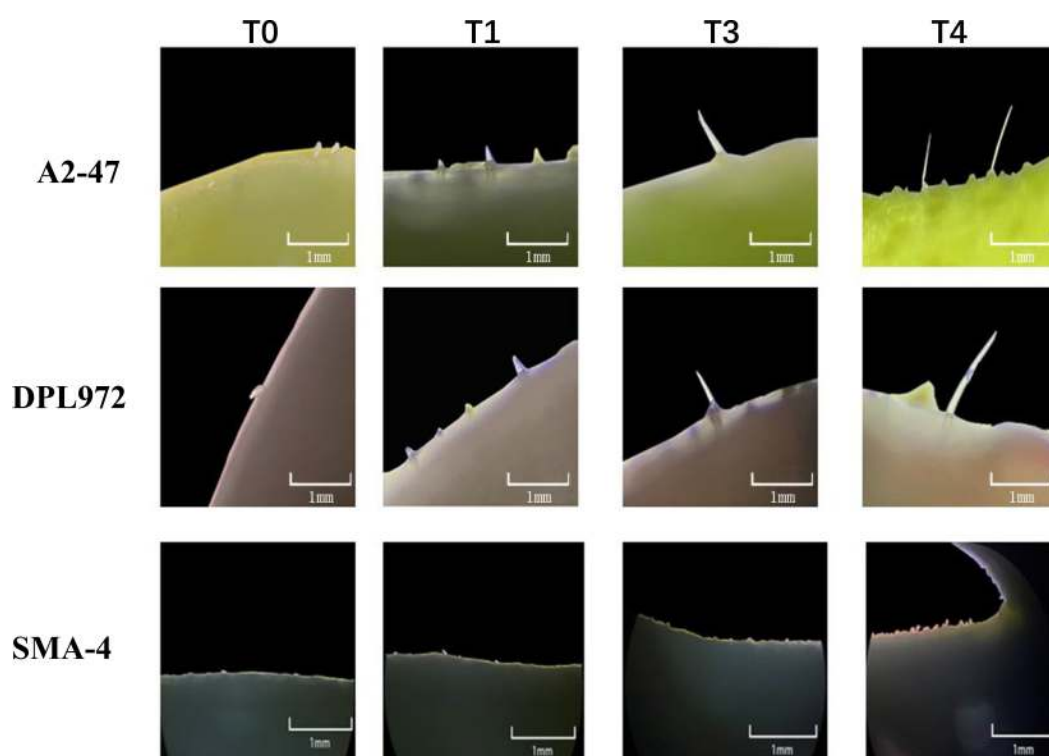


FIGURE 2
Images of trichome initiation process on hypocotyl of A2-47, DPL972, and SMA-4 under stereomicroscope.

developmental stages were more pronounced than those between the varieties. All samples were categorized into four main clusters, corresponding to the preinitial stage (24 h), initial stage (60 h), elongation stage (72 h), and fixed length stage (104 h) of trichome

development, with samples at the same developmental stage clustering closely together ([Supplementary Figure S3](#)).

Cotton trichome development is a dynamic and complex process regulated by numerous networks and a large array of

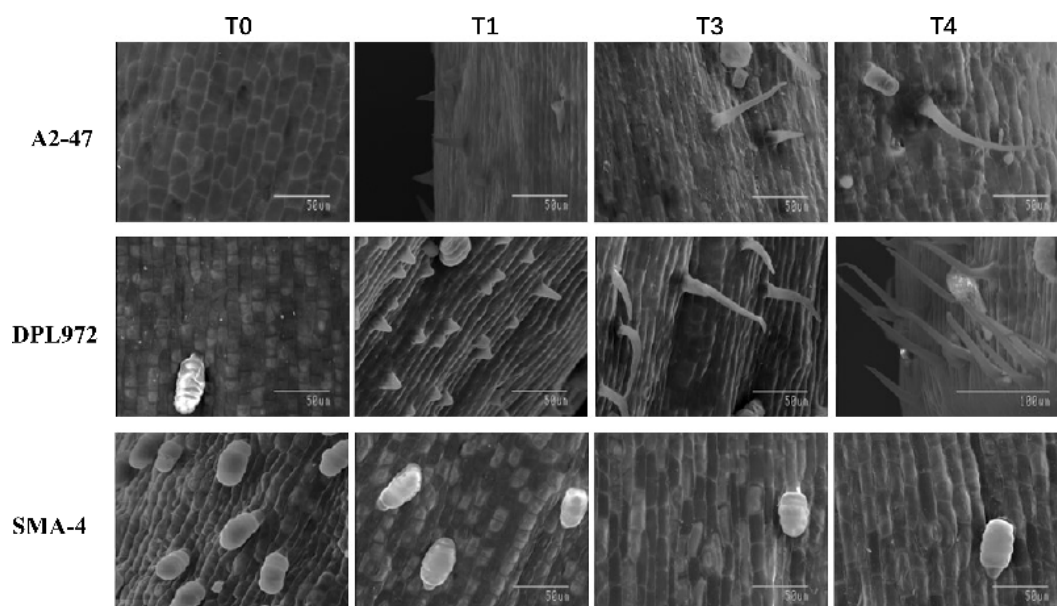


FIGURE 3
Images of trichome initiation process on hypocotyl of A2-47, DPL972 and SMA-4 under scanning electron microscopy.

genes. To identify the differentially expressed genes (DEGs) involved in the initiation and development of Asian cotton hypocotyls, we compared and analyzed two adjacent developmental stages across three materials. In A2-47, a total of 1,837 DEGs were identified by comparing stages T0 and T1, with 1,020 DEGs up-regulated in T1, potentially facilitating the transformation of epidermal cells into trichomes, while 817 DEGs were down-regulated, likely inhibiting changes in epidermal cell morphology. As development progressed, we observed 953 DEGs between T1 and T3, and as many as 3,969 DEGs between T3 and T4, with these DEGs exhibiting varying upward and downward trends across T3 and T4 (Figure 4A; Supplementary Table S4). To further investigate the biological functions of these DEGs with different regulatory patterns, we conducted a KEGG enrichment analysis, which revealed that all identified DEGs were primarily associated with phenylpropanoid biosynthesis, phenylalanine metabolism, and the biosynthetic pathways of secondary metabolites (Figure 4D).

Similarly, in SMA-4, 748 DEGs were obtained in T0 and T1, 431 in T1 and T3, and 558 in T3 and T4 (Figure 4B; Supplementary Table S5). KEGG enrichment analysis showed that these genes were significantly enriched in the biosynthesis of valine, leucine and isoleucine, biosynthesis of secondary metabolites, and sphingolipid metabolism pathway (Figure 4E).

For DPL972, a total of 462 DEGs were identified between T0 and T1, 988 DEGs between T1 and T3, and 3,459 DEGs between T3 and

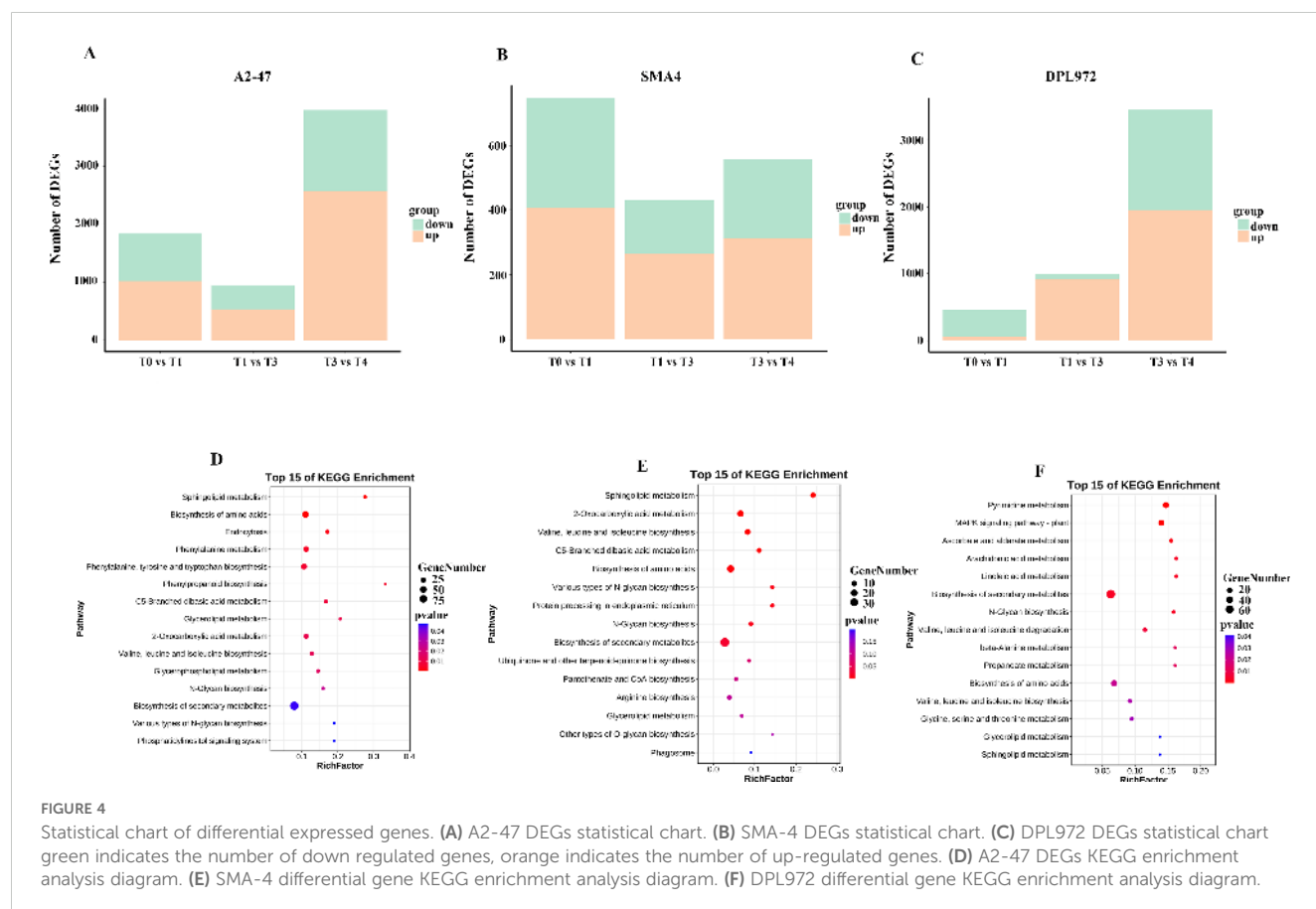
T4 (Figure 4C; Supplementary Table S6). KEGG enrichment analysis revealed that these genes were significantly associated with pathways involved in secondary metabolite biosynthesis, the MAPK signaling pathway (in plants), and pyrimidine metabolism (Figure 4F).

In conjunction with the aforementioned results, the number of DEGs at the T0-T1 and T3-T4 stages was found to be greater than that observed in the other stages, indicating that these two critical periods may play a significant role in regulating the initiation and development of hypocotyl trichomes.

3.3 Screening candidate genes for trichome initiation based on maSigpro time series analysis

3.3.1 Time series analysis of SMA-4 and A2-47 to screen candidate genes for trichome initiation and development:

In order to understand the expression patterns of DEGs at different stages of puberty initiation and development. SMA-4 and A2-47 were further explored through maSigpro time series analysis (Supplementary Figure S4). According to the difference of expression trend, these genes were divided into nine modules, cluster1-cluster9. The genes in cluster6, cluster7 and cluster9 modules (Supplementary Table S7) have opposite expression trends before and during the trichome initiation of the two



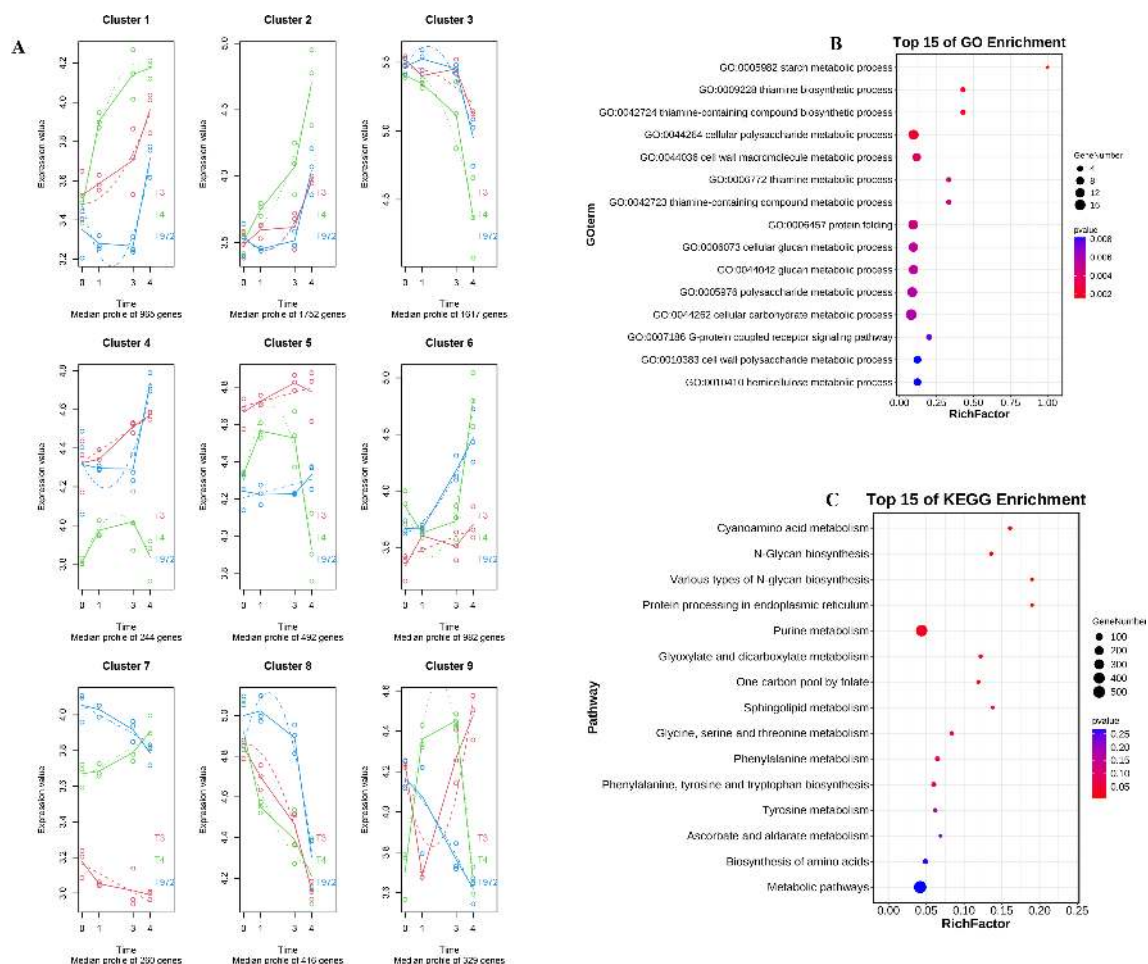


FIGURE 5

SMA-4 and A2-47 timing analysis. (A) Time sequence analysis Multi-segment line graph of differential genes in A2-47 and SMA-4. (B) GO enrichment analysis of differential genes in C2 and C3 modules. (C) KEGG enrichment analysis of C2 and C3 module differential genes.

materials (Figure 5A). In order to clarify the biological functions and pathways of these three module genes, we performed GO enrichment and KEGG enrichment analysis. GO enrichment analysis results showed that these genes were closely related to starch metabolism, cell wall Macromolecule Metabolism, cell wall polysaccharide metabolism and hemicellulose metabolism. These processes had significant effects on plant cell division, differentiation and elongation, as well as the construction and maintenance of cell wall structure. At the same time, KEGG enrichment analysis also showed that these differential genes were significantly enriched in the biosynthesis of phenylalanine, tyrosine and tryptophan, phenylalanine metabolism, glycine, serine and threonine metabolism and other pathways, which played an important role in plant growth and development, metabolism and stress response pathways (Figures 5B, C).

3.3.2 Screening candidate genes for trichome initiation by DPL972 and SMA-4 time series analysis

MaSigPro time series analysis is also performed on DPL972 and SMA-4 (Supplementary Figure S5). The results showed that the genes in cluster3, cluster4 and cluster5 modules (Supplementary Table S8) had opposite expression trends before and during the trichome initiation of the two materials (Figure 6A). GO enrichment and KEGG enrichment analysis were performed. The results of GO enrichment analysis showed that the DEGs were related to biological processes such as oxygenic acid metabolism, carboxylic acid metabolism and small molecule metabolism. KEGG enrichment analysis is closely related to biosynthesis of secondary metabolites, biosynthesis of amino acids, and degradation pathways of valine, leucine, and isoleucine (Figures 6B, C).

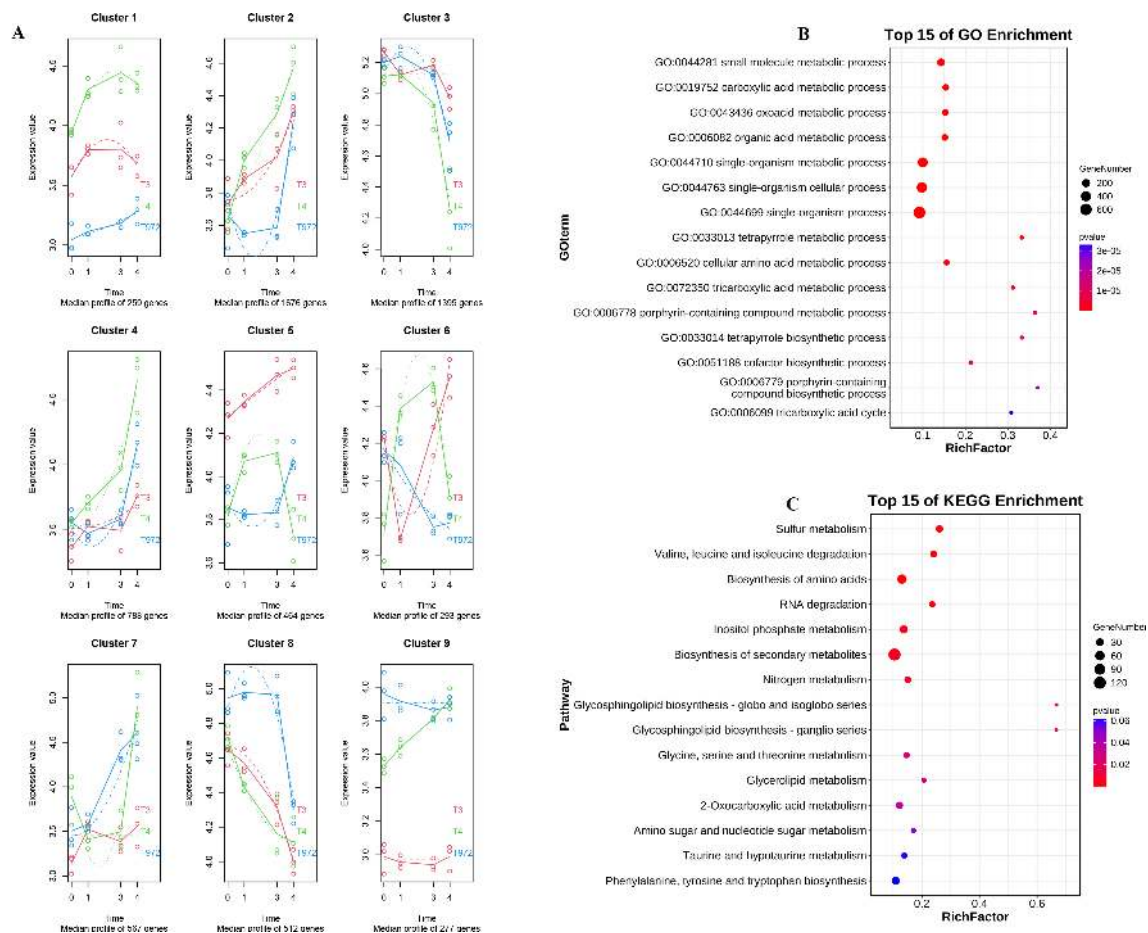


FIGURE 6

DPL972 and SMA-4 timing analysis. (A) Time sequence analysis Multi-segment line graph of differential genes in DPL972 and SMA-4. (B) GO enrichment analysis of differential genes. (C) KEGG enrichment analysis of differential gene.

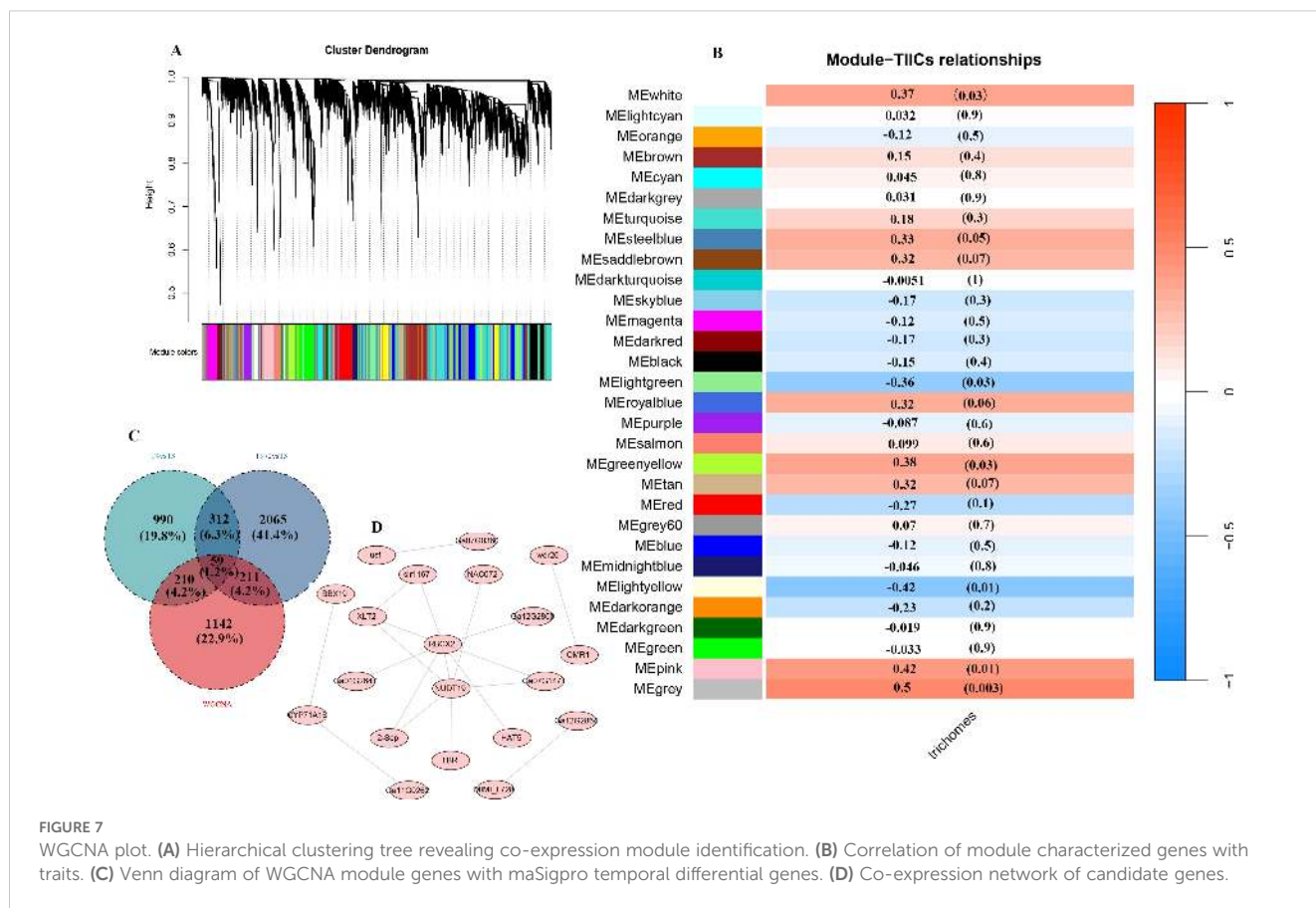
3.4 Analysis of the regulatory network of cotton hairy morphogenesis related genes based on WGCNA

The transcriptome data were analyzed globally to explore the phenotypic association between DEGs and trichome morphogenesis and transformation. The transcriptome data were processed by WGCNA. All genes were divided into 30 modules, and each module was represented by a separate branch (Figure 7A). These modules are composed of genes that exhibit similar expression patterns, reflecting their close relationship in function or regulation. Compared with other modules, MEpink module was positively correlated with trichome morphogenesis, while MELight yellow module was negatively correlated with trichome morphogenesis ($\text{cox} = 0.42$, $r = -0.42$, $P \text{ value} < 0.05$) (Figure 7D). MEpink module contains *GaHDI* (*Ga06G1792*), *GaRDL1* (*Ga05G0755*), *GaEXPA1* (*Ga10G1805*) and other genes closely related to the initiation and development of epidermal hair (Supplementary Table S9) (Xu et al., 2013).

To further investigate the functional characteristics of these modules, we conducted KEGG enrichment analysis on the genes associated with the MEpink and MELightyellow modules. The

results indicated that the genes in the MEpink module are primarily involved in the biosynthesis of thiometabolism, as well as the metabolism of valine, leucine, and isoleucine, along with the C5-branched-chain dicarboxylic acid metabolic pathway (Supplementary Figure S6A). In contrast, the genes in the MELightyellow module exhibit significant enrichment in pathways related to the degradation of valine, leucine, and isoleucine, the phosphatidylinositol signaling pathway, RNA polymerase activity, and carbon metabolism (Supplementary Figure S6B).

To identify candidate genes directly associated with the initiation and development of puberty, we analyzed the intersection of differentially expressed genes (DEGs) and genes from weighted gene co-expression network analysis (WGCNA) modules, ultimately identifying 59 intersecting genes (Figure 7C). These genes are considered potential candidates for involvement in trichome initiation and development. We constructed a regulatory network for these 59 genes using Cytoscape software and found that only 20 of these genes may have regulatory interactions. Notably, the proteins encoded by *Ga13G2472* (NUDT19) and *Ga05G0743* (RBCX2) are more likely to interact with other proteins (Figure 7D), suggesting their significant roles within the regulatory network.



3.5 Machine learning accurately screens key genes for hypocotyl epidermal hair development

To accurately identify the key genes closely associated with the initiation and development of puberty, we utilized a combination of three machine learning methodologies SVM-RFE analysis, Boruta's algorithm, and LASSO regression curve analysis based on the 59 candidate genes previously identified through weighted gene co-expression network analysis (WGCNA). The integration of these three approaches enhances the stability and accuracy of our feature selection process, thereby constructing a robust predictive model.

In the LASSO regression curve analysis, when the optimal λ value (0.04107188) is reached, six genes are not punished to 0, which indicates that these six genes are more important in the process of pubescence growth and development than other genes (Figures 8A, B). In SVM-RFE analysis, when the model reached the highest correlation for the first time, 19 genes passed the algorithm verification (Figure 8C), which showed that these genes had a high contribution in predicting the initial development of trichome. At the same time, Boruta's algorithm was also used to analyze. The results showed that 13 genes passed the model validation (Figure 8D). The intersection results are shown in three machine learning models, *Ga02G1392*(*TBR*), *Ga03G0474*(*OMR1*), *Ga12G2860*(*ACO1*), *Ga11G2117*(*BBX19*) and *Ga12G2864*(*CUE*) both can pass the

verification (Figure 8E). These results indicate that these five genes play an important role in the initial development of trichome.

3.6 Expression trend analysis of DEGs during the initial development of epidermal hair

In order to further verify the reliability and accuracy of RNA seq data, we used real-time fluorescent quantitative PCR to analyze the expression levels of five screened genes *Ga02G1392* (*TBR*), *Ga03G0474* (*OMR1*), *Ga12G2860* (*ACO1*), *Ga11G2117* (*BBX19*), and *Ga12G2864* (*CUE*). The RT qPCR results were compared with the transcriptome analysis data. The results showed that the relative expression level obtained by RT qPCR analysis was consistent with the TPM value calculated by RNA seq (Figure 9), indicating that the bioinformatics analysis method used in RNA seq data processing was correct.

At the same time, the expression level trend of five genes during the initial development of hypocotyl was analyzed (Figure 9). The results showed that the expressions of *GaTBR* and *GaBBX19* in A2–47 were up-regulated at the initial stage, and the expressions of *GaACO1* and *GaCUE* gradually increased at the secondary cell wall thickening stage, while the expressions of *GaOMR1* and *GaTBR* increased again at the mature stage, indicating that the expression

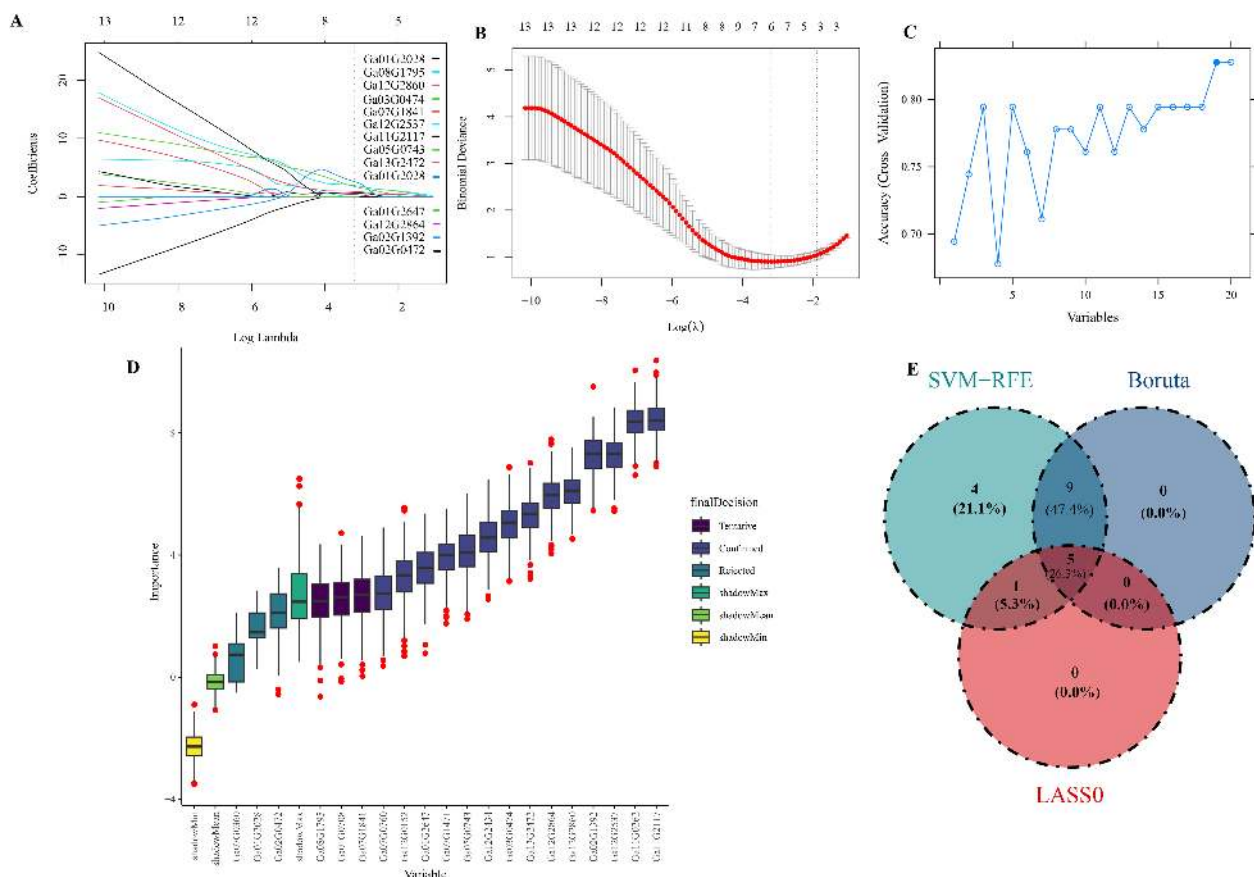


FIGURE 8

LASSO regression curves. **(A)** LASSO regression roadmap: the abscissa is the logarithm of Lambdas, and the ordinate is the variable coefficient. With the increase of Lambdas, the variable coefficient approaches zero. When the optimal lambda is reached, variables with a coefficient equal to 0 are eliminated. **(B)** LASSO cross validation diagram: the abscissa is the logarithm of Lambdas, and the ordinate model error. The best lambda value is at the lowest point of the red curve. **(C)** SVM-RFE analysis chart: the abscissa represents the performance index of the model, the ordinate represents the number of genes. **(D)** Boruta analysis chart: the abscissa represents the name of variables, and the ordinate represents the importance of features. **(E)** Three machine learning Venn charts.

trend of the five genes in normal gene expression materials showed obvious stage specificity, which was highly consistent with the four stages of development. The expression levels of *GaTBR*, *GaBBX19*, *GaOMR1* and *GaACO1* genes in SMA-4 were significantly lower than those in A2-47, while *GaTBR* was highly expressed in DPL972, while *GaCUE* was very low. These results suggest that the disorder of gene expression sequence (inhibition or continuous activation) in the mutant may be the reason for the difference of trichome phenotype.

4 Discussion

4.1 Cotton hypocotyl epidermal trichome: a good system for studying the fate of epidermal cells

Analyzing the molecular mechanisms that drive the differentiation and development of plant epidermal cells remains a central focus of studies in plant organogenesis and morphological

diversity. In this research domain, establishing an optimal experimental system is crucial for ensuring successful outcomes. Plant epidermal hairs, commonly referred to as trichomes, have emerged as a paradigmatic model for investigating plant cell differentiation—specifically, the determination of cell fate and the regulation of polarity, owing to their ease of visualization and single-cell nature (Hulskamp, 2004). Trichomes of *Arabidopsis thaliana* have been extensively utilized as model systems, capitalizing on their genomic characteristics and the organism's status as a model plant (Han et al., 2022), thereby elucidating the molecular mechanisms underlying trichome development.

Although a single-cell development model for cotton fibers has been established (Qin et al., 2022), the sampling of ovule epidermal cells in cotton is feasible only during specific growth stages, namely pre- and post-flowering. The small size of these cells, as well as their encapsulation within the ovary, presents significant challenges for sampling and observation. Both cotton fiber and stem fuzz are composed of single cells, and the cytological processes through which ovule cells extend to form cotton fibers display similarities to the formation of stem fuzz. Research has revealed notable parallels

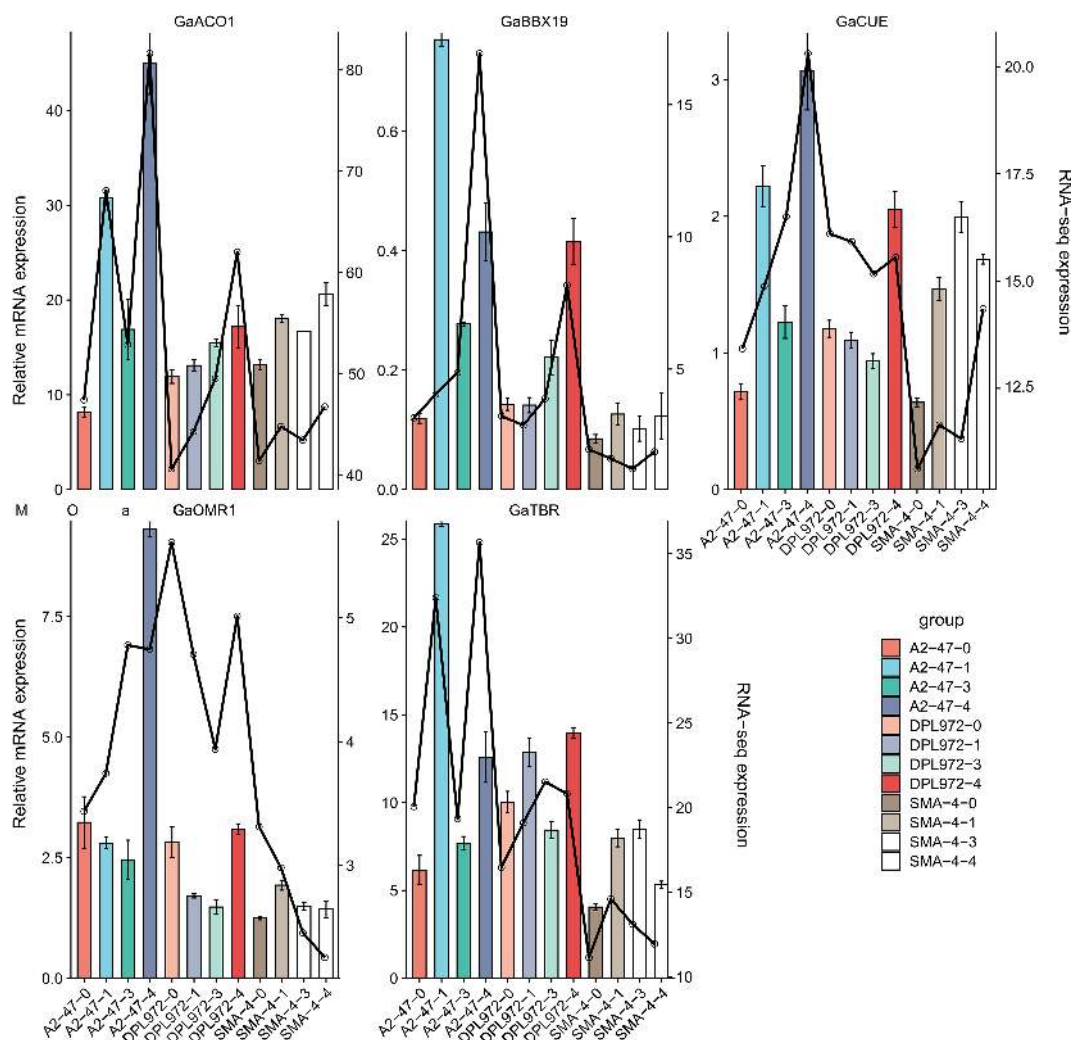


FIGURE 9

Trend of differential gene expression level. The abscissa is the initial development stage of the epidermis, and the ordinate is the relative expression. The histogram shows the relative expression of RT qPCR; The line graph represents the relative expression in transcriptome data.

in their molecular regulatory mechanisms; thus, exploring the molecular pathways involved in the development of stem and leaf hairs is crucial for advancing our understanding of the mechanisms underlying fiber development.

To date, research on cotton stem trichomes has primarily focused on measuring trichome length and classifying trichome types in their mature states (Yuan et al., 2021). The findings from these studies may be somewhat outdated and could overlook the underlying molecular mechanisms governing epidermal cell differentiation. This study reveals that during the germination of Asian cotton seeds, specific epidermal cells in the hypocotyl initiate protrusion and develop into trichomes, which provide an accessible subject for observation and sampling. Consequently, we present a straightforward and efficient system for investigating plant epidermal cell differentiation and development, offering a model for exploring the fate of essential epidermal cells in crops. Utilizing this system, we compared dynamic changes in epidermal hair

phenotypes between the wild-type A2-47 and the mutants SMA-4 (GaHDI mutation) and DPL972 (which exhibits high expression of GaGIR1), identifying four distinct stages in the differentiation and development of epidermal cells into trichomes. Each stage was characterized by unique features: the preinitial stage (T0) exhibited condensed basal chromatin without nuclear enlargement; the initial stage (T1) displayed polar protrusions at the cell apex and nuclear translocation to the sub-apex; while the elongation stage (T3) and the fixed length stage (T4) were distinguished by ongoing cell elongation and specific nuclear localization.

Furthermore, this study first observed that the formation of clustered, non-single epidermal trichomes involves two pathways: proliferation and rearrangement of epidermal hair cells, alongside simultaneous development and growth of adjacent cells. This offers a novel research perspective on plant cell division and regulation of trichome polarity, providing crucial morphological evidence to further unravel the molecular mechanisms underlying cotton fuzz development.

4.2 Innovative application of key gene screening technology for epidermal hair initiation and development

The innovative application of key gene screening technology for epidermal hair initiation and development in this study has achieved a dual breakthrough in the methodology for identifying such genes.

4.2.1 Dynamic capture of DEGs at multiple stages

Utilizing the precise time series system (T0-T4) of Asian cotton epidermal hair development, this study employed the MaSigPro two-step regression method to effectively analyze dynamic DEGs between the wild-type A2-47 and mutants (SMA-4, DPL972). Compared to traditional differential analysis, this approach significantly enhances the recognition sensitivity of developmental sequence-specific genes, as previously reported (Nueda et al., 2014). This analysis method is widely used in medicine to identify the gene expression with time changes in different cell types. The common response of immune cells to virus infection at different times (O'Leary and Zheng, 2024). In this study, 1572 and 2648 DEGs were identified between SMA-4 and A2-47 and between DPL972 and A2-47, which improved the capture efficiency of developmental stage specific genes.

4.2.2 Multimodal machine learning fusion screening

In pepper (*Capsicum annuum* L.), WGCNA was used to screen important modules related to trichome formation, and six differential genes related to trichome development in Pepper epidermis (Shen et al., 2024) were identified. To mitigate the biases inherent in a single algorithm, this study innovatively integrates WGCNA with three machine learning algorithms: LASSO, SVM-RFE and Boruta feature selection analysis. Machine learning can integrate genome, transcriptome, epigenome and other multi omics data, find cross level marker combinations, and improve predictability. Machine learning can mine potential markers (such as non coding RNA and microbiome characteristics) ignored by traditional methods, and reveal their biological mechanisms (GuyonIsabelle and ElisseeffAndré, 2003; Kursa and Rudnicki, 2010; Xiaokang, 2020). However, in bioinformatics, data sets are often complex and the number of samples is limited. A single model is prone to fitting problems, that is, it performs well on the training set, but has poor generalization ability on unknown data (Bolón-Canedo et al., 2012; Saeys et al., 2007). In addition, a single machine learning algorithm may be sensitive to some characteristics of the data or noise, resulting in different results of the screening basis (Statnikov et al., 2008; Subramanian et al., 2005) given by different algorithms. Therefore, by selecting the intersection of different machine learning algorithms and combining the advantages of various machine learning algorithms (SVM-RFE, Lasso, Boruta), we can reduce the possible deviation caused by a single algorithm. A more

comprehensive and diversified result is obtained, which helps to reduce noise, avoid over fitting, improve biological credibility, and ensure that the five different genes screened can be consistently verified in multiple algorithms, thus enhancing the reliability of the results (Statnikov et al., 2008; Subramanian et al., 2005).

4.3 Functional network analysis of key candidate genes

We conducted an analysis of the roles played by five key genes identified through our screening in the regulatory network governing cotton trichome initiation and development. The formation of epidermal hair papillae is a crucial aspect during the cell wall maturation stage (Esch et al., 2003; Hulskamp et al., 1994; Kawashima et al., 2014). In *Arabidopsis*, the *TBR* protein regulates the deposition of crystalline cellulose in the papillae region, which in turn affects trichome transparency and structural integrity (Sinclair et al., 2020; Suo et al., 2013). Our findings suggest that the acetyltransferase encoded by *GaTBR* may also participate in cotton trichome development through a similar mechanism. Xyloglucan, an essential structural polysaccharide in plant cell walls, forms cross-linking structures with cellulose microfibrils that restrict cell elongation (Smith and Hake, 1992; Vissenberg et al., 2005). Consequently, *GaTBR* may influence cell wall loosening and cell elongation by modulating xyloglucan acetylation levels, thereby affecting the activity of enzymes such as *XTH* (Vissenberg et al., 2000, 2005). Importantly, *GaTBR* exhibits strictly developmental stage-specific expression in wild-type plants, while its expression is disrupted in mutants, indicating potential direct or indirect regulation by the upstream transcription factor *GaHDI*.

Branched-chain amino acids (BCAAs) are essential building blocks for protein synthesis, and their metabolites also function as signaling molecules that regulate cell cycle progression. In *Arabidopsis*, the threonine deaminase/dehydratase (TD) encoded by *OMR1* plays a critical role in coordinating cell proliferation and expansion during root hair development by catalyzing isoleucine biosynthesis (Yu et al., 2013). This study posits that *GaOMR1* may influence cell proliferation and morphogenesis in cotton trichome development through the modulation of isoleucine biosynthesis.

The BBX protein family, identified as core regulators of photomorphogenesis (Gangappa and Botto, 2014), is capable of modulating trichome initiation by integrating various endogenous hormone signaling pathways (Wang et al., 2024; Xin et al., 2025; Xu et al., 2024). In cotton, the DELLA protein GhSLR1, which acts as a negative regulator of gibberellin (GA) signaling, suppresses fiber cell elongation by destabilizing the GhHOX3-GhHDI complex (Hao et al., 2012). Conversely, *GhTCP14* promotes GA biosynthesis, induces the degradation of GhSLR1 protein, and activates downstream cell wall loosening genes such as *GhRDL1* and *GhEXPA1*, ultimately facilitating fiber elongation (Shan et al., 2014). Building on these findings, we hypothesize that *GaBBX19* in Asiatic cotton (*Gossypium arboreum*) is

involved in the hormone-mediated regulation of trichome development. Its responsiveness to photoperiod signals (Wang et al., 2014) may account for the phenotypic variation observed in trichome density under different light conditions.

Ethylene, a well-known inhibitory hormone, displays a negative correlation between the activity of its key biosynthetic enzyme, 1-aminocyclopropane-1-carboxylate oxidase (ACO), and cell elongation (Zarembinski and Theologis, 1994). In *Arabidopsis*, mitogen-activated protein kinase 6 (MPK6) phosphorylates 1-aminocyclopropane-1-carboxylate synthase (ACS) proteins to enhance ethylene biosynthesis, thereby influencing stress-responsive growth (Liu and Zhang, 2004). These findings suggest that *GaACO1* may regulate trichome cell elongation through the activation of MAPK cascades, the downregulation of cell cycle-related genes, and its involvement in the ethylene biosynthesis pathway (Wang et al., 2011).

In summary, this study employed an established model of cotton hypocotyl trichome initiation to gain deeper insights into the complete life cycle of hypocotyl epidermal trichome cells in cotton. By integrating temporal progression analysis (MaSigPro), co-expression networks (WGCNA), and machine learning approaches, we systematically and comprehensively analyzed the differential expression profiles during the initiation and development of cotton trichomes. Our analysis identified five candidate genes associated with cell wall formation and the gibberellin and abscisic acid signaling pathways, which are likely to serve as key regulators of trichome initiation in Asiatic cotton (*Gossypium arboreum*) hypocotyls. This research provides a detailed morphological framework for investigating the molecular mechanisms underlying cotton trichome development and establishes a robust data and theoretical foundation for subsequent functional validation of these candidate genes.

Data availability statement

The original data has been uploaded to NCBI: “PRJNA1282329 : Transcriptomic Exploration Yields Novel Perspectives on the Regulatory Network Underlying Trichome Initiation in *Gossypium arboreum* Hypocotyl.”

Author contributions

YX: Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. LY: Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. ZZ: Validation, Visualization, Writing – original draft. MD: Conceptualization, Supervision, Writing – review &

editing. YC: Conceptualization, Supervision, Writing – review & editing. XH: Conceptualization, Supervision, Writing – review & editing. JR: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the grants from the National Key R&D Program of China (Grant No. 2024YFD1200300 to Junkang Rong), and the National Natural Science Foundation of China (Grant No. 32171923 to Mingquan Ding; Grant No. 31501349 to Yuefen Cao).

Acknowledgments

We thank Prof. Xiongming Du at the Institute of Cotton Research, Chinese Academy of Agricultural Science (ICR, CAAS) for providing us with *G. arboreum* materials.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1604186/full#supplementary-material>

References

- Aleem, M., Raza, M. M., Haider, M. S., Atif, R. M., Ali, Z., Bhat, J. A., et al. (2021). Comprehensive RNA-seq analysis revealed molecular pathways and genes associated with drought tolerance in wild soybean (*Glycine soja* Sieb. and Zucc.). *Physiol. Plant* 172, 707–732. doi: 10.1111/pp1.v172.2
- Bolón-Canedo, V., Sánchez-Marroño, N., and Alonso-Betanzos, A. (2012). A review of feature selection methods on synthetic data. *Knowl. Inf. Syst.* 34, 483–519.
- Bouyer, D., Kirik, V., and Hulskamp, M. (2001). Cell polarity in *Arabidopsis trichomes*. *Semin. Cell Dev. Biol.* 12, 353–356. doi: 10.1006/scdb.2001.0263
- Cao, J. F., Zhao, B., Huang, C. C., Chen, Z. W., Zhao, T., Liu, H. R., et al. (2020). The miR319-targeted GhTCP4 Promotes the transition from cell elongation to wall thickening in cotton fiber. *Mol. Plant* 13, 1063–1077.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, 1884–1890. doi: 10.1093/bioinformatics/bty560
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10. doi: 10.1093/gigascience/giab008
- Desai, A., Chee, P. W., May, O. L., and Paterson, A. H. (2008). Correspondence of trichome mutations in diploid and tetraploid cottons. *J. Hered.* 99, 182–186. doi: 10.1093/jhered/esm112
- Ding, M., Cao, Y., He, S., Sun, J., Dai, H., Zhang, H., et al. (2020). GaHD1, a candidate gene for the *Gossypium arboreum* SMA-4 mutant, promotes trichome and fiber initiation by cellular H₂O₂ and Ca²⁺ signals. *Plant Mol. Biol.* 103, 409–423. doi: 10.1007/s11103-020-01000-3
- Ding, M., Ye, W., Lin, L., He, S., Du, X., Chen, A., et al. (2015). The hairless stem phenotype of cotton (*Gossypium barbadense*) is linked to a copia-like retrotransposon insertion in a homeodomain-leucine zipper gene (HD1). *Genetics* 201, 143–154. doi: 10.1534/genetics.115.178236
- Dong, M., Xue, S., Bartholomew, E. S., Zhai, X., Sun, L., Xu, S., et al. (2022). Transcriptomic and functional analysis provides molecular insights into multicellular trichome development. *Plant Physiol.* 189, 301–314. doi: 10.1093/plphys/kiac050
- Edward, P. T. (2013). A fertile field: the mutual influence and parallel histories of auxin and flavonoids. *Signaling communication Plants*.
- Esch, J. J., Chen, M., Sanders, M., Hillestad, M., Ndkium, S., Idelkope, B., et al. (2003). A contradictory GLABRA3 allele helps define gene interactions controlling trichome development in *Arabidopsis*. *Development* 130, 5885–5894. doi: 10.1242/dev.00812
- Feng, X., Cheng, H., Zuo, D., Zhang, Y., Wang, Q., Liu, K., et al. (2019). Fine mapping and identification of the fuzzless gene GaFz1 in DPL972 (*Gossypium arboreum*). *Theor. Appl. Genet.* 132, 2169–2179. doi: 10.1007/s00122-019-03330-3
- Feng, H., Li, X., Chen, H., Deng, J., Zhang, C., Liu, J., et al. (2018). GhHUB2, a ubiquitin ligase, is involved in cotton fiber development via the ubiquitin-26S proteasome pathway. *J. Exp. Bot.* 69, 5059–5075.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Software* 33, 1–22. doi: 10.18637/jss.v033.i01
- Gangappa, S. N., and Botto, J. F. (2014). The BBX family of plant transcription factors. *Trends Plant Sci.* 19, 460–470.
- GuyonIsabelle & Elisseeff André 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research*.
- Han, G., Li, Y., Yang, Z., Wang, C., Zhang, Y., and Wang, B. (2022). Molecular mechanisms of plant trichome development. *Front. Plant Sci.* 13, 910228. doi: 10.3389/fpls.2022.910228
- Hao, J., Tu, L., Hu, H., Tan, J., Deng, F., Tang, W., et al. (2012). GbTCP, a cotton TCP transcription factor, confers fiber elongation and root hair development by a complex regulating system. *J. Exp. Bot.* 63, 6267–6281. doi: 10.1093/jxb/ers278
- Hu, H., Wang, M., Ding, Y., Zhu, S., Zhao, G., Tu, L., et al. (2018). Transcriptomic repertoires depict the initiation of lint and fuzz fibers in cotton (*Gossypium hirsutum* L.). *Plant Biotechnol. J.* 16, 1002–1012. doi: 10.1111/pbi.2018.16.issue-5
- Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., et al. (2019). *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* 51, 739–748. doi: 10.1038/s41588-019-0371-5
- Hulskamp, M. (2004). Plant trichomes: a model for cell differentiation. *Nat. Rev. Mol. Cell Biol.* 5, 471–480. doi: 10.1038/nrm1404
- Hulskamp, M., Misra, S., and Jurgens, G. (1994). Genetic dissection of trichome cell development in *Arabidopsis*. *Cell* 76, 555–566. doi: 10.1016/0092-8674(94)90118-X
- Kawashima, T., Maruyama, D., Shagirov, M., Li, J., Hamamura, Y., Yelagandula, R., et al. (2014). Dynamic F-actin movement is essential for fertilization in *Arabidopsis thaliana*. *Elife* 3. doi: 10.7554/eLife.04501.023
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317
- Kursa, M. B., and Rudnicki, W. R. (2010). Feature selection with the Boruta Package. *J. Stat. Software* 36. doi: 10.18637/jss.v036.i11
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559
- Li, H. B., Qin, Y. M., Pang, Y., Song, W. Q., Mei, W. Q., and Zhu, Y. X. (2007). A cotton ascorbate peroxidase is involved in hydrogen peroxide homeostasis during fiber cell development. *New Phytol.* 175, 462–471. doi: 10.1111/j.1469-8137.2007.02120.x
- Li, P. T., Wang, M., Lu, Q. W., Ge, Q., Rashid, M. H. O., Liu, A. Y., et al. (2017). Comparative transcriptome analysis of cotton fiber development of Upland cotton (*Gossypium hirsutum*) and Chromosome Segment Substitution Lines from *G. hirsutum* x *G. barbadense*. *BMC Genomics* 18, 705. doi: 10.1186/s12864-017-4077-8
- Liu, Y., and Zhang, S. (2004). Phosphorylation of 1-aminocyclopropane-1-carboxylic acid synthase by MPK6, a stress-responsive mitogen-activated protein kinase, induces ethylene biosynthesis in *Arabidopsis* W. *Plant Cell* 16, 3386–3399.
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods* 25, 402–408.
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8
- Ma, J., Pei, W., Ma, Q., Geng, Y., Liu, G., Liu, J., et al. (2019). QTL analysis and candidate gene identification for plant height in cotton based on an interspecific backcross inbred line population of *Gossypium hirsutum* x *Gossypium barbadense*. *Theor. Appl. Genet.* 132, 2663–2676. doi: 10.1007/s00122-019-03380-7
- Nueda, M. J., Tarazona, S., and Conesa, A. (2014). Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics* 30, 2598–2602.
- Naoumkina, M., Thyssen, G. N., and Fang, D. D. (2015). RNA-seq analysis of short fiber mutants Ligon-lintless-1 (Li 1) and -2 (Li 2) revealed important role of aquaporins in cotton (*Gossypium hirsutum* L.) fiber elongation. *BMC Plant Biol.* 15, 65. doi: 10.1186/s12870-015-0454-0
- O'Leary, K., and Zheng, D. (2024). Metacell-based differential expression analysis identifies cell type specific temporal gene response programs in COVID-19 patient PBMCs. *NPJ Syst. Biol. Appl.* 10, 36. doi: 10.1038/s41540-024-00364-2
- Otake, D., Morris, J. H., Boucas, J., Pico, A. R., and Demchak, B. (2019). Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol.* 20, 185. doi: 10.1186/s13059-019-1758-4
- Qin, Y., Sun, M., Li, W., Xu, M., Shao, L., Liu, Y., et al. (2022). Single-cell RNA-seq reveals fate determination control of an individual fiber cell initiation in cotton (*Gossypium hirsutum*). *Plant Biotechnol. J.* 20, 2372–2388. doi: 10.1111/pbi.v20.12
- Saeys, Y., Inza, I., and Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics* 23, 2507–2517.
- Schwab, B., Folkers, U., Ilgenfritz, H., and Hulskamp, M. (2000). Trichome morphogenesis in *Arabidopsis*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 879–883. doi: 10.1098/rstb.2000.0623
- Shan, C. M., Shangguan, X. X., Zhao, B., Zhang, X. F., Chao, L. M., Yang, C. Q., et al. (2014). Control of cotton fiber elongation by a homeodomain transcription factor GhHOX3. *Nat. Commun.* 5, 5519. doi: 10.1038/ncomms5519
- Shankar, R., Bhattacharjee, A., and Jain, M. (2016). Transcriptome analysis in different rice cultivars provides novel insights into desiccation and salinity stress responses. *Sci. Rep.* 6, 23719. doi: 10.1038/srep23719
- Shen, Y., Mao, L., Zhou, Y., Sun, Y., Lv, J., Deng, M., et al. (2024). Transcriptome analysis reveals key genes involved in trichome formation in pepper (*Capsicum annuum* L.). *Plants (Basel)* 13. doi: 10.3390/plants13081090
- Silva-Navas, J., Moreno-Risueno, M. A., Manzano, C., Tellez-Robledo, B., Navarro-Neila, S., Carrasco, V., et al. (2016). Flavonols mediate root phototropism and growth through regulation of proliferation-to-differentiation transition. *Plant Cell* 28, 1372–1387. doi: 10.1105/tpc.15.00857
- Sinclair, S. A., Gille, S., Pauly, M., and Kramer, U. (2020). Regulation of acetylation of plant cell wall components is complex and responds to external stimuli. *Plant Signal Behav.* 15, 1687185. doi: 10.1080/15592324.2019.1687185
- Smith, L. G., and Hake, S. (1992). The initiation and determination of leaves. *Plant Cell* 4, 1017–1027.
- Statnikov, A., Wang, L., and Aliferis, C. F. (2008). A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinf.* 9, 319. doi: 10.1186/1471-2105-9-319
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102
- Suo, B., Seifert, S., and Kirik, V. (2013). *Arabidopsis* GLASSY HAIR genes promote trichome papillae development. *J. Exp. Bot.* 64, 4981–4991. doi: 10.1093/jxb/ert287
- Tan, J., Tu, L., Deng, F., Hu, H., Nie, Y., and Zhang, X. (2013). A genetic and metabolic analysis revealed that cotton fiber cell development was retarded by flavonoid naringenin. *Plant Physiol.* 162, 86–95. doi: 10.1104/pp.112.212142

- Taylor-Teeples, M., Lin, L., de Luca, M., Turco, G., Toal, T. W., Gaudinier, A., et al. (2015). An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* 517, 571–575.
- Vissenberg, K., Martinez-Vilchez, I. M., Verbelen, J.-P., Miller, J. G., and Fry, S. C. (2000). *In vivo* colocalization of xyloglucan endotransglycosylase activity and its donor substrate in the elongation zone of arabidopsis roots. *Plant Cell* 12. doi: 10.1105/tpc.12.7.1229
- Vissenberg, K., Oyama, M., Osato, Y., Yokoyama, R., Verbelen, J. P., and Nishitani, K. (2005). Differential expression of *AtXTH17*, *AtXTH18*, *AtXTH19* and *AtXTH20* genes in *Arabidopsis* roots. Physiological roles in specification in cell wall construction. *Plant Cell Physiol.* 46, 192–200. doi: 10.1093/pcp/pci013
- Walford, S. A., Wu, Y., Llewellyn, D. J., and Dennis, E. S. (2011). *GhMYB25-like*: a key factor in early cotton fiber development. *Plant J.* 65, 785–797. doi: 10.1111/j.1365-313X.2010.04464.x
- Walford, S. A., Wu, Y., Llewellyn, D. J., and Dennis, E. S. (2012). Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, *GhHD-1*. *Plant J.* 71, 464–478. doi: 10.1111/j.1365-313X.2012.05003.x
- Wan, Q., Guan, X., Yang, N., Wu, H., Pan, M., Liu, B., et al. (2016). Small interfering RNAs from bidirectional transcripts of *GhMML3_A12* regulate cotton fiber development. *New Phytol.* 210, 1298–1310. doi: 10.1111/nph.2016.210.issue-4
- Wan, Q., Zhang, Z., Hu, M., Chen, L., Liu, D., Chen, X., et al. (2007). T1 locus in cotton is the candidate gene affecting lint percentage, fiber quality and spiny bollworm (*Earias* spp.) resistance. *Euphytica* 158, 241–247. doi: 10.1007/s10681-007-9446-y
- Wang, Y., Chai, C., Valliyodan, B., Maupin, C., Annen, B., and Nguyen, H. T. (2015). Genome-wide analysis and expression profiling of the *PIN* auxin transporter gene family in soybean (*Glycine max*). *BMC Genomics* 16, 951. doi: 10.1186/s12864-015-2149-1
- Wang, L., Cook, A., Patrick, J. W., Chen, X. Y., and Ruan, Y. L. (2014). Silencing the vacuolar invertase gene *GhVIN1* blocks cotton fiber initiation from the ovule epidermis, probably by suppressing a cohort of regulatory genes via sugar signaling. *Plant J.* 78, 686–696. doi: 10.1111/tpj.2014.78.issue-4
- Wang, L., Kartika, D., and Ruan, Y. L. (2021). Looking into 'hair tonics' for cotton fiber initiation. *New Phytol.* 229, 1844–1851. doi: 10.1111/nph.v229.4
- Wang, J., Meng, Z., He, H., Du, P., Dijkwel, P. P., Shi, S., et al. (2024). Genome-Wide Analysis of BBX Gene Family in Three Medicago Species Provides Insights into Expression Patterns under Hormonal and Salt Stresses. *Int. J. Mol. Sci.* 25. doi: 10.3390/ijms25115778
- Wang, X., Miao, Y., Cai, Y., Sun, G., Jia, Y., Song, S., et al. (2021). Large-fragment insertion activates gene GaFZ (Ga08G0121) and is associated with the fuzz and trichome reduction in cotton (*Gossypium arboreum*). *Plant Biotechnol. J.* 19, 1110–1124. doi: 10.1111/pbi.13532
- Wang, H., Mei, W., Qin, Y., and Zhu, Z. (2011). 1-Aminocyclopropane-1-carboxylic acid synthase 2 is phosphorylated by calcium-dependent protein kinase 1 during cotton fiber elongation. *Acta Biochim. Biophys. Sin (Shanghai)*. 43, 654–661. doi: 10.1093/abbs/gmr056
- Wu, Y., MaChado, A. C., White, R. G., Llewellyn, D. J., and Dennis, E. S. (2006). Expression profiling identifies genes expressed early during lint fiber initiation in cotton. *Plant Cell Physiol.* 47, 107–127. doi: 10.1093/pcp/pci228
- Xiaokang, Z. (2020). Biomarker discovery using statistical and machine learning approaches on gene expression data.
- Xie, Q., Gao, Y., Li, J., Yang, Q., Qu, X., Li, H., et al. (2020). The HD-Zip IV transcription factor *SHDZIV8* controls multicellular trichome morphology by regulating the expression of *Hairless-2*. *J. Exp. Bot.* 71, 7132–7145. doi: 10.1093/jxb/eraa428
- Xin, Y., Pan, W., Zhao, Y., Yang, C., Li, J., Wang, S., et al. (2025). The NAC transcription factor LpNAC48 promotes trichome formation in *Lilium pumilum*. *Plant Physiol.*
- Xu, B., Gou, J. Y., Li, F. G., Shangguan, X. X., Zhao, B., Yang, C. Q., et al. (2013). A cotton BURP domain protein interacts with α -expansin and their co-expression promotes plant growth and fruit production. *Mol. Plant.* 6, 945–958.
- Xu, Z., Zhang, G., Chen, J., Ying, Y., Yao, L., Li, X., et al. (2024). Role of *Rubus chingii* BBX gene family in anthocyanin accumulation during fruit ripening. *Front. Plant Sci.* 15, 1427359. doi: 10.3389/fpls.2024.1427359
- Yang, S., Xue, S., Shan, L., Fan, S., Sun, L., Dong, Y., et al. (2025). The *CsTM* alters multicellular trichome morphology and enhances resistance against aphid by interacting with *CsTIP1;1* in cucumber. *J. Adv. Res.* 69, 17–30. doi: 10.1016/j.jare.2024.04.008
- Yu, H., Zhang, F., Wang, G., Liu, Y., and Liu, D. (2013). Partial deficiency of isoleucine impairs root development and alters transcript levels of the genes involved in branched-chain amino acid and glucosinolate metabolism in *Arabidopsis*. *J. Exp. Bot.* 64, 599–612. doi: 10.1093/jxb/ers352
- Yuan, R., Cao, Y., Li, T., Yang, F., Yu, L., Qin, Y., et al. (2021). Differentiation in the genetic basis of stem trichome development between cultivated tetraploid cotton species. *BMC Plant Biol.* 21, 115. doi: 10.1186/s12870-021-02871-4
- Zarembinski, T. I., and Theologis, A. (1994). Ethylene biosynthesis and action: a case of conservation. *Plant Mol. Biol.* 26, 1579–1597.
- Zhang, M., Xiao, Y., Zeng, J., and Pei, Y. (2017a). PIN-formed protein, a door to reveal the mechanism for auxin-triggered initiation of cotton fiber. *Plant Signal Behav.* 12, e1319031. doi: 10.1080/15592324.2017.1319031
- Zhang, N., Yang, L., Luo, S., Wang, X., Wang, W., Cheng, Y., et al. (2018). Genetic evidence suggests that *GIS* functions downstream of *TCL1* to regulate trichome formation in *Arabidopsis*. *BMC Plant Biol.* 18, 63. doi: 10.1186/s12870-018-1271-z
- Zhang, M., Zeng, J. Y., Long, H., Xiao, Y. H., Yan, X. Y., and Pei, Y. (2017b). Auxin regulates cotton fiber initiation via ghPIN-mediated auxin transport. *Plant Cell Physiol.* 58, 385–397. doi: 10.1093/pcp/pcw203