BMC Plant Biology

RESEARCH Open Access

Transcriptomics and proteomics insights into carotenoid differentiation in tissue cultured *Rehmannia glutinosa* root cambial meristematic cells: REG-CMC1 and non-somaclonal REG-CMC2

Pengfei Zhou[1], Haihua Li[2], Jiahe Li[3], Haoke Zhang[4], Xiaonan Deng[3], Zhiye Huang[3], Yunshuang Yuan[3], Zemei Lai[3], Xiaoyan Zhang[3], Tong Wang[3], Yafei Huang[3], Wener Guo[3], Zhuchao Zhong[3] and Chun Yang[1*]

Abstract

Plant bioactive compounds, particularly carotenoids, anthocyanins and flavonoids, plays a pivotal role in the pigmentation of plant tissues. In the current study, two distinct cambial meristematic cell (CMC) lines named REG-CMC-1 and REG-CMC-2 were derived from *R. glutinosa* roots through tissue culture, with significant differences in carotenoid content. REG-CMC-1 showed a significantly higher carotenoid content compared to REG-CMC-2 cell lines. A total of 1.78 mg/g DW (dry weight) of carotenoids were detected from REG-CMC-1 cells, compared to only 0.1 mg/g DW from REG-CMC-2 cells. Comprehensive transcriptomics and proteomics analysis were conducted further to investigate the difference in carotenoid content between two cell lines. Transcriptomics analysis of REG-CMC-1 and REG-CMC-2 cell lines identified 14,773 differentially expressed genes (DEGs) including genes encoding xanthoxin dehydrogenase, zeta-carotene desaturase, zeaxanthin epoxidase and 9-cis-epoxycarotenoid dioxygenase, which may be involved in carotenoid production. Significantly enriched genes were involved in secondary metabolite biosynthesis, plant hormone signaling and stress responses. Proteomics study identified 5,361 unique proteins, of which 1,710 were differentially expressed, with critical categories related to proteolysis, carbohydrate metabolism and phosphorylation highlighting various enzymes critical to carotenoid biosynthesis pathway, including lycopene beta-cyclase and phytoene desaturase. These findings indicate the presence of carotenoid biosynthesis-related components in *R. glutinosa* root cambial meristematic cells and suggest potential relevance for biotechnological studies of carotenoid production.

Keywords Pigments, Carotenoid, Secondary metabolites, Tissue culture, Transcriptomics, Proteomics

*Correspondence:
Chun Yang
yangchunangel@163.com
[1]Dongguan Key Laboratory of Pharmaceutical Biotechnology, School of Basic Medical Sciences, Guangdong Medical University, Dongguan 523808, China
[2]School of Medicine and Health, Guangdong Innovative Technical College, Dongguan 523946, China
[3]School of Pharmacy, Guangdong Medical University, Dongguan 523808, China
[4]The First Dongguan Affiliated Hospital, Guangdong Medical University, Dongguan 523808, China

Introduction

Rehmannia glutinosa, a member of the Orobanchaceae family, is a traditional herb commonly used in East Asia, particularly in countries like Japan, China and Korea [1, 2]. This herb is utilized as dried, fresh and prepared root forms. Whether fresh, dried or steaming, tuberous roots have long been used to treat various ailments. It can be used to treat rheumatoid arthritis, fibromyalgia, asthma, hypertension, menopause symptoms, kidney problems, adrenal gland abnormalities, impotence and hair loss. In addition, *R. glutinosa* has been used to treat diabetes since ancient times. The polysaccharides and oligosaccharides of *R. glutinosa* are the active ingredients that have been studied most for treating diabetes [3]. The health benefits are primarily attributed to various phytochemicals, such as iridoid glycosides, phenylethanoid glycosides, polysaccharides, phenolic acids and rehmannoside A [4].

Research on the bioactive compounds of *R. glutinosa* has identified several novel and known compounds with potential therapeutic properties [5, 6]. Previous phytochemical investigations of *R. glutinosa* species have isolated and identified several chemicals, such as iridoid glycosides, ionone glycosides, phenethyl alcohol glycosides, frehmaglutoside G, frehmaglutoside H, a novel carotenoid glycoside named neorehmannioside etc [7–10]. , collectively underscoring the rich chemical diversity of *R. glutinosa* roots and highlighting their potentiality as a source for developing novel pharmacological agents. Active substances of *R. glutinosa* roots affect bone remodeling, the neurological, circulatory, cardiovascular and immunological systems. Secondary metabolites are usually produced in small quantities within specific tissues of certain plant species. Key classes of these compounds include alkaloids, phenylpropanoids, flavonoids and terpenoids, all of which are synthesized through complex biosynthetic pathways [2]. To date, approximately 200,000 metabolites have been identified, with several utilized in pharmaceuticals, food coloring and flavoring, pest and disease control agents, as well as in cosmetics and fragrances [11–13].

Carotenoids, pivotal terpenoid compounds, serve dual roles as photosynthetic pigments and antioxidant agents in plants, while also providing protective functions against abiotic and biotic stresses. Carotenoids are the primary source of yellow, orange and red pigments in plants and widely found in nature [14, 15]. The majority of carotenoids are found in higher plants, particularly in their leaves, flowers, and fruits [16]. In medicinal plants like *R. glutinosa*, the carotenoid biosynthetic pathway intersects with iridoid production, influencing both pigment accumulation and secondary metabolite synthesis [17, 18]. Consumption of carotenoids led to a reduced risk of several diseases, including cancers, cardiovascular disease, age-related macular degeneration and photosensitivity due to UV exposure [19]. The carotenoids including beta-carotene, lycopene, lutein and zeaxanthin are most commonly studied for these effects. These benefits are partly attributed to the antioxidant properties. Additionally, beta-carotene has the added advantage of being converted into vitamin A in the body. Natural food colorants like palm oil, paprika, annatto and saffron also derive their color from carotenoids [20].

The global population is growing, making it increasingly challenging to sustainably supply fresh and nutritious foods for future generations. Therefore, it is necessary to produce compounds like carotenoids with pharmacological value in alternative ways to meet the higher demand. One of the most reliable, environmentally friendly, sustainable and predictable processes for creating plant metabolites for the pharmaceutical sector is plant cell culture. The biosynthesis of plant metabolites like carotenoids, is well studied at the genetic and post-translational levels [21], but some knowledge gaps still exist. Transcriptional and post-transcriptional regulations are recognized as primary mechanisms controlling carotenoid biosynthesis while deeper research is needed to better understand.

In the present study, two distinct *R. glutinosa* cell lines, designated as REG-CMC-1 and REG-CMC-2, were investigated. REG-CMC-1 represents the original *R. glutinosa* CMC line (cell line 1), whereas REG-CMC-2 denotes a subsequently established non-somaclonal line (cell line 2), developed following the previously applied procedure for REG-CMC-1 [22]. These two cell lines differ in pigmentation color and this work aims to identify high-confidence candidate genes, proteins and pathways associated with carotenoid differentiation. The proposed regulatory networks represent hypothesis-generating models that can inform and prioritize future functional studies.

Results

Morphological and cytological features of REG-CMC-1 and REG-CMC-2 cell lines

For induction of REG-CMC-2 cell line, fresh roots of *R. glutinosa* (three months old) were utilized as explants for the initiation of cell cultures (Fig. 1A). The root tissues were sectioned into thin slices and cultured on induction medium to promote callus formation (Fig. 1B). After 7–10 days of incubation (24 h darkness at 25 °C), visible callus development was observed along the cut surfaces of the explants (Fig. 1C). The actively proliferating calli were subsequently transferred to liquid medium, where they formed compact yet friable aggregates, indicative of vigorous growth (Fig. 1D). Microscopic examination of the resulting suspension cultures revealed dispersed and elongated cell clusters (Fig. 1E). Furthermore, light

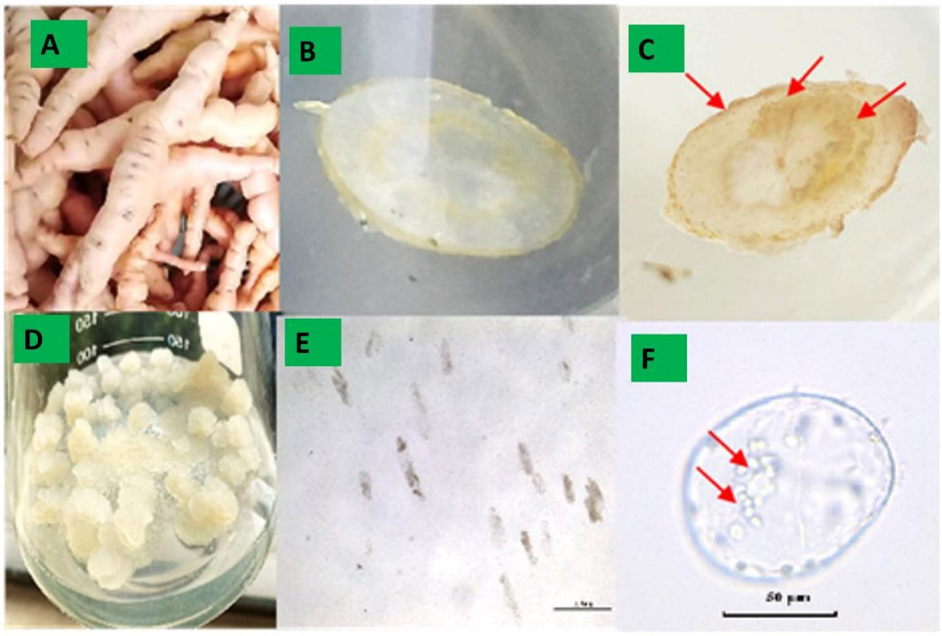

Fig. 1 Induction of *R. glutinosa* REG-CMC-2 cell line. **A** Fresh root of *R. glutinosa* (three month old); **B** Sliced R. glutinosa; **C** Sliced *R. glutinosa* cultured for 10 days (red arrows indicate developing callus); **D** Callus; **E**. Micrographs of REG-CMC-2 cell clusters. **F** REG-CMC-2 cells showing vacuoles (red arrows) under a light microscope (Scale bar = 50 µm)

microscopy of the REG-CMC-2 cell line showed distinct vacuoles within the cells suggesting high cell viability and active metabolic status (Fig. 1F).

The REG-CMC-1 and REG-CMC-2 cell cultures predominantly consisted of individual cells or small clusters containing only a few cells. In contrast, dedifferentiated cell (DDC) cultures were characterized by dense, compact aggregates composed of numerous tightly packed cells [22]. Anatomical examination further indicated that REG-CMC-1 and REG-CMC-2 cells (CMCs) contained multiple small vacuoles, whereas DDCs exhibited a single large vacuole occupying most of the cell volume. These cytological characteristics are consistent with previously reported features of *R. glutinosa* cell cultures [23, 24].

Growth characteristics, carotenoid accumulation and carotenoid content across successive cell passages

The growth kinetics of the two *Rehmannia glutinosa* cell lines, REG-CMC-1 and REG-CMC-2, were assessed over an 18-day culture period (Fig. 2A). Both lines exhibited a similar sigmoidal growth pattern, with steady biomass accumulation from day 6 to day 15, followed by a slight decline at day 18. REG-CMC-1 consistently showed slightly higher dry weight compared with REG-CMC-2 throughout the culture period, reaching maximum values of approximately 12–13 g/L at day 15. These results indicate that both cell lines maintain stable proliferation under identical culture conditions, with REG-CMC-1 demonstrating marginally greater growth potential.

Carotenoid content was quantified in both REG-CMC-1 and REG-CMC-2 during the same culture period using a gravimetric method (Fig. 2B). Carotenoid levels in REG-CMC-1 increased gradually from day 6 to day 15, reaching a maximum of approximately 1.78 mg/g, followed by a slight decrease at day 18. A similar pattern was observed in REG-CMC-2, with peak accumulation (~0.1 mg/g) also occurring at day 15. However, carotenoid content in REG-CMC-1 was significantly higher than in REG-CMC-2 across all time points, suggesting line-specific differences in carotenoid biosynthesis capacity.

To evaluate the stability of carotenoid production during long-term maintenance, carotenoid content was analyzed across multiple subcultures (passages 25–45) for both cell lines (Fig. 2C). REG-CMC-1 maintained carotenoid concentrations between 1.6 and 2.0 mg/g DW, while REG-CMC-2 showed relatively stable values ranging from 0.08 to 0.10 mg/g DW. No significant fluctuations were observed in either line across passages, indicating that both REG-CMC-1 and REG-CMC-2 retained stable biosynthetic potential during continuous sub culturing.

Differences in carotenoids contents in REG-CMC-1 and REG-CMC-2 cell lines

REG-CMC-1 cell lines were characterized by their yellow pigmentation and REG-CMC-2 cell lines were characterized by the absence of coloured pigments. Callus cultures

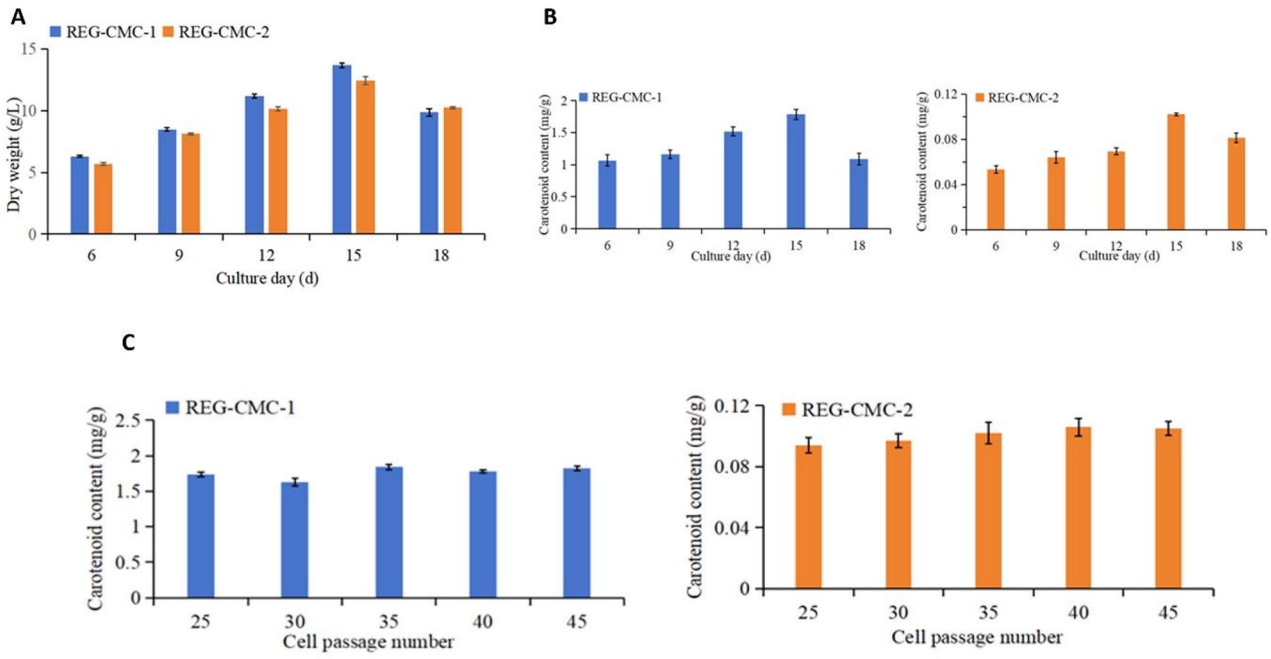

Fig. 2 Growth and Carotenoid Accumulation Characteristics of *R. glutinosa* Cell Suspension Lines. **A** Growth kinetics (g/L) of the two *R. glutinosa* cell lines, REG-CMC-1 and REG-CMC-2 during culture period; **B** Detected carotenoid content in REG-CMC-1 and REG-CMC-2 cell lines; **C** Carotenoid content across multiple subcultures

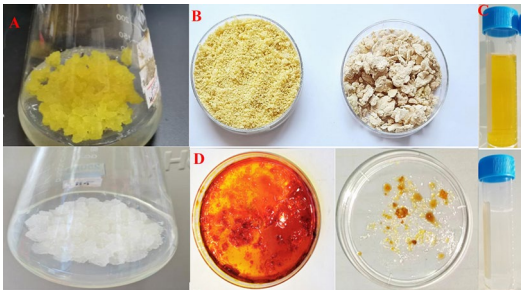

Fig. 3 Extraction of metabolites from REG-CMC-1 and REG-CMC-2 cell lines; **A** Callus from REG-CMC-1 (yellow colour, up) and REG-CMC-2 (white colour, down) cells; **B** REG-CMC-1 lyophilized powder (left) and REG-CMC-2 lyophilized powder (right). **C** Extracts from REG-CMC-1 (up) and REG-CMC-2 (down) cell lines; **D** Carotenoid crude extracts (REG-CMC-1) (left) and (REG-CMC-2) (right)

of *Rehmannia glutinosa* REG-CMC-1 and REG-CMC-2 cell lines exhibited distinct morphological and pigmentation characteristics. The REG-CMC-1 callus displayed a bright yellow coloration, whereas the REG-CMC-2 callus appeared white and compact (Fig. 3A). Lyophilization of the cultured cells produced fine powders, with the REG-CMC-1 powder showing a yellow colour and the REG-CMC-2 powder exhibiting off-white appearance (Fig. 3B). Metabolite extraction from the lyophilized powders resulted in liquid extracts that retained the characteristic coloration of each cell line yellow for REG-CMC-1 and nearly colorless for REG-CMC-2 (Fig. 3C). Further extraction and concentration yielded carotenoid-rich crude extracts, with REG-CMC-1 producing an intense orange-red pigment and REG-CMC-2 showing a much lighter off-white extract (Fig. 3D). These visual differences suggest distinct metabolite profiles between the two cell lines, potentially reflecting variations in pigment accumulation and secondary metabolite biosynthesis. REG-CMC-1 cells produced 1.78 mg of total carotenoids per g of DW, whereas REG-CMC-2 cells yielded 0.10 mg/g DW. These values represent the quantified carotenoid content extracted from the dried cell biomass. Meanwhile, the carotenoid contents of REG-CMC-1 and REG-CMC-2 cells determined by Visible Spectrophotometry peaked on 15th day, with values of 13.14 mg/g and 0.0096 mg/g, respectively.

Transcriptomics analysis of REG-CMC1 and REG-CMC2 cell lines

Quality control of transcriptomics data

Cultured REG-CMC1 and REG-CMC2 cell lines were utilized to construct libraries for high-throughput sequencing to elucidate the molecular basis of carotenoid production. Transcriptome sequencing yielded 46.56 Gb of clean data, with each sample producing 7 Gb and Q30 bases at 93%. Three samples of REG-CMC1 yielded raw reads of approximately 52.9 to 50.9 million, clean reads ranging from 51.5 to 49.9 million and clean bases between 7.73 G and 7.48 G respectively. The average Q20, Q30 and GC contents were 98%, 93% and 45% respectively. In REG-CMC2 cell lines, raw reads varied

from 55.2 to 51.1 million, clean reads were between 53.7 and 49.9 million and clean bases ranged from 8.05 G to 7.48 G. Q20 values were approximately 98.4%, Q30 values exceeded 94% and GC content was around 45% respectively.

REG-CMC1 cell lines produced 97.47% clean reads, 0.19% low-quality reads and 2.34% adapter-related reads. Meanwhile, REG-CMC2 cell lines generated 97.27% clean reads, 0.17% low-quality reads and 2.56% adapter-related reads. Regional distribution of reads across genomic features is presented in Fig. 4A. The regional distribution of reads in REG-CMC1 cell lines included 52.44% exon, 8.87% intron and 38.67% intergenic regions. In REG-CMC2 cell lines, 52.74% exon, 8.92% intron and 38.33% intergenic regions were identified. Principal Component Analysis (PCA) performed on the REG-CMC1 and REG-CMC2 cell lines revealed that the first principal component (PC1) accounted for 62.07% of the total variation in the dataset, representing the most significant axis along which the samples differ. The second principal component (PC2) explained an additional 10.35% of the variation, capturing the next most important differences not described by PC1. Together, these two components accounted for approximately 72.42% of the overall variance, indicating that the majority of differences between the two cell lines can be effectively represented using just these two dimensions. This level of explained variance suggests a clear distinction between REG-CMC1 and REG-CMC2, which is often illustrated through a PCA plot showing strong separation along PC1 and PC2 (Fig. 4B).

Identified genes from REG-CMC1 and REG-CMC2 cell lines

Transcriptomics analysis between REG-CMC1 and REG-CMC1-2 cell lines identified 14,773 DEGs, with 8,776 and 5,997 up regulated and down regulated genes respectively. List of some of the identified genes involved in carotenoid biosynthesis pathways is presented in Table 1. A large number of genes involved in carotenoid synthesis were significantly upregulated, reflecting their potential role in metabolic processes related to secondary metabolites. Many of the upregulated genes encode enzymes such as xanthoxin dehydrogenase and zeta-carotene desaturase, which are crucial in the biosynthesis of carotenoids, as well as secondary metabolites involved in cellular processes. These enzymes are also associated with KEGG pathways involved in carotenoid biosynthesis and general metabolic processes, showing a strong link to the biosynthesis of secondary metabolites. Several transcripts encoding zeta-carotene desaturase and beta-carotene isomerase, are similarly involved in the carotenoid biosynthesis pathway. Genes encoding zeaxanthin

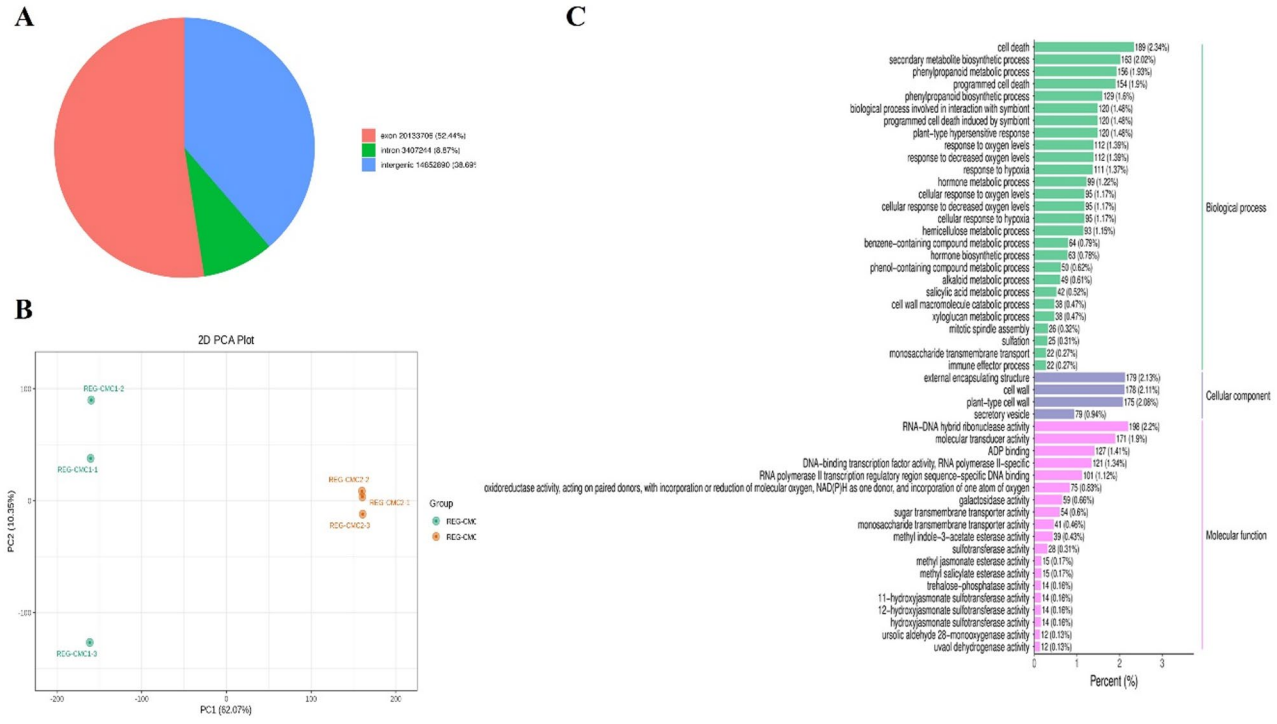

Fig. 4 Overview of transcriptomics profiling and functional enrichment analysis of REG-CMC1 vs. REG-CMC2 cells. **A** The distribution map of the comparison area; Reads that map to introns could come from precursor mRNA or introns that retain variable splicing events. Meanwhile, reads aligned to intergenic regions might originate from non-coding RNA or minor DNA contamination. **B** Principal Component Analysis (PCA) of REG-CMC1 and REG-CMC2 cell lines; **C** GO enrichment diagram of differentially expressed genes from the transcriptomics analysis of REG-CMC1 vs. REG-CMC2 cells

Table 1 List of some of the identified genes associated with carotenoid biosynthesis and other metabolic pathways

Gene ID	Gene Name	log_2 Fold Change	Regulation	KEGG pathways
IMPGRGL1N62432	*Xanthoxin dehydrogenase*	5.388	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N46692	*Xanthoxin dehydrogenase-like*	2.01201	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N38937	*Zeaxanthin epoxidase*	1.98867	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N15039	*Zerumbone synthase-like*	-1.4442	Down (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N11711	*Phytoene dehydrogenase*	2.454676	Up (REG-CMC1)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N40609	*15-cis-zeta-carotene isomerase*	1.70329	Up (REG-CMC1)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N43793	*Geranylgeranyl pyrophosphate synthase*	2.11556	Up (REG-CMC1)	Terpenoid backbone biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N954	*Zerumbone synthase*	1.42646	Up (REG-CMC1)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N64635	*Xanthoxin dehydrogenase*	3.64784	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N31768	*Xanthoxin dehydrogenase*	5.27378	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N48381	*Indole-3-acetaldehyde oxidase*	2.39483	Up (REG-CMC1)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N21853	*9-cis-epoxycarotenoid dioxygenase NCED5*	-2.3747	Down (REG-CMC1)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N49156	*Zeaxanthin epoxidase*	1.1864	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N47039	*Xylulose kinase*	2.59515	Up (REG-CMC1)	Pentose and glucuronate interconversions; Metabolic pathways
IMPGRGL1N32688	*9-cis-epoxycarotenoid dioxygenase NCED1*	3.76058	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N32067	*9-cis-epoxycarotenoid dioxygenase NCED1*	3.47968	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N10527	*Carotenoid cleavage dioxygenase 4*	3.75904	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N38940	*Zeaxanthin epoxidase*	6.10078	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N32702	*9-cis-epoxycarotenoid dioxygenase NCED1*	6.51771	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N31611	*Probable carotenoid cleavage dioxygenase 4*	-4.1789	Down (REG-CMC1)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N57599	*Geranylgeranyl pyrophosphate synthase*	-1.0107	Down (REG-CMC2)	Terpenoid backbone biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N31767	*Xanthoxin dehydrogenase*	1.42457	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N5897	*Caffeoylshikimate esterase-like*	1.70753	Up (REG-CMC2)	Phenylpropanoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N41127	*Probable carotenoid cleavage dioxygenase 4*	-1.5563	Down (REG-CMC1)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N61345	*Prolycopene isomerase*	-1.0609	Down (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites
IMPGRGL1N13155	*Xanthoxin dehydrogenase-like*	1.36413	Up (REG-CMC2)	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites

epoxidase, are essential for the conversion of carotenoid intermediates into more complex molecules, linking them to important physiological processes like photoprotection and plant development. There are also important downregulated genes which encodes 9-cis-epoxycarotenoid dioxygenase, an enzyme involved in the breakdown of carotenoids. This downregulation might indicate a reduction in the catabolic pathways of carotenoids

under specific conditions, potentially highlighting an increased focus on carotenoid biosynthesis rather than degradation. Downregulated gene encoding *abscisic acid 8'-hydroxylase*, are associated with phytohormone regulation, particularly abscisic acid metabolism.

This analysis also identified genes encoding for proteins that belong to diverse functional categories, including proteins with oxidase activity, chaperonins involved in RNA degradation and cytochrome P450 enzymes associated with various secondary metabolite pathways. A transcript encodes for a carotenoid 9, 10 (9', 10')-cleavage dioxygenase, a crucial enzyme in carotenoid degradation, which might be regulated under certain environmental or metabolic conditions [25]. Another identified gene encoding caffeoyl-CoA O-methyltransferase, which is involved in the biosynthesis of lignin and phenolic compounds and plays important roles in plant pigmentation and defense mechanisms [26]. It catalyzes the methylation of caffeoyl-CoA, a key step in the formation of secondary metabolites, including flavonoids and anthocyanins which are responsible for the blue, red, and purple colors in flowers and fruits [27].

In contrast, several other identified genes encode enzymes unrelated to pigment biosynthesis, such as ribonucleoside-diphosphate reductase and DNA polymerase theta. These enzymes are likely involved in DNA repair and cell division, essential processes for maintaining cellular functions in pigment-producing cells, although they do not directly participate in pigment biosynthesis. Another gene which encodes a MYB transcription factor (Myb4-like) associated with pigment regulation and known regulators of flavonoid biosynthesis, a pathway crucial for the production of anthocyanins [28]. The presence of a MYB-related protein suggests that this gene may play a role in regulating pigment production, specifically within the flavonoid biosynthetic pathway. Additionally, a homeobox-leucine zipper protein (HAT4-like) encoding gene, may influence pigment production indirectly through its involvement in plant development and stress responses. Another gene of interest encodes a multidrug resistance protein (DETOXIFICATION 29 isoform X1), which may be involved in the transport of secondary metabolites, including pigments. Similarly, a multidrug resistance protein from the MATE family and acetolactate synthase 2 (chloroplastic-like), are likely involved in broader metabolic functions, such as amino acid synthesis and transport, which could indirectly influence pigment production.

Identified transcription factor families from the transcriptomics analysis

Transcriptomics analysis revealed a diverse array of TFs potentially involved in regulating pigment production in REG-CMC1 and REG-CMC2 cell lines. Among the most prominent TF groups identified are the ethylene-responsive transcription factors (ERFs), which include isoforms from the RAP2, RAV1, CBF, and ERF families. These TFs are implicated in regulating pigmentation in response to environmental stress signals, such as drought or pathogen attack. Such stress often triggers the production of anthocyanins and other pigments as part of the defense mechanisms of plants [29]. Another significant group is the dehydration-responsive element-binding proteins (DREBs), including DREB1C and DREB1B which activate stress-related genes that influence pigment production. Similarly, PHD finger proteins, such as ALFIN-LIKE, are involved in various developmental processes, including those regulating pigment formation [30, 31]. AP2-like factors, including APETALA 2 (AP2) and WRI1, also play essential roles in flower and fruit development, processes that are closely linked to pigment synthesis [32]. Moreover, the ARID family of TFs, represented by genes such as *separase* and *aminotransferase*, are primarily involved in cell cycle regulation and metabolic control, indirectly influencing pigment production [33].

B3 domain-containing TFs, including RAV-like factors, The BES1 (brassinosteroid signaling) TF family, BES1 (brassinosteroid signaling) were also identified which are essential for regulating pigment biosynthesis [34]. Additionally, bHLH (basic helix-loop-helix) transcription factors, such as MYC2, play key roles in regulating stress responses and secondary metabolism, including jasmonate signaling, which is important for pigmentation under environmental stress [35]. Other groups including bZIP (basic leucine zipper) transcription factors, including TGA7, C2H2 zinc-finger domain TF family and the C2C2-CO-like family, which includes CONSTANS-like proteins are implicated in stress response pathways and plant hormone signaling, all of which are crucial for regulating pigment production [36]. Similarly, NAC transcription factors, MADS-box transcription factors, such as MADS-MIKC and MADS-M-type which are involved in various developmental and stress-response processes, may indirectly regulate genes related to secondary metabolite production, including carotenoids [37].

Identified key GO and KEGG enrichment pathways

The transcriptomics analysis identified a range of biological processes associated with secondary metabolite biosynthesis and plant responses. GO enrichment analysis of the transcriptomics analysis of REG-CMC1 vs. REG-CMC2 cell lines is presented in Fig. 4C. Key processes included the secondary metabolite biosynthetic process (GO: 0044550) and the phenylpropanoid biosynthetic process (GO: 0009699), both essential for producing plant secondary metabolites. Several processes related to oxygen levels and responses to hypoxia were also noted, including cellular response to hypoxia (GO: 0071456)

and response to decreased oxygen levels (GO: 0036293), highlighting the plant's adaptive mechanisms to low-oxygen environments. Cell death (GO: 0008219) and programmed cell death (GO: 0012501) also indicated critical cellular events that may be linked to stress responses or development. The involvement of pathways related to plant-type hypersensitive response (GO: 0009626) and immune effector processes (GO: 0002252) suggests mechanisms for plant defense against pathogens.

Furthermore, metabolic processes involving various compounds, including salicylic acid (GO: 0009696), jasmonic acid (GO: 0009694) and alkaloids (GO: 0009820), indicated a potential complex network of hormonal and secondary metabolite pathways. The presence of processes related to cell wall macromolecule catabolism (GO: 0016998) and hormone biosynthesis (GO: 0042446) further emphasized the role of these pathways in plant structure and function. These findings illustrated the intricate biological processes underlying plant metabolism, stress responses and interactions with symbionts, reflecting the complexity of plant biology and the importance of secondary metabolites in growth and defense. The GO enrichment circular diagram of differential genes is presented in Fig. 5A for better clarification of GO analysis.

The distribution of differential genes annotated to KEGG pathways is illustrated in Fig. 5B. Among the differentially expressed genes, 11,741 common DEGs between the REG-CMC1 and REG-CMC1-2 cell lines were linked to 145 pathways. Key categories include "Environmental Information Processing", highlighting pathways such as plant hormone signal transduction (353 genes, 7.28%) and the MAPK signaling pathway in plants (211 genes, 4.35%). Under "Metabolism", several pathways were identified, including the biosynthesis of various plant secondary metabolites (1432 genes, 59.55%), starch and sucrose metabolism (198 genes, 4.09%), and phenylpropanoid biosynthesis (142 genes, 2.93%), indicating diverse metabolic activities. Additional pathways focused on specific compounds, such as isoquinoline alkaloid biosynthesis (67 genes, 1.38%), anthocyanin biosynthesis (16 genes, 0.33%), and flavonoid biosynthesis (62 genes, 1.28%), emphasizing the complexity of plant secondary metabolite production. Other metabolic processes, including fatty acid elongation (31 genes, 0.64%), were also noted, alongside pathways related to plant-pathogen interactions (395 genes, 8.15%) and ribosome biogenesis (154 genes, 3.18%) in eukaryotes.

Proteomics analysis of REG-CMC1 and REG-CMC2 cell lines

Identified differentially expressed proteins (DEPs)

The comprehensive protein identification yielded 23,895 peptides, which successfully identified 5,361 unique proteins. To investigate the proteomic differences between

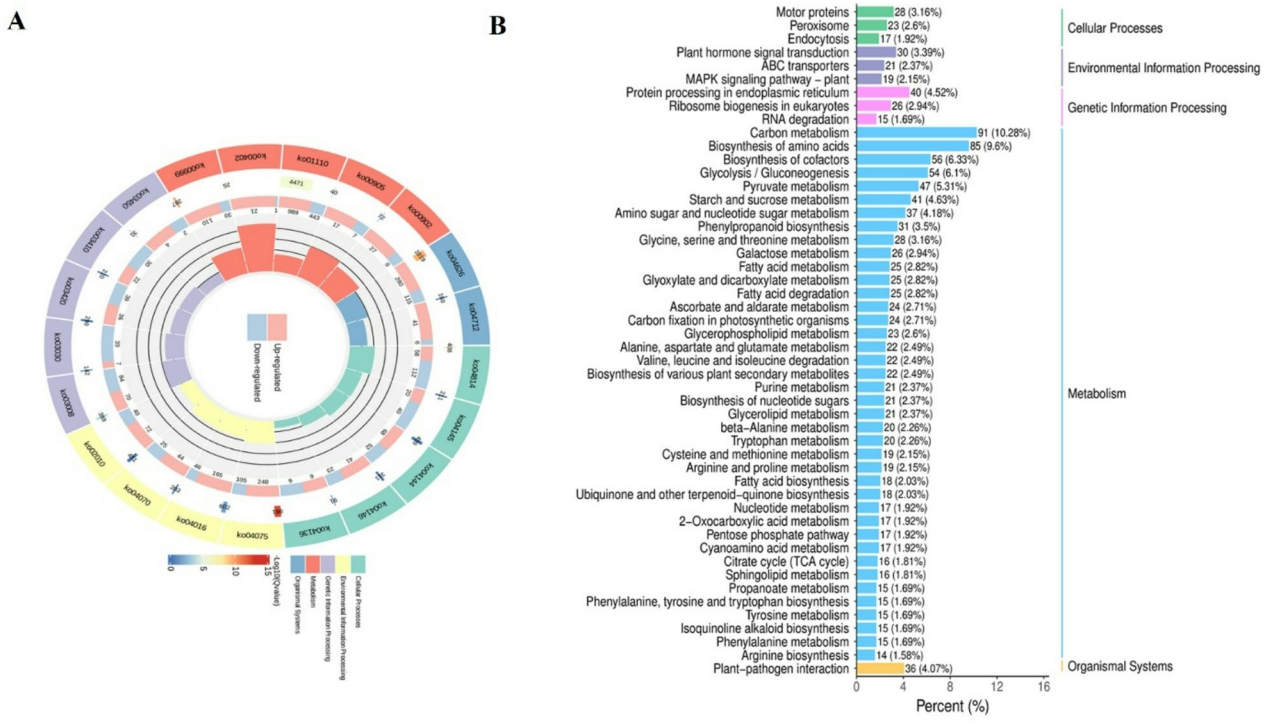

Fig. 5 GO enrichment and KEGG pathway annotation of differentially expressed genes. **A** Circular diagram presenting GO enrichment of DEGs. The figure displays a circular diagram illustrating GO enrichment analysis; **B** Distribution of differential genes annotated to KEGG pathways. The numbers within the figure indicated the count of differential genes annotated to each pathway

the REG-CMC1 and REG-CMC2 cell lines, principal component analysis (PCA) was performed. The PCA plot (Fig. 6A) demonstrated a clear separation between the two groups along the first principal component (PC1), which explained 84% of the total variance. This distinct clustering pattern indicates substantial differences in the overall protein expression profiles between REG-CMC1 and REG-CMC2 cell lines. The differential expression analysis identified a total of 1,710 differentially expressed proteins (DEPs) between REG-CMC1 and REG-CMC2 cell lines (Fig. 6B). Among these, 1,005 proteins were significantly up regulated, while 705 were down regulated in REG-CMC1 relative to REG-CMC2 cell lines indicating a widespread reprogramming of protein expression between the two cell lines. A volcano plot was used to visualize the distribution of DEPs based on $\log_2$ fold change and $-\log_{10}$ (p-value) (Fig. 6C). Red and green dots represent significantly up regulated and down regulated proteins, respectively, whereas gray dots indicate non-significant changes. The clear asymmetry between up- and down regulated proteins reinforces the observation that REG-CMC1 cells exhibit marked up regulation of numerous proteins compared to REG-CMC2. To further characterize the expression patterns of DEPs, hierarchical clustering analysis was conducted, and the results were visualized in a heatmap (Fig. 6D). The heatmap revealed distinct clustering of samples consistent with the PCA results, with REG-CMC1 and REG-CMC2 cell lines forming two separate branches. The differential expression patterns displayed strong internal consistency within each group, further confirming the reproducibility of the proteomics data.

Domain annotation results of identified proteins

The domain annotations of identified proteins are illustrated as bar chart in Fig. 7A. Protein domains were categorized into active site, binding site, domain, family, homologous superfamily and repeat. In the active site category, 30 proteins (1.96%) were classified as serine/threonine protein kinases, while 21 proteins (1.37%) contained the ATP-binding domain associated with protein kinases. Within the domain category, the protein kinase domain was identified in 45 proteins (2.94%), followed by the ABC transporter-like ATP-binding domain and AAA+ family with 17 proteins (1.11%) and the ATPase domain with 16 proteins (1.05%). The cytochrome P450 family exhibited the highest representation, encompassing 16 proteins (1.05%), succeeded by the short-chain dehydrogenase/reductase (SDR) and peptidase S10 families, each containing 12 proteins (0.78%) and the aquaporin transporter with 10 proteins (0.65%). In the homologous superfamily category, significant entries included the P-loop containing nucleoside triphosphate hydrolase with 106 proteins (6.93%), the NAD (P) binding domain superfamily with 61 proteins (3.99%), the alpha/beta hydrolase fold with 54 proteins (3.53%) and the protein kinase-like domain superfamily with 53 proteins (3.46%). Additionally, the glycoside hydrolase superfamily comprised 40 proteins (2.61%) and the armadillo-type fold was represented by 26 proteins (1.7%). The WD40 repeat identified 16 proteins (1.05%) as a

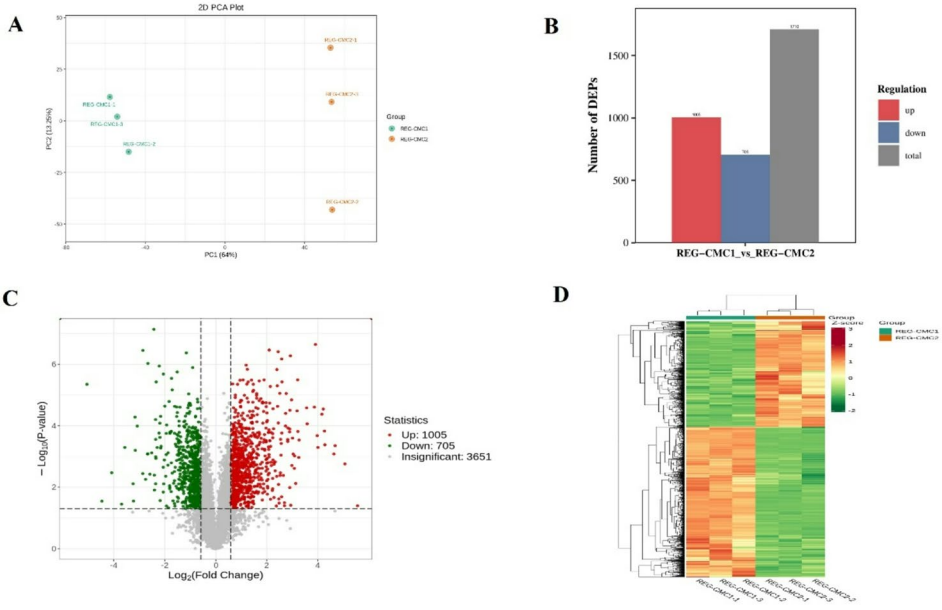

Fig. 6 Overview of the proteomic analysis of differentially expressed proteins. **A** PCA of proteomics samples; **B** Statistical chart of differences between protein groups; **C** Volcano plot of differential protein expression; **D** Heat map showing differential protein clusters; the rows represent protein clusters, while the columns correspond to sample clusters

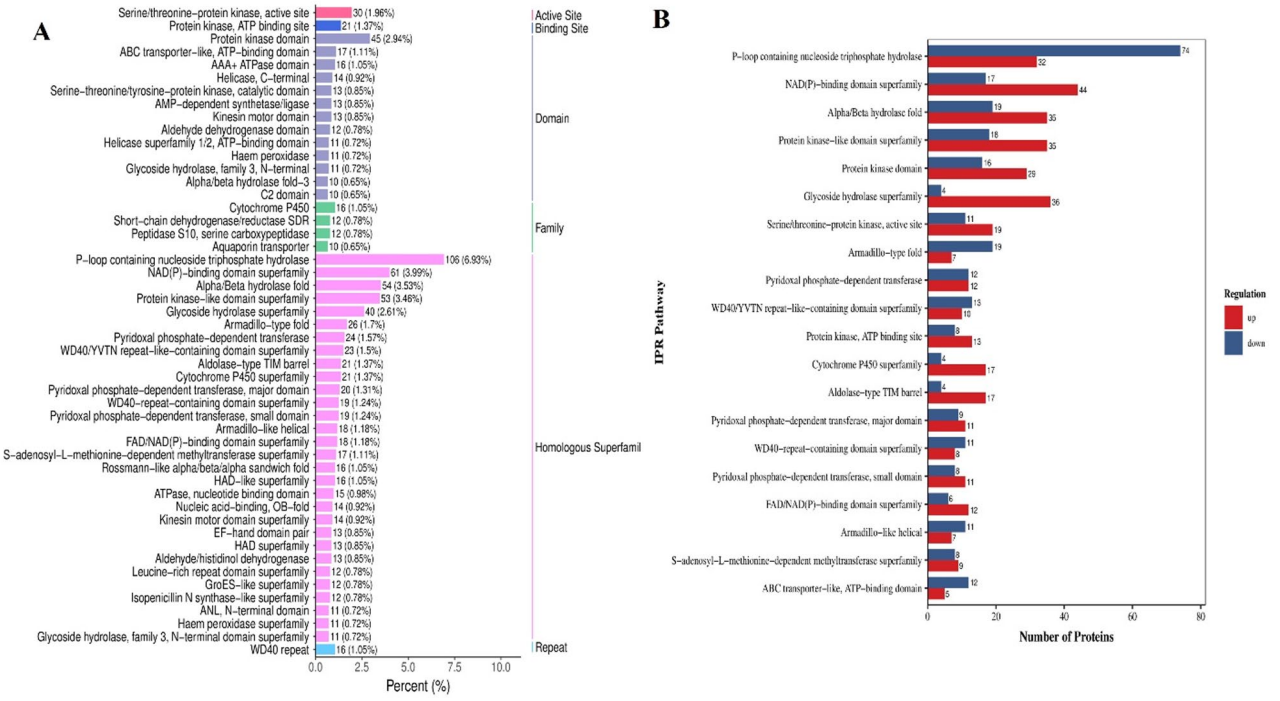

Fig. 7 Domain analysis of DEPs identified from proteomics analysis. **A** Domain annotations of differentially expressed proteins; **B** Comparison of domain annotations for up regulated and down regulated differentially expressed proteins

significant repeat domain in this analysis. The comparison of domain annotations for up regulated and down regulated differentially expressed proteins is presented in Fig. 7B.

Functional classification of deps in REG-CMC1 and REG-CMC2 cell lines

The identified proteins were functionally classified using the GO enrichment analysis. Among the biological processes, a total of 56 proteins (6.09%) were enriched in proteolysis, followed closely by carbohydrate metabolic processes (53 proteins, 5.76%) and protein phosphorylation (52 proteins, 5.65%). Additional categories included phosphorylation (25 proteins, 2.72%), methylation (22 proteins, 2.39%), microtubule-based movement (18 proteins, 1.96%), and intracellular protein transport (16 proteins, 1.74%). Regarding cellular compartment localization, the nucleus and cytoplasm contained the highest number of identified proteins, totaling 80 (11.08%). Other significant compartments included the chloroplast (37 proteins, 5.12%), membrane (32 proteins, 4.43%), plant membrane (28 proteins, 3.88%), cellular region (28 proteins, 3.88%), cytosol (23 proteins, 3.19%) and nucleolus (18 proteins, 2.49%). In terms of molecular functions, the largest category was metal ion binding, with 134 proteins (10.82%). Other important categories included hydrolase activity (51 proteins, 4.12%), oxidoreductase activity (45 proteins, 3.63%), DNA binding (43 proteins, 3.47%), RNA binding (36 proteins, 2.91%), and heme binding (34 proteins, 2.74%) (Fig. 8A). KOG analysis (Fig. 8B) revealed that, within the general function prediction category, approximately 107 proteins were up regulated, while 79 proteins were down regulated. Importantly, in the categories related to post-transcriptional modifications, protein turnover, and chaperones, 97 proteins were up regulated and 44 proteins were down regulated. Similarly, in the carbohydrate transport and metabolism category, 108 proteins were up regulated, whereas 20 proteins were down regulated.

KEGG pathway enrichment analysis of deps in REG-CMC1 and REG-CMC2 cell lines

KEGG analysis indicated significant enrichment across five major pathways in REG-CMC1 and REG-CMC2 cell lines, including cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems. Within the cellular processes category, motor proteins represented the largest group, comprising 28 proteins (3.16%), followed by pathways such as peroxisome (23 proteins, 2.6%) and endocytosis (17 proteins, 1.92%). In the environmental information processing category, plant hormone signal transduction was the most prominent pathway, with 30 proteins (3.39%), followed by ABC transporters (21 proteins, 2.37%) and MAPK signaling pathways in plants (19 proteins, 2.27%). For genetic information processing, significant pathways included protein processing in the endoplasmic reticulum (40 proteins, 4.52%), ribosome

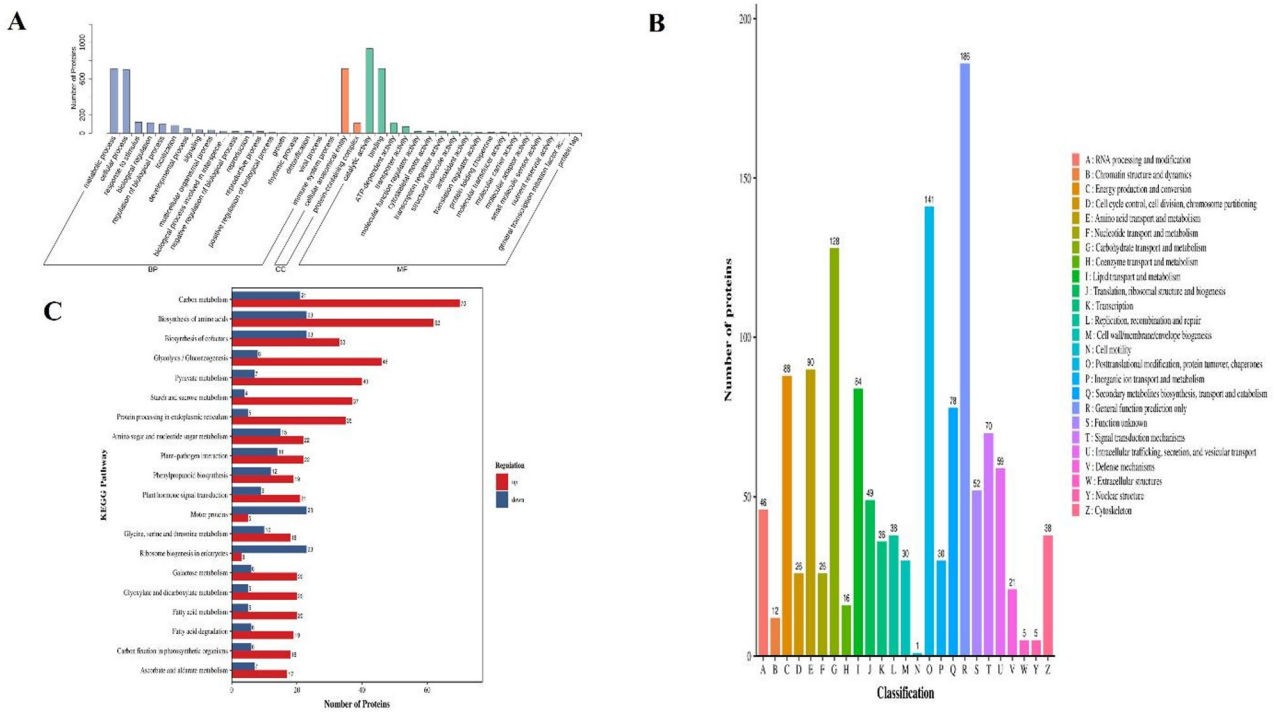

Fig. 8 Functional classification and pathway annotation of DEPs. **A** GO classification of differentially expressed proteins; **B** KOG functional annotation of differentially expressed proteins; **C** Bar chart comparing KEGG pathway classifications

biogenesis in eukaryotes (21 proteins, 2.37%) and RNA degradation (19 proteins, 2.15%). The metabolism category was the largest, featuring key pathways such as carbon metabolism (91 proteins, 10.28%), biosynthesis of amino acids (85 proteins, 9.6%), biosynthesis of cofactors (56 proteins, 6.33%), glycolysis/gluconeogenesis (54 proteins, 6.1%), pyruvate metabolism (47 proteins, 5.31%), starch and sucrose metabolism (41 proteins, 4.63%), and amino sugar and nucleotide sugar metabolism (37 proteins, 4.18%). Furthermore, plant-pathogen interactions (36 proteins, 4.07%) were recognized as the exclusive pathway inside organismal systems.

The differential expression of proteins across various KEGG pathways was quantified, with a bar graph (Fig. 8C). In the carbon metabolism pathway, 70 proteins were up regulated, while 21 were down regulated. For the biosynthesis of amino acids, 62 proteins were up regulated, and 23 were down regulated. In the biosynthesis of cofactors, 33 proteins exhibited up regulation and 23 down regulation. The glycolysis/gluconeogenesis pathway showed 46 proteins up regulated and 8 down regulated. In pyruvate metabolism, there were 40 up regulated and 7 down regulated proteins. For starch and sucrose metabolism, 37 proteins were up-regulated and 4 down regulated. Finally, in the protein processing in the endoplasmic reticulum pathway, 35 proteins were up regulated and 5 down regulated.

Proteins involved in carotenoid synthesis in REG-CMC1 and REG-CMC2 cell lines

A list of some of the identified proteins that may be involved in the carotenoid biosynthesis process is presented in Table 2. Multiple proteins which may contribute to carotenoid metabolism, with several having possible roles in key processes such as the carotenoid biosynthetic process and ubiquinone biosynthesis. Many of the identified proteins may contribute in broader metabolic pathways, such as secondary metabolite biosynthesis. A significant portion of the identified proteins were may be involved in various steps of carotenoid biosynthesis. A hypothetical protein from *Paulownia fortunei*, identified with the carotenoid biosynthesis process and associated with phytoene desaturase activity, which catalyzes the conversion of phytoene into carotenoids [38]. Similarly, a plastid lycopene beta-cyclase 2 from *Bixa orellana*, is involved in the conversion of lycopene into other carotenoids and is localized in the cytoplasm. This protein plays a key role in the synthesis of carotenoids such as lutein and zeaxanthin, critical for photosynthesis and plant pigmentation [39, 40].

Several other proteins, such as amine oxidase and indole-3-acetaldehyde oxidase, are involved in the oxidation of amines and aldehydes, processes that may support broader secondary metabolism, including the biosynthesis of carotenoids and other metabolic intermediates. Further, the identification of aldehyde oxidase enzymes

Table 2 List of some of the identified proteins with possible roles in carotenoid biosynthesis, their potential metabolic pathways and subcellular localizations in REG-CMC1 and REG-CMC2 cell lines

Accession Number	Description	KEGG Pathways	Subcellular localization
IMPGRGL1N11711	Hypothetical protein Pfo_012356 [*Paulownia fortunei*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N20863	Plastid lycopene beta-cyclase 2 [*Bixa orellana*]; lycopene beta-cyclase 2 [*Bixa orellana*]	Carotenoid biosynthesis; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N23314	Amine oxidase [*Handroanthus impetiginosus*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	chloroplast
IMPGRGL1N26012	Hypothetical protein Pfo_017880 [*Paulownia fortunei*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N30391	Hypothetical protein Pfo_008364 [*Paulownia fortunei*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	chloroplast
IMPGRGL1N31036	(-)-Isopiperitenol/ (-)-carveol dehydrogenase, mitochondrial-like [*Olea europaea var. sylvestris*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	chloroplast
IMPGRGL1N31612	Hypothetical protein Pfo_010563 [*Paulownia fortunei*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoskeleton
IMPGRGL1N33469	Prolycopene isomerase [*Buddleja davidii*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	chloroplast
IMPGRGL1N35243	PREDICTED: Indole-3-acetaldehyde oxidase [*Camelina sativa*]; PREDICTED: Indole-3-acetaldehyde oxidase [*Camelina sativa*]; PREDICTED: Indole-3-acetaldehyde oxidase [*Camelina sativa*]; PREDICTED: Indole-3-acetaldehyde oxidase [*Camelina sativa*]; PREDICTED: Indole-3-acetaldehyde oxidase [*Camelina sativa*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N17765	Hypothetical protein Pfo_018186 [*Paulownia fortunei*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Nucleus
IMPGRGL1N20863	Plastid lycopene beta-cyclase 2 [*Bixa orellana*] Lycopene beta-cyclase 2 [*Bixa orellana*]	Carotenoid biosynthesis; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N23314	Amine oxidase [*Handroanthus impetiginosus*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	chloroplast
IMPGRGL1N26012	Hypothetical protein Pfo_017880 [*Paulownia fortunei*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N31612	Hypothetical protein Pfo_010563 [*Paulownia fortunei*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoskeleton
IMPGRGL1N35243	PREDICTED: Indole-3-acetaldehyde oxidase [*Camelina sativa*]; PREDICTED: Indole-3-acetaldehyde oxidase [*Camelina sativa*]; PREDICTED: Indole-3-acetaldehyde oxidase [*Camelina sativa*]; PREDICTED: indole-3-acetaldehyde oxidase [*Camelina sativa*]; PREDICTED: Indole-3-acetaldehyde oxidase [*Camelina sativa*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N38937	hypothetical protein Pfo_001613 [Paulownia fortunei]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	chloroplast
IMPGRGL1N40609	Hypothetical protein Pfo_016404 [*Paulownia fortunei*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N46692	(-)-Isopiperitenol/ (-)-carveol dehydrogenase, mitochondrial-like [*Sesamum indicum*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Endoplasmic reticulum
IMPGRGL1N47347	Reductase [*Handroanthus impetiginosus*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N48375	Indole-3-acetaldehyde oxidase [*Phtheirospermum japonicum*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N48381	Abscisic-aldehyde oxidase [*Striga hermonthica*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N59860	Phytoene desaturase 1 [*Rehmannia glutinosa*]; Phytoene desaturase 1 [*Rehmannia glutinosa*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N62432	(-)-Isopiperitenol/ (-)-carveol dehydrogenase mitochondrial [*Phtheirospermum japonicum*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	chloroplast
IMPGRGL1N954	Hypothetical protein BUALT_Bualt15G0037800 [*Buddleja alternifolia*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoplasm

Table 2 (continued)

Accession Number	Description	KEGG Pathways	Subcellular localization
IMPGRGL1N20863	Plastid lycopene beta-cyclase 2 [*Bixa orellana*]; lycopene beta-cyclase 2 [*Bixa orellana*]	Carotenoid biosynthesis; Biosynthesis of secondary metabolites	Cytoplasm
IMPGRGL1N26012	Hypothetical protein Pfo_017880 [*Paulownia fortunei*]	Carotenoid biosynthesis; Metabolic pathways; Biosynthesis of secondary metabolites	Cytoplasm

suggests a functional diversity in this protein family, with possible roles in both ferredoxin metabolism and carotenoid biosynthesis. These enzymes are commonly associated with the production of aldehyde intermediates in secondary metabolism and are distributed across the cytoplasm and chloroplasts. Additionally, the identification of proteins with domains related to short-chain dehydrogenases reflects a role in the possible reduction of various metabolites, including carotenoid intermediates, which is essential for maintaining the balance of carotenoid biosynthesis. These enzymes are localized to the cytoplasm, chloroplast, and endoplasmic reticulum, indicating that these cellular compartments are involved in the synthesis and modification of carotenoids [41–43].

Moreover, a large number of identified proteins are hypothetical, originating from diverse plant species such as *Paulownia fortunei*, *Bixa Orellana* and *Handroanthus impetiginosus*, *Paulownia fortunei* are associated to KEGG pathways related to carotenoid biosynthesis (ko00906) and biosynthesis of secondary metabolites (ko01110). Proteins like lycopene beta-cyclase 2 from *Bixa orellana* and phytoene desaturase 1 from *R. glutinosa* are potentially involved in the carotenoid biosynthesis pathway, supporting their significance in the production of carotenoid pigments. Some identified proteins such as indole-3-acetaldehyde oxidase and abscisic-aldehyde oxidase are predicted to have roles in hormone-related pathways, potentially linking carotenoid derivatives to plant growth and stress responses [44–47].

Integrated transcriptomics and proteomics analysis revealed key molecular determinants of carotenoid variation in REG-CMC cell lines

Integrated transcriptomics and proteomics analysis revealed key molecular determinants of carotenoid variation in REG-CMC cell lines

Venn analysis of genes and proteins

To explore the relationship between transcriptomic and proteomic changes, an integrative analysis was conducted combining differentially expressed transcripts (DETs) and differentially expressed proteins (DEPs). The Venn diagram (Fig. 9A) illustrates the overlap among all detected transcripts (all_tran), all detected proteins (all_prot), differentially expressed transcripts (diff_tran) and differentially expressed proteins (diff_prot). A total of 291 entities were commonly differentially expressed at both the transcript and protein levels, while 492 were unique to the transcriptome and 414 were unique to the proteome. This overlap indicates a moderate correlation between mRNA and protein expression profiles, suggesting post-transcriptional regulation may contribute to the observed discrepancies.

To visualize the expression patterns of the overlapping differentially expressed entities, hierarchical clustering analysis was performed (Fig. 9B). The resulting heatmap revealed two distinct clusters corresponding to the experimental groups, indicating consistent trends in expression between transcriptomics and proteomics datasets.

GO enrichment analysis of integrated transcriptomic and proteomic data

GO enrichment analysis of the overlapping gene-protein set revealed significant enrichment in multiple biological processes, molecular functions and cellular components (Fig. 9C). In the biological process category, pathways related to "secondary metabolite biosynthetic process" (GO: 0044550) and "phenylpropanoid biosynthetic process" (GO: 0009699) were highly enriched, underscoring their potential roles in carotenoid biosynthesis and pigmentation differences between REG-CMC1 and REG-CMC2. Additional enriched processes included responses to hypoxia, programmed cell death, and various metabolic pathways. Among cellular components, "cell wall" (GO: 0005618), "plant-type cell wall" (GO: 0009505), and "anchored component of membrane" (GO: 0031225) showed strong enrichment, suggesting structural differences between cell lines. Molecular function enrichment revealed key regulatory activities such as "RNA polymerase II transcription factor activity" (GO: 0000981), "sulfotransferase activity" (GO: 0008146), and "ADP binding" (GO: 0043531), highlighting their possible involvement in gene regulation, metabolic modification and signaling.

Protein-specific GO enrichment further emphasized roles in microtubule-associated processes. Terms such as "microtubule-based movement" (GO:0007018), "microtubule binding" (GO:0008017), and "microtubule motor activity" (GO:0003777) were significantly enriched, suggesting a role in cellular transport and organization. Ion-binding functions such as "magnesium ion binding"

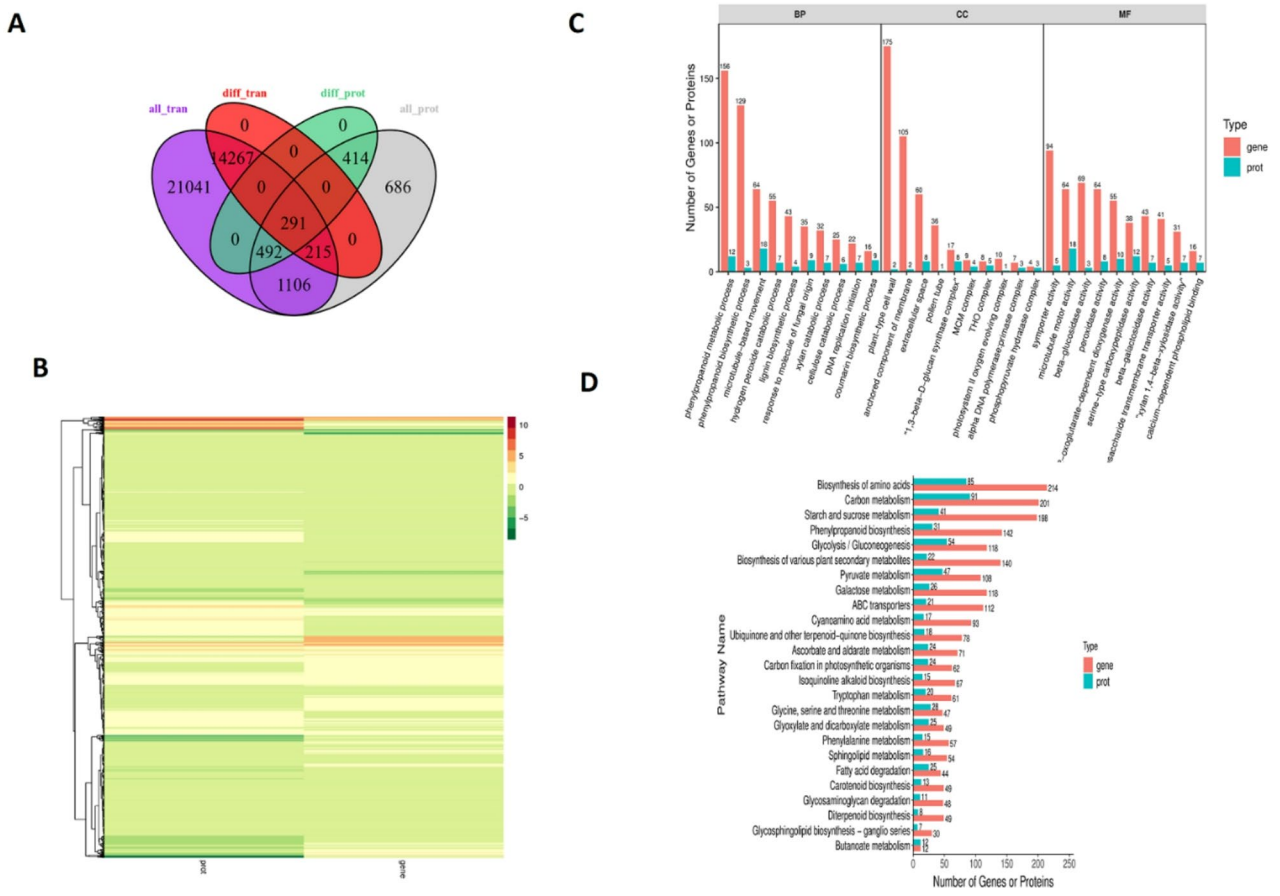

Fig. 9 Integrated analysis of transcriptomics and proteomics data to identify shared functional features between DEGs and DEPs. **A** Venn diagram of differentially expressed proteins and transcripts; **B** Protein and gene fold clustering heat map; **C** GO enrichment analysis of integrated genes and proteins; **D** KEGG pathway enrichment of differential proteins and genes

and "calcium ion binding" were also notably enriched. Additionally, protein enrichment pointed to significant involvement in carbohydrate metabolism ("carbohydrate metabolic process" (GO:0005975), proteolysis and energy production. Structural components including the "extracellular region" (GO:0005576), "chromoplast" (GO:0009509), and "pollen tube" (GO:0090406) were also enriched, indicating relevance to plant-specific functions such as photosynthesis and reproduction. The integrated GO enrichment analysis highlights the involvement of numerous biological processes and molecular functions, particularly those are associated with secondary metabolism, cytoskeletal dynamics, and ion-mediated signaling, which may contribute to the phenotypic and metabolic differences observed between the REG-CMC1 and REG-CMC2 cell lines.

Additionally, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis identified several key metabolic and biosynthetic pathways that were significantly enriched at both the transcript and protein levels (Fig. 9D). The most enriched pathways included "biosynthesis of amino acids," "carbon metabolism," "starch and sucrose metabolism," and "phenylpropanoid biosynthesis." Many of these pathways are associated with energy production and primary metabolism, reflecting their potential involvement in regulation between transcriptional and translational responses.

Distribution of Genes and Proteins in Metabolic Pathways Related to Carotenoid Biosynthesis in REG-CMC1 and REG-CMC2 Cell Lines

KEGG pathway enrichment analysis revealed significant differences in the distribution of genes and proteins across various metabolic pathways in REG-CMC1 and REG-CMC2 cell lines (Fig. 10A and B). Notably, pathways associated with second

Screening, functional annotation and enrichment analysis of DEGs

ary metabolism particularly phenylpropanoid biosynthesis (ko00940) and biosynthesis of various plant secondary metabolites (ko00999) were significantly enriched, suggesting their potential roles in the regulation of carotenoid and other pigment-related compounds. The

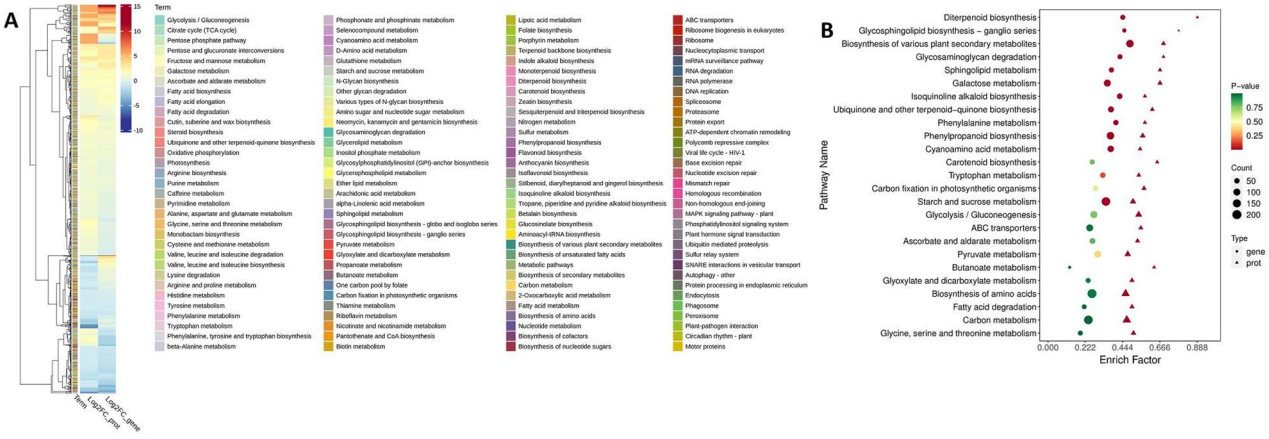

Fig. 10 Visualization of differential expression and pathway enrichment from integrated transcriptomic and proteomic analysis. **A** Heatmap showing differential expression of proteins and genes. Red indicates up regulation, while blue represents down regulation; **B** Dot plot of enriched KEGG pathways from integrated transcriptomics and proteomics analysis

phenylpropanoid biosynthesis pathway included 142 genes but only 31 corresponding proteins, indicating a potential post-transcriptional regulation or temporal control of protein expression in this biosynthetic route.

Similarly, the phenylpropanoid metabolic process contained 156 genes but only 12 proteins, while the phenylpropanoid biosynthetic process included 129 genes and just 3 proteins. This disparity between gene and protein representation suggests a complex regulation, possibly reflecting tissue-specific or developmentally regulated carotenoid biosynthesis. Other relevant pathways, such as diterpenoid biosynthesis (ko00904) and sphingolipid metabolism (ko00600), were also enriched, supporting the broader involvement of lipid-derived secondary metabolites in pigment production. In contrast, primary metabolic pathways like glycolysis/gluconeogenesis (ko00010) and carbon metabolism (ko01200) showed higher protein representation but lacked significant enrichment, indicating they may not directly involved in carotenoid-specific regulation.

Comprehensive proteomics analysis revealed key changes in protein expression across metabolic pathways, secondary metabolism and cellular regulation

Proteomic analysis revealed notable alterations in protein expression levels across key biological processes, with a particular focus on pathways involved in amino acid metabolism and secondary metabolite biosynthesis which may contribute carotenoid production. Several enzymes associated with phenylpropanoid biosynthesis, a pathway closely linked to carotenoid accumulation, were down regulated in REG-CMC2 compared to REG-CMC1. These include 4-coumarate: CoA ligase and phenylalanine ammonia-lyase, both of which play essential roles in the conversion of L-phenylalanine to precursors of secondary metabolites. Their reduced abundance suggests a possible suppression of phenylpropanoid-derived metabolic flux, potentially impacting carotenoid synthesis.

In addition, tyrosine decarboxylase, involved in aromatic amino acid metabolism (tyrosine, phenylalanine, tryptophan pathways), was also down regulated, indicating a broader downshift in organonitrogen compound biosynthesis. Similarly, decreased expression of an ATP-binding cassette (ABC) C transporter suggests reduced transmembrane transport of key metabolites, possibly affecting intracellular distribution of carotenoid precursors. Conversely, several proteins showed increased expression in REG-CMC2, including shikimate O-hydroxycinnamoyl transferase, an enzyme associated with phenylpropanoid and secondary metabolite biosynthesis, indicating a compensatory or cell-type-specific regulatory mechanism. Up regulation of a putative LOV domain-containing protein and R2R3-MYB transcription factor both involved in signaling and gene regulation suggests enhanced transcriptional control of pathways related to growth and metabolic differentiation. Furthermore, primary-amine oxidase, up regulated in REG-CMC2, may participate in the metabolism of phenylalanine and related pathways, further linking amino acid catabolism to secondary metabolite production.

Figure 11A illustrates the carotenoid biosynthesis pathway highlighting the differentially expressed genes identified through transcriptomics analysis of the REG-CMC1 and REG-CMC2 cell lines. Key up regulated nodes are concentrated around the early and middle stages of carotenoid biosynthesis, including enzymes involved in the conversion of phytoene to lycopene (such as *crtI*, *crtB*, *crtY*) and in the subsequent cyclization and hydroxylation reactions leading to β-carotene, zeaxanthin and lutein formation. Down regulated enzymes were primarily associated with the xanthophyll and abscisic acid

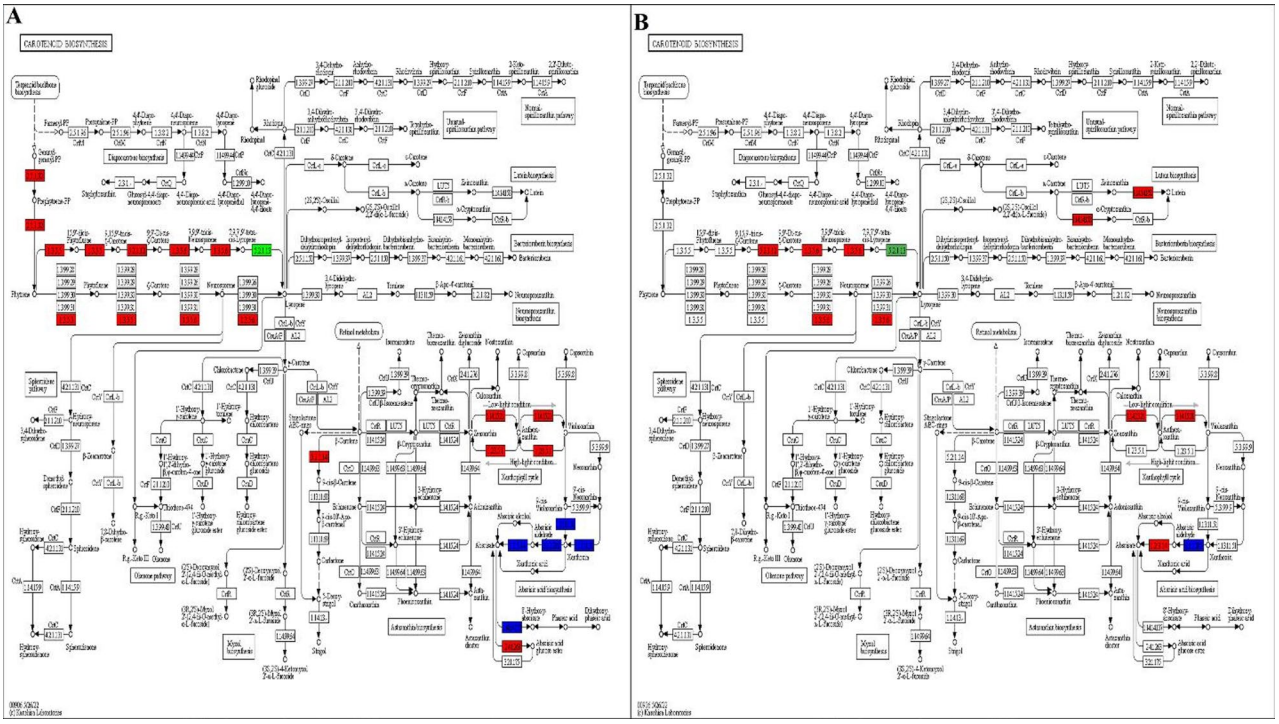

Fig. 11 Carotenoid biosynthesis pathway maps. **A** Pathway map from transcriptomics analysis. Genes or enzymes with up regulated expression in REG-CMC1 and/or REG-CMC2 are indicated in red, whereas those down regulated are shown in blue. The green highlighted enzymes represent genes that were uniquely expressed or significantly induced in only one of the two cell lines; **B** Pathway map from proteomics analysis. Proteins showing up regulation are highlighted in red, while down regulated proteins are indicated in blue. Green shading denotes proteins with significant expression unique to one of the two REG-CMC lines

(ABA) branches of the pathway (such as *ABA2*, *NCED*), suggesting reduced carotenoid catabolism and ABA synthesis in the analyzed cells.

Carotenoid biosynthesis pathway displays the enzymatic cascade converting precursor molecules from the terpenoid backbone pathway into carotenoids and their downstream derivatives (Fig. 11B). The proteomic profile reveals pronounced up regulation of enzymes involved in the early and intermediate stages of carotenoid biosynthesis, particularly those catalyzing the sequential desaturation and cyclization reactions from phytoene to lycopene and β-carotene (such as phytoene desaturase, ζ-carotene desaturase, lycopene β-cyclase). Additionally, proteins related to xanthophyll formation, such as β-carotene hydroxylase and zeaxanthin epoxidase, also exhibited elevated abundance, indicating an active conversion toward zeaxanthin and lutein production. Conversely, down regulated proteins were mainly localized within the abscisic acid (ABA) branch of the pathway, such as abscisic-aldehyde oxidase and 9-cis-epoxycarotenoid dioxygenase (NCED), consistent with reduced carotenoid catabolism and lower ABA biosynthetic activity in these cell lines.

qRT-PCR verification

All 6 selected DEGs including *Xanthoxin dehydrogenase (IMPGRGL1N62432)*, *Zeaxanthin epoxidase (IMPGRGL1N38937)*, *Zeta-carotene desaturase (novel.19984)*, *Phytoene dehydrogenase (IMPGRGL1N11711)*, *15-cis-zeta-carotene isomerase (IMPGRGL1N40609)* and *Carotenoid cleavage dioxygenase 4* (*IMPGRGL1N10527)* were successfully validated by qRT-PCR. Quantitative RT-PCR analysis was performed to evaluate the expression profiles of potential carotenoid pathway genes under two cell lines (CMC1 and CMC2). Mean relative expression values were calculated for six representative genes, and the correlation between the two conditions was assessed using Pearson's correlation coefficient. A moderate positive correlation ($r = 0.36$) was observed between CMC1 and CMC2 cell lines, suggesting partial co-regulation of carotenoid biosynthesis genes. The expression trends observed in REG-CMC-1 and REG-CMC-2 cell lines were consistent with the RNA-Seq data (Fig. 12A), supporting the accuracy and reliability of the transcriptomics analysis. Specifically, *Phytoene dehydrogenase* showed higher expression in CMC1, whereas *Carotenoid cleavage dioxygenase 4* was more strongly expressed in CMC2, implying potential differential regulation of synthesis and degradation pathways under the two cell lines.

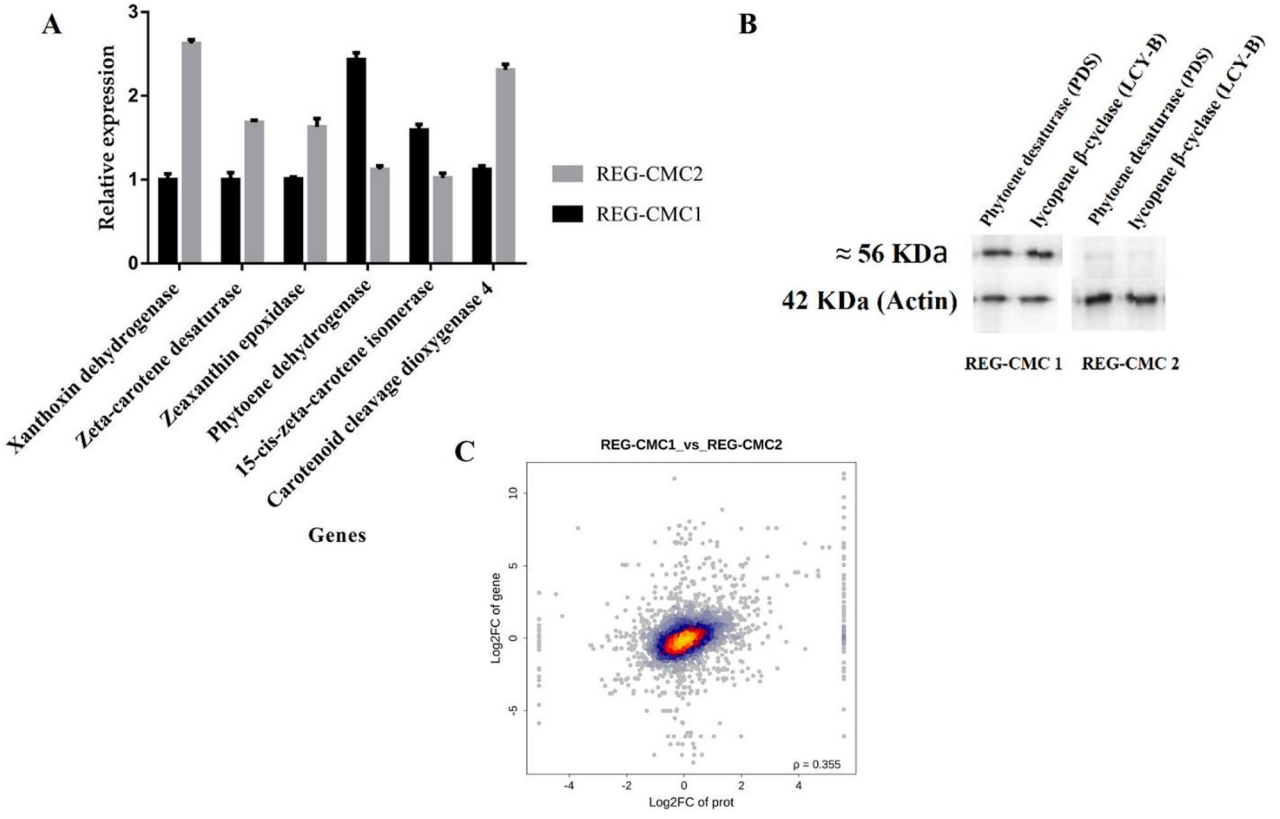

Fig. 12 Validation and correlation analysis of gene and protein expression between *R. glutinosa* cell lines REG-CMC-1 and REG-CMC-2. **A** Validation of RNA-Seq data through qRT-PCR analysis; **B** Protein expression validation through Western blotting; **C** scatter plot comparing changes in gene expression (transcriptomics) and protein expression (proteomics) between REG-CMC1" and "REG-CMC2 cell lines

Protein expression profiles of PDS and LCY-B in REG-CMC-1 and REG-CMC-2 cells

Western blot analysis demonstrated that both phytoene desaturase (PDS) and lycopene β-cyclase (LCY-B) proteins were abundantly expressed in REG-CMC-1 cells, whereas no or only minimal expression was detected in REG-CMC-2 cells (Fig. 12B). These findings are consistent with and corroborate the results obtained from the proteomic analysis.

Expression correlation analysis of genes and proteins in REG-CMC1 and REG-CMC2 cell lines

The Spearman correlation coefficient ($\rho = 0.355$) quantifies the relationship between gene and protein fold changes (Fig. 12C). This moderate positive correlation suggests that genes showing increased expression at the mRNA level generally tend to show increased protein expression as well, though the relationship is far from perfect. The correlation indicates an improved alignment between transcript and protein changes compared to individual sample analyses, but still reflects the influence of biological processes such as post-transcriptional regulation, protein turnover and translation efficiency.

Discussion

The biosynthesis of secondary metabolites in *R. glutinosa*, particularly carotenoids is important for root pigmentation. ABC transporters facilitate the export of these secondary metabolites from the cytoplasm to vacuoles, where they accumulate and contribute to root coloration [48, 49]. REG-CMC-1 cells produced significantly higher yellow pigments compared to REG-CMC-2 cells. The dramatic difference in carotenoid content suggests profound regulatory divergence at both transcriptional and proteomic levels. The production of yellow pigments in REG-CMC-1 cells may be regulated by carotenoid biosynthesis pathway as large number of identified genes and proteins are associated with carotenoid biosynthesis.

The current study provides evidence that integrated transcriptomic and proteomic analyses can offer insights into the molecular mechanisms associated with carotenoid biosynthesis in *Rehmannia glutinosa* CMCs. The identification of genes and proteins such as phytoene synthase (PSY), 15-cis-phytoene desaturase, zeta-carotene desaturase (ZDS), beta-carotene isomerase, and zeaxanthin epoxidase suggests the possibility of coordinated regulation across multiple steps of the carotenoid biosynthetic pathway, from precursor formation to downstream

modifications. These findings aligned with known roles of these enzymes in model plants, where they catalyzed essential steps in carotenoid formation and conversion, contributing to the diversity of carotenoid-derived compounds such as lutein, zeaxanthin, and abscisic acid [50–52]. Xanthoxin dehydrogenase and abscisic-aldehyde oxidase encoding genes which are involved in carotenoid catabolism and related pathways were also identified in this research. These results suggest a tightly regulated network which may be responsible for carotenoid accumulation and may control its conversion to bioactive metabolites such as abscisic acid. The detection of enzymes including geranylgeranyl diphosphate synthase (GGPPS) and tocopherol cyclase indicates a potential association between carotenoid biosynthesis and pathways involved in isoprenoid and secondary metabolite metabolism, confirming earlier reports on the diversity of terpenoid products in CMCs of *R. glutinosa* [53].

Comparison with previous studies reinforces these findings. Zhou et al. [22] reported elevated levels of iridoid glycosides (IAs) and terpenoids in *R. glutinosa* CMCs, along with identification of genes crucial for catalpol and IA biosynthesis. In this study, the REG-CMC-1 and REG-CMC-2 cell lines exhibited distinct transcriptional and proteomic profiles. Compared with REG-CMC-2, REG-CMC-1 showed higher expression of carotenoid-related genes and greater carotenoid content, which may reflect differences in metabolic characteristics between CMC subpopulations. Enrichment of secondary metabolic pathways, including carotenoid and phenylpropanoid biosynthesis, as well as galactose metabolism, indicates a degree of biochemical diversity between the two cell lines. KEGG and GO enrichment analyses further showed that transcript levels were generally higher than the corresponding protein abundance, particularly in the carotenoid biosynthesis pathway. This observation suggests that post-transcriptional or translational regulation may contribute to differences in metabolite accumulation, a finding consistent with previous studies that reported fluctuations in gene expression and metabolite levels during developmental transitions in other plants [54].

Furthermore, the subcellular localization and enzyme complex formation observed in other species such as PSY, PDS, and ZDS associating with plastidic thylakoid membranes [52] and GGPPS accumulating in plastidic globules [55] support the importance of spatial organization in carotenoid biosynthesis. These insights are supported by studies in grapevine, where expression of *VvDXS* and *VvDXR* genes, key upstream regulators in the methylerythritol phosphate (MEP) pathway, was positively correlated with carotenoid accumulation during fruit ripening [56–58]. Similar transcriptional patterns may be observed in *R. glutinosa*, which could be related to carotenoid biosynthesis in CMCs, with possible involvement of the MEP pathway in terpenoid and carotenoid metabolism.

Proteomics analysis of *R. glutinosa* CMC lines revealed the presence of proteins related to carotenoid biosynthesis. Among these, phytoene desaturase 1 and lycopene β-cyclase were observed at the protein level, indicating a possible role in early and intermediate stages of carotenoid biosynthesis. These enzymes catalyze the desaturation of phytoene and the cyclization of lycopene into beta-carotene, respectively critical reactions that ultimately lead to the production of carotenoids and xanthophylls essential for pigmentation and photosynthesis [52]. In addition, the detection of prolycopene isomerase and indole-3-acetaldehyde oxidase suggests a potentially interconnected protein network in which redox processes and energy metabolism may be associated with carotenoid accumulation. Although several proteins were annotated as hypothetical, their co-expression with known carotenoid biosynthetic enzymes indicates that further functional studies may help clarify their roles within this pathway. This complexity reflects previous findings that in plants, many carotenoid biosynthetic enzymes form multi-enzyme complexes in plastidic compartments such as the stroma and thylakoid membranes, where localization and post-translational modifications play crucial roles in functional activity [51, 55].

In addition to carotenoids, the results of this study indicate the potential involvement of the phenylpropanoid pathway in *R. glutinosa* root pigmentation. Proteins and gene expression related to phenylalanine ammonia-lyase (PAL), cinnamate-4-hydroxylase (C4H), and 4-coumarate-CoA ligase (4CL) were enriched, consistent with previous research highlighting their roles in the biosynthesis of flavonoids and anthocyanins secondary metabolites known for their pigmentation and antioxidative functions [59, 60]. The presence and regulation of MYB transcription factors, particularly R2R3-MYBs, further support their involvement in the activation of anthocyanin biosynthetic genes, aligning with the regulatory mechanisms described in other plant species [61, 62].

The co-regulation of both carotenoid and anthocyanin biosynthesis pathways appears to be coordinated, at least in part, by the MBW transcriptional complex (MYB–bHLH–WD40), which has been implicated in modulating a wide range of pigmentation-related processes [63]. In *R. glutinosa*, this shared regulatory control likely underpins the complex coloration of roots, where both carotenoids and phenylpropanoids contribute to the visual phenotype and offer protective benefits through their antioxidant properties [64]. Furthermore, cytoskeletal dynamics and primary metabolic processes such as starch and sucrose metabolism may indirectly influence secondary metabolite biosynthesis. Microtubule

organization, for instance, has been shown to modulate the expression of anthocyanin biosynthetic genes, possibly by affecting intracellular trafficking or transcriptional regulation [63]. Similarly, the breakdown of sucrose into glucose and fructose supplies essential carbon skeletons for flavonoid and anthocyanin biosynthesis, as well as general energy demands of highly active metabolic cells [65, 66]. The regulatory networks proposed in this study are derived from correlation-based omics analyses combined with previously reported functional evidences. Collectively, the results indicate a possible coordination between carotenoid and phenylpropanoid pathways in *R. glutinosa* CMCs at transcriptional and proteomic levels. Several key carotenoid-associated genes and regulatory factors showed concordant expression trends at both the mRNA and protein levels, suggesting that these candidates may play important roles in the divergence of carotenoid accumulation between REG-CMC1 and REG-CMC2.The involvement of regulatory factors such as MBW transcription factors, together with metabolic flux from primary carbohydrate metabolism, may contribute to the regulation of pigment biosynthesis in root tissues. These findings contribute to a better understanding of root pigmentation and secondary metabolite accumulation in *R. glutinosa*, and may support future efforts to optimize the production of bioactive compounds.

Materials and methods

Plant materials

R. glutinosa plant was identified by Dr. Pengfei Zhou of School of Basic Medical Sciences, Guangdong Medical University, China. About 1–3-month-old fresh roots of *R. glutinosa* plants were collected in Jiaozuo, Henan Province, China. *Rehmannia glutinosa* roots those growing naturally in the wild were collected for cambium tissue culture. No special permission was needed for research involving this species. Plant specimens were stored in the gene bank herbarium at Guangdong Medical University, China, with the reference number of GDD230XT99.

Cells lines and culture conditions

The two cell lines, REG-CMC-1 and REG-CMC-2, were derived from explants of two different *R. glutinosa* roots. According to the iPlant Database (https://www.iplant.cn/info/Rehmannia) maintained by the Institute of Botany, Chinese Academy of Sciences, the genus *Rehmannia* comprises six species: *R. glutinosa*, *R. henryi*, *R. chrysantha*, *R. piasezkii*, *R. solanifolia*, and *R. chingii*. Based on morphological assessment and comparison with diagnostic characteristics described in this database, the plant specimens collected for this study were taxonomically confirmed as *Rehmannia glutinosa*. Two wild *R. glutinosa* individuals were collected from Jiaozuo, Henan Province. These plants served as the sources for establishing two independent cell lines. One specimen exhibited a distinctly yellow root cross-section, from which the REG-CMC-1 cell line was induced. The second specimen displayed an almost white root cross-section with a slightly yellowish epidermis, from which the REG-CMC-2 cell line was derived. The fresh rhizomes of *R. glutinosa* were harvested in Jiaozuo, Henan Province, China. The surface soil of *R. glutinosa* roots washed and then rinsed with tap water for 3 h. The cleaned explants were placed on a clean bench, and filter paper was used to absorb the surface water, followed by sterilization with 3% hydrogen peroxide (2 min), rinsing with sterile distilled water (3 times), sterilization with 75% medical ethanol (2 min), and 3 times rinsing using sterile distilled water. Then, 15 min sterilization with mercuric chloride solution (0.1%, w/v), followed by rinsing with sterile distilled water (5 times), and finally soaking the sterilized explants in 150 mg/l citric acid solution to prevent browning.

The methodology used to establish REG-CMC-2 in the present study was adapted directly from the procedures described for REG-CMC-1 [22]. Sterilized *R. glutinosa* explants were cultured in 250 mL conical flasks containing 50 mL of basal Murashige and Skoog (MS) medium supplemented with 30 g L^{-1} sucrose, 2 mg L^{-1} α-naphthalene acetic acid (NAA), 2 mg L^{-1} 6-benzylaminopurine (6-BA), and 1 mg L^{-1} 2,4-dichlorophenoxyacetic acid (2,4-D). Within approximately 7–12 days of culture, pale-yellow (REG-CMC-1) or white (REG-CMC-2) cell masses developed around the root cambium region. After 14 days, the newly formed calli were transferred to fresh subculture medium (100 mL of MS medium supplemented with 2 mg L^{-1} naphthalene acetic acid (NAA), 1 mg L^{-1} 6-benzylaminopurine (6-BA), and 30 g L^{-1} sucrose; pH: 6.0) for further growth. The initial callus tissue derived from the roots appeared as firm, light-yellow aggregates. Both REG-CMC-1 and REG-CMC-2 cells subsequently developed into soft, friable layers of cells exhibiting yellow-green or whitish coloration.

Following induction, the initially generated cells were designated as the first generation. Both REG-CMC-1 and REG-CMC-2 were maintained on MS solid medium supplemented with 2 mg/L NAA, 1 mg/L 6-BA, 30 g/L sucrose, and 4 g/L gellan gum (pH 6.0), and sub cultured at 12-day intervals at 25 °C in darkness. Microscopic observations revealed that CMC cultures consisted predominantly of single cells or small clusters, whereas callus cultures formed dense, compact clusters with many cells, consistent with our previous observations of REG-CMC-1. For the present study, REG-CMC-1 at the 35th subculture and REG-CMC-2 at the 30th subculture were employed. To establish a suspension culture system, 6.0 g of REG-CMC-1 and REG-CMC-2 were transferred into

250 mL Erlenmeyer flasks containing 100 mL of MS medium supplemented with 2 mg/L NAA, 1 mg/L 6-BA, and 30 g/L sucrose (pH was adjusted to 6.0). The flasks were placed on a shaker with 120 rpm at 25 °C under continuous dark conditions.

Metabolites extraction, refinement and carotenoid determination

For the carotenoid extraction, 10.0 g of dried cell mass was placed into a 500 mL beaker. A total of 250 mL of anhydrous ethanol was added, and the mixture was sealed and soaked in the dark for 10 h. The extraction process was further enhanced using an ultrasonic cleaning machine with 20 kHz frequency and 10 s on / 10 s off pulse for 30 min at 30 °C. This extraction was repeated twice, with the filtrate from each extraction being collected and filtered. After completing all three extractions, the filtrates were combined. The combined solution was then processed using a rotary evaporator to remove the ethanol, yielding a crude extract of carotenoids including other metabolites.

The crude extract was weighed and W_1 represents the measured weight of the crude extract obtained after the extraction process. This value indicates the total amount of material extracted from the sample, which contains the carotenoids along with other possible compounds. Because the initial sample weight (10 g) and extraction conditions were fixed for all samples, crude extract weight was normalized by dividing by 10 to express carotenoid content (mg) on a per-gram sample basis, following the normalization principle used in standard carotenoid quantification methods [67].

The carotenoid content of REG-CMC-1 and REG-CMC-2 was determined using a commercial Carotenoid Assay Kit (Shanghai Yuanye Bio-Technology Co., Ltd.). Fresh cells (0.1 g; dry weight 0.05 g) were homogenized with 50 mg of extraction powder and 1 mL of Carotenoid Assay Buffer. The homogenate was transferred into a 10 mL centrifuge tube, and the mortar or homogenizer was rinsed several times with a small volume of Carotenoid Assay Buffer. The washings, together with the residues, were combined in the same centrifuge tube. The volume was then adjusted to 10 mL with Carotenoid Assay Buffer, mixed thoroughly, and incubated in the dark for 15 min. Subsequently, the mixture was centrifuged at 4000 rpm for 5 min, and the supernatant was collected for analysis. Prior to measurement, the spectrophotometer was preheated for at least 30 min and set to a wavelength of 440 nm. The carotenoid extract was placed into a 1 cm path-length cuvette, with Carotenoid Assay Buffer used as the blank. The absorbance at 440 nm ($A_{44}0$) was recorded.

Carotenoid content was calculated using the following equation:

Carotenoid content (mg/g) = A_{440}/ (250 × 1)×V×1000/W = 0.04×A440/W [68].

Where, V = volume of carotenoid extract (0.01 L); W = sample fresh weight or dry weight (g), 250 = empirical extinction coefficient of carotenoids (L/g/cm); 1000 = unit conversion factor (g to mg); 1 = path length of the cuvette (cm). This extraction was repeated twice, with the filtrate from each extraction being collected and filtered. After completing all three extractions, the filtrates were combined.

RNA extraction, cDNA library construction and sequencing

Total RNA was extracted from REG-CMC-1 and REG-CMC-2 cell lines with the RNA Easy Fast Plant Tissue Kit (TIANGEN Biotech Co., Ltd., Beijing, China) by the manufacturer's instructions. Three independent cell cultures were maintained and total RNA was extracted from each, resulting in three biological replicates. Messenger RNA (mRNA) was isolated using two standard approaches such as enrichment of polyadenylated transcripts via Oligo (dT) magnetic beads and depletion of ribosomal RNA from total RNA to retain non-rRNA species. Agarose gel electrophoresis was used to assess the quality and integrity of RNA and a NanoDrop ™ 8000 spectrophotometer (Thermo Scientific, USA) was used to analyze the purity of RNA. High-quality RNA of each sample was then used to create cDNA libraries in accordance with the Illumina platform procedure (Illumina, USA) and the cDNA libraries were sequenced using the Illumina HiSeq™ 2000 platform.

Screening, functional annotation and enrichment analysis of DEGs

Raw sequencing reads were subjected to stringent quality control using FastP [69] to ensure high data integrity for downstream analyses. Filtering criteria included the removal of reads containing adapter sequences and exclusion of reads in which the proportion of undetermined bases (N) exceeded 10%; and as well as elimination of reads where over 50% of bases had a quality score ≤ Q20. High-quality clean reads were then aligned to the reference genome to generate mapped data. Differential gene expression analysis between REG-CMC-1 and REG-CMC-2 cell lines were performed using DESeq2 [70]. Benjamini-Hochberg method was employed to modify the p-values for multiple hypothesis testing, resulting in the false discovery rate (FDR). Genes were categorized as differentially expressed if the absolute value of the $\log_2$ fold change ($\log_2$ FC) was ≥ 1 and the false discovery rate (FDR) was < 0.05. KEGG [71] and GO [72] were utilized for the functional annotation, classification, and pathway enrichment analysis of DEGs.

Table 3 List of primers used in this study for qRT-PCR validation

SL. No.	Primer ID	Primer Sequence (5′-3′)
1.	Xde-Fw	TCTAGCGATTCGCCTCTTGT
2.	Xde-Rv	TTGCACCACCTGTTACCAAA
3.	Zds-Fw	TCGTCACGAGGCCTAGAAGT
4.	Zds-Rv	TTTGGAGGCTGTTCTTGCTT
5.	ZEP-Fw	GCAATGGTATGCATTTCACG
6.	ZEP-Rv	CAGAATGGCTTCCTCCTCAG
7.	PDS-Fw	AAGGCGCTGTCTTATCAGGA
8.	PDS-Rv	CTCCCGCCAAATTAATGAGA
9.	Z-Iso-Fw	TTCCATCGCTAAATCCCTTG
10.	Z-Iso-Rv	GATACGGCGGATTAAGGTGA
11.	CCD4-Fw	TACACTCTGCAGCCGATACG
12.	CCD4-Rv	AAGCCAAGCTTGTGTTTGCT

Validation of RNA-Seq data by quantitative real-time PCR (qRT-PCR)

To validate the RNA-Seq results and examine the expression patterns of genes involved in carotenoid biosynthesis, 6 DEGs including *Xanthoxin dehydrogenase (IMPGRGL1N62432)*, *Zeaxanthin epoxidase (IMPGRGL1N38937)*, *Zeta-carotene desaturase (novel.19984)*, *Phytoene dehydrogenase (IMPGRGL1N11711)*, *15-cis-zeta-carotene isomerase (IMPGRGL1N40609)* and *Carotenoid cleavage dioxygenase 4* (*IMPGRGL1N10527)* were randomly selected for qRT-PCR analysis. For each gene, three independent biological replicates were used, with each biological replicate further tested in three technical replicates.

Total RNA was extracted from the cell lines using the RNAprep Pure Plant Plus Kit (Tiangen, Beijing, China), following the manufacturer's protocol. Complementary DNA (cDNA) synthesis was carried out using the TransScript® All-in-One First-Strand cDNA Synthesis SuperMix for qPCR (TransGen Biotech, China), according to the manufacturer's instructions. Gene-specific primers were designed using PRIMER5 software (version 5.2) and are detailed in Table 3. The housekeeping gene β-actin was used as an internal reference for normalization of gene expression levels. qRT-PCR assays were performed using the ABI PRISM® 9600 Sequence Detection System (Applied Biosystems, USA) with SYBR Green Supermix, following the manufacturer's instructions. The thermal cycling program included an initial denaturation at 94 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 30 s (annealing), and a final extension at 72 °C for 10 min. Relative gene expression levels were calculated using the $2^{-\Delta\Delta CT}$ method [73].

Extraction, purification and identifcation of protein

Protein extraction

Three biological replicates were prepared by culturing cells independently and processing them separately through protein extraction and further processes. The sample was grounded into a powder in liquid ammonia. An appropriate amount of the powder was transferred to a 1.5 ml centrifuge tube and an appropriate amount of L3 lysis buffer (1% SDS, 100 mM Tris-HCl, 7 M urea, 2 M thiourea, 1 mM PMSF, and 2 mM EDTA) was added. After oscillation and mixing, the sample was ultrasonically lysed on ice for 10 min and the supernatant was centrifuged to obtain a protein solution. After that, 4 volumes of chilled acetone were added to the protein solution, and the solution was precipitated at -20 °C overnight. The precipitate was centrifuged at 4 °C and retained. The precipitate was washed with cold acetone and redissolved in 8 M urea. The total protein concentration was determined by BCA protein quantification assay.

Protease desalting

Equal amounts of protein solution were taken based on protein concentration and brought to a total volume of 200 µl with 8 M urea. The solution was then treated with 10 mM DTT at 37 °C for 45 min and alkylated with 50 mM iodoacetamide (AM) for 15 min in the dark at room temperature. Later protein solution (4 times) was added with pre-chilled acetone, and the protein was precipitated at -20 °C for 2 h. After centrifugation, the protein pellet was precipitated and resuspended in 200 µl of 25 mM ammonium bicarbonate solution and 3 µl of trypsin (Promega) and digested overnight at 37 °C. Following digestion, peptides from each sample were desalted on a C18 column, concentrated by vacuum centrifugation, and re-dissolved in 0.1% (v/v) formic acid.

Liquid chromatography with tandem mass spectrometry (LC-MS/MS)

The samples were analysed using a NanoElute UHPLC apparatus at nano-flow parameters. Mobile phase A consisted of 0.1% formic acid in water, whereas mobile phase B included 0.1% formic acid in pure acetonitrile. Samples were introduced through an autosampler onto a C18 analytical column (25 cm × 75 µm, 1.6 µm particle size; Ion Opticks, Australia). To maintain optimal separation efficiency, the column temperature was regulated at 50 °C using an integrated oven. Each run involved a sample load of 200 ng, with a flow rate set at 300 nL/min. During gradient elution, the proportion of mobile phase A gradually decreased while mobile phase B increased, enabling peptide separation based on hydrophobicity. The gradient began with 98% mobile phase A. From 0 to 25 min, mobile phase A decreased linearly to 78%. Between 25 and 30 min, it was further reduced to 65%. From 30 to 35 min, a sharper decrease brought it down to 20%. Finally, a hold at 20% mobile phase A from 35 to 40 min ensured complete elution of strongly retained peptides. A 40-minute gradient elution began with a linear increase of mobile phase B from 2% to 22% over the

first 25 min. This was followed by a continued gradient from 22% to 35% between 25 and 30 min. Subsequently, the concentration of mobile phase B was further elevated from 35% to 80% over the next 5 min (30–35 min). The gradient concluded with mobile phase B maintained at a constant 80% for the final 5 min (35–40 min) to complete the separation process.

Mass spectrometry analysis and data acquisition

Mass spectrometry analysis yielded the mass-to-charge ratio (m/z) and signal intensity of peptides, along with the m/z and signal intensity of the fragment ions generated post-peptide fragmentation. Following chromatographic separation, the mixed peptide sample was initially analyzed using a timsTOF Pro2 mass spectrometer operating in data-dependent acquisition parallel accumulation serial fragmentation (ddaPASEF) mode. This preliminary step was conducted to optimize the acquisition windows for subsequent data-independent acquisition (diaPASEF) analysis. The effective chromatographic gradient used during MS acquisition was 40 min, with detection performed in positive ion mode. The precursor ion mass range was set from 100 to 1700 m/z, and ion mobility values ranged from 0.85 to 1.3 V·s/cm². Ions were accumulated and released within 100 ms, achieving an ion utilization efficiency close to 100%. Instrument settings included a capillary voltage of 1500 V, a drying gas flow rate of 3 L/min, and a drying temperature of 180 °C.

In ddaPASEF mode, the system acquired four MS/MS scans per cycle, with a total cycle time of 0.53 s. The scans were performed across a charge state range of 0–5, with a dynamic exclusion time of 0.4 min to prevent redundant fragmentation. Additional parameters included an ion density threshold of 10,000, a minimum intensity threshold of 1500, and collision energy ramping based on ion mobility starting at 27 eV for 1/K = 0.85 V·s/cm² and increasing to 45 eV for 1/K = 1.3 V·s/cm². The quadrupole isolation width was set at 2 Å for m/z values below 700, and 3 Å for m/z above 800.

For the diaPASEF acquisition, data were collected using a mass range of 400–1200 m/z and a mobility range of 0.85–1.3 V·s/cm². The method employed 24 mass steps per cycle, each with a mass window of 25 Da and an overlap of 0.1 Da, across two ion mobility windows, resulting in a total of 48 acquisition windows. The average acquisition cycle time for diaPASEF was approximately 1.17 s.

To identify proteins, spectral library searching was employed in addition to conventional sequence database searches. In this approach, experimental tandem mass spectra are matched against spectra in a curated spectral library, enabling peptide sequence identification via similarity scoring from the NIST Mass Spectrometry Data Center (proteomicsresource.washington.edu). A false discovery rate (FDR) of 1% was implemented for both precursor ion and protein levels.

Protein differential expression analysis

Proteins with a fold change (FC) ≥ 1.5 and a P-value < 0.05 were considered significantly differentially expressed [74]. These proteins were subsequently normalized using z-scores, and hierarchical clustering was performed to generate heat maps for visualization.

Functional classification of proteins

The detected proteins linked to different biological process categories were classified according to the GO (http://www.geneontology.org) and UniProtKB (http://www.uniprot.org) databases. Hierarchical clustering of protein expression patterns was performed using Cluster software version 3.0, utilizing $\log_2$-transformed spot abundance ratios to compare REG-CMC-1 and REG-CMC-2 cell lines. The KOG database was categorized into 26 functional classifications. The identified proteins were analyzed against the comprehensive KEGG species database to derive KEGG annotation results. The number of differentially expressed proteins associated with each KEGG pathway was quantified, and the results were visualized using a bar graph.

Validation of protein expression through Western blotting

Total protein was extracted from REG-CMC-1 and REG-CMC-2 cells, and protein concentrations were determined using a BCA assay kit (Vazyme, China). Equal amounts of protein were resolved by SDS–PAGE and subsequently transferred onto polyvinylidene difluoride (PVDF) membranes using an electrotransfer system (Bio-Rad, USA). The membranes were blocked with 5.0% (w/v) nonfat milk at room temperature for 1 h and then incubated overnight at 4 °C with primary antibodies against phytoene desaturase (PDS) and lycopene β-cyclase (LCY-B) (Phyto AB, USA). After incubation with horseradish peroxidase (HRP)-conjugated secondary antibodies, immunoreactive signals were detected using Clarity™ Western ECL substrate (Bio-Rad, USA). β-actin served as an internal loading control and was detected using a β-actin antibody (Phyto AB, USA), followed by incubation with a peroxidase-conjugated secondary antibody. Band intensities were quantified using ImageJ software.

Conclusions

R. glutinosa continues to emerge as a valuable resource in both traditional and modern medicine due to its diverse bioactive compounds, particularly its carotenoids, glycosides and phenolic compounds. The exploration of its chemical profile has opened new avenues for understanding its therapeutic effects, including its ability to manage chronic conditions such as diabetes and

hypertension. The innovative application of plant cell culture systems, especially in the context of producing secondary metabolites like carotenoids, offers a promising strategy for sustainable and environmentally friendly production of bioactive compounds, reducing dependency on traditional agricultural practices and enabling controlled, high-yield production under optimized conditions. Two cell lines, REG-CMC-1 and REG-CMC-2, were successfully induced and established from the cambial ring region of *R. glutinosa* roots. The induction of REG-CMC-1 was initiated from root slices in which the cambial area exhibited a distinct yellow pigmentation, which was subsequently reflected in the yellow coloration of the derived cell line. In contrast, REG-CMC-2 originated from root slices where the cambial region appeared distinctly white, resulting in a white-colored cell line. That identified genes and enzymes including *phytoene synthase*, *zeta-carotene desaturase* and *lycopene β-cyclase* may contribute to the increased production of yellow pigments observed in REG-CMC-1 cells. The consistency between omics data and independent experimental validation confirms the reliability of the identified differentially expressed genes and proteins. This analysis reveals distinct expression patterns related to carotenoid biosynthesis and associated regulatory pathways, underscoring fundamental molecular differences between REG-CMC1 and REG-CMC2 cell lines. Functional validation experiments are beyond the scope of the present study and we proposed a putative regulatory network that provides a hypothesis-generating framework for carotenoid differentiation.

Abbreviations

CMC	Cambial meristematic cell
DEGs	Differentially expressed genes
GGPP	Geranylgeranyl pyrophosphate
MEP	Methylerythritol phosphate
PSY	Phytoene synthase
PCA	Principal Component Analysis
TF	Transcription factor
ERFs	Ethylene-responsive transcription factors
DREBs	Dehydration-responsive element-binding proteins
bHLH	Basic helix-loop-helix
bZIP	Basic leucine zipper
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
DEPs	Differentially expressed proteins
KOG	EuKaryotic Orthologous Groups
MAPK	Mitogen-activated protein kinase
ZDS	Zeta-carotene desaturase
IAs	Iridoid glycosides
PAL	Phenylalanine ammonia-lyase
C4H	Cinnamate-4-hydroxylase
4CL	4-coumarate-CoA ligase
LC-MS/MS	Liquid chromatography with tandem mass spectrometry
FDR	False discovery rate
FC	Fold change

Acknowledgements
This research was supported by the Guang Dong Basic and Applied Basic Research Foundation, the Funds for PhD Researchers of Guangdong Medical University and the Science and Technology Development Special Fund Competitive Allocation Project of Zhanjiang City. Additional support was provided by the Shenzhen Science and Technology Plan Project and the Student Innovation Research and Entrepreneurship Training (SRT) programs. The authors gratefully acknowledge the financial support that made this work possible.

Authors' contributions
**Pengfei Zhou: ** Conceptualization, Data curation, Methodology, Formal analysis, Funding acquisition, Software, Validation, Writing – original draft; **Haihua Li: ** Investigation, Methodology, Writing – original draft; **Jiahe Li: ** Investigation, Writing – original draft; **Haoke Zhang: ** Investigation; draft; **Xiaonan Deng: ** Investigation, Writing – original draft; **Zhiye Huang: ** Data curation, Methodology, Formal analysis; **Yunshuang Yuan: ** Investigation, Writing – original draft; **Zemei Lai: ** Methodology, Investigation; **Xiaoyan Zhang: ** Data curation, Methodology; **Tong Wang: ** Methodology, Investigation, Writing – original draft, **Yafei Huang: ** Methodology; **Wener Guo: ** Methodology, Writing – review & editing; **Zhuchao Zhong: ** Investigation, Writing – original draft; **Chun Yang: ** Project administration, Software, Supervision, Writing – review & editing.

Funding
This research was funded by Guang Dong Basic and Applied Basic Research Foundation (NO. 2021A1515012583 and 2025A1515012615), Funds for PHD researchers of Guangdong Medical University in 2019 (NO.B2019002), Science and Technology Development Special Fund Competitive Allocation Project of Zhanjiang City (NO. 2020A01038), Shenzhen Science and Technology Plan Project (NO. JCYJ20210324132211028), Student Innovation Research and Entrepreneurship Training (SRT) (NO.S202310571099, ZZDM020, GDMU2022237, GDMU2022235, ZYDM012, ZYDM013, S202410571049), The Medical Scientific Research Foundation of Guangdong Province (No. A2024398).

Data availability
The datasets generated during the current study are available in the NCBI Sequence Read Archive (SRA) repository, under accession number PRJNA1189061 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1189061).

Declarations

Ethics approval and consent to participate
This study does not involve human participants, clinical trials or patient data.

Consent for publication
This study involved plant tissue culture techniques to produce phytochemicals and did not involve human or animal subjects, and therefore ethical approval and participant consent were not required. All applicable institutional and national guidelines for the care and use of plant materials were followed.

Competing interests
The authors declare no competing interests.

Received: 20 June 2025 / Accepted: 3 February 2026
Published online: 13 February 2026

References

1. Group AP, Chase MW, Christenhusz MJ, Fay MF, Byng J, Judd W, Soltis D, Mabberley D, Sennikov A, Soltis P. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016;181:1–20.
2. Zhang R-X, Li M-X, Jia Z-P. *Rehmannia glutinosa*: review of botany, chemistry and Pharmacology. J Ethnopharmacol. 2008;117:199–214.
3. Bian Z, Zhang R, Zhang X, Zhang J, Xu L, Zhu L, Liu Y. Extraction, structure and bioactivities of polysaccharides from *Rehmannia glutinosa*: A review. J Ethnopharmacol. 2023;305:116132.
4. WangY, Liao D, Qin M, Li X. Simultaneous determination of catalpol, aucubin, and Geniposidic acid in different developmental stages of *Rehmannia*

glutinosa leaves by high performance liquid chromatography. J Anal Methods Chem. 2016;4956589:1–6. https://doi.org/10.1155/2016/4956589.
5. Li X, Zhou M, Shen P, Zhang J, Chu C, Ge Z, Yan J. Chemical constituents from *Rehmannia glutinosa*. China J Chin Mater Med. 2011;36(22):3125–9.
6. Shah BN, Patel PB, Patel AB, Nayak BS, Modi DC. *REHMANNIA GLUTINOSA*–A PHYTO-PHARMACOLOGICAL. Pharmacologyonline. 2010;1:737–53.
7. Liu YF, Shi GR, Wang X, Zhang CL, Wang Y, Chen RY, Yu DQ. Bioactive iridoid glycosides from the whole plants of *Rehmannia Chingii*. J Nat Prod. 2016;79(2):428–33.
8. Feng WS, Li M, Zheng XK, Zhang N, Song K, Wang JC, Kuang HX. Two new ionone glycosides from the roots of *Rehmannia glutinosa* libosch. Nat Prod Res. 2015;29(1):59–63.
9. Fu GM, Shi SP, Ip FC, Pang HH, Ip. N.Y. A new carotenoid glycoside from *Rehmannia glutinosa*. Nat Prod Res. 2011;25(13):1213–8.
10. Oksman-Caldentey K-M, Inzé D. Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. Trends Plant Sci. 2004;9:433–40.
11. Fiehn O. Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol. 2002;48:155–71.
12. Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod. 2012;75:311–35.
13. Ambati RR, Phang SM, Ravi S, Aswathanarayana RG. Astaxanthin: sources, extraction, stability, biological activities and its commercial applications – a review. Mar Drugs. 2014;12:128–52.
14. Howitt CA, Pogson BJ. Carotenoid accumulation and function in seeds and non-green tissues. Plant Cell Environ. 2006;29:435–45. https://doi.org/10.1111/j.1365-3040.2005.01492.x).
15. Sugawara T, Yamashita K, Asai A, Nagao A, Shiraishi T, Imai I, Hirata T. Esterification of xanthophylls by human intestinal Caco-2 cells. Arch Biochem Biophys. 2009;483:205–12.
16. Mattea F, Martin A, Cocero MJ. Carotenoid processing with supercritical fluids. J Food Eng. 2009;93:255–65.
17. Kang JN, Han JW, Yang SH, Lee SM. Co-expression analysis reveals differential expression of homologous genes associated with specific terpenoid biosynthesis in *Rehmannia glutinosa*. Genes. 2022;13(6):1092.
18. Shumskaya M, Wurtzel ET. The carotenoid biosynthetic pathway: thinking in all dimensions. Plant Sci. 2013;208:58–63.
19. CooperstoneJL, Schwartz SJ. Recent insights into health benefits of carotenoids. Handbook on natural pigments in food and beverages. Elsevier. 2016; pp. 473– 497. https://doi.org/10.1016/B978-0-08-100371-8.00020-8.
20. Ong AS, Tee ES. In: Methods E, editor. Natural sources of carotenoids from plants and oils. Volume 213. London, UK: Academic; 1992. pp. 142–67.
21. Moran NA, Jarvik T. Lateral transfer of genes from fungi underlies carotenoid production in aphids. Science. 2010;328:624–7.
22. Zhou P, Li H, Lin Y, Zhou Y, Chen Y, Li Y, Qiu X. Omics analyses of *Rehmannia glutinosa* dedifferentiated and cambial meristematic cells reveal mechanisms of Catalpol and Indole alkaloid biosynthesis. BMC Plant Biol. 2023;23(1):463.
23. Lee E-K, Jin Y-W, Park JH, Yoo YM, Hong SM, Amir R, Yan Z, Kwon E, Elfick A, Tomlinson S. Cultured cambial meristematic cells as a source of plant natural products. Nat Biotechnol. 2010;28:1213–7.
24. Zhou P, Yang J, Zhu J, He S, Zhang W, Yu R, Zi J, Song L, Huang X. Effects of β-cyclodextrin and Methyl jasmonate on the production of vindoline, catharanthine, and ajmalicine in *Catharanthus roseus* cambial meristematic cell cultures. Appl Microbiol Biotechnol. 2015;99:7035–45.
25. Bouvier F, D'Harlingue A, Camara B. Carotenoid cleavage enzymes in plants and in the human diet: potential implications in health and disease. Biochim Et Biophys Acta (BBA) - Lipids Lipid Metabolism. 2005;1740(2):227–33.
26. Zhao J, Dixon RA. Molecular genetic dissection of the lignin biosynthetic pathway. Plant Physiol. 2010;153(3):1046–56. https://doi.org/10.1104/pp.110.157411.
27. Liu M, Xu Y, Ma X. Flavonoids and anthocyanins as natural colorants in food and beverages. Front Chem. 2019;7:75. https://doi.org/10.3389/fchem.2019.00075.
28. Zhang L, Duan Z, Ma S, Sun S, Sun M, Xiao Y, Sun Y. SlMYB7, an AtMYB4-like R2R3-MYB transcription factor, inhibits anthocyanin accumulation in *Solanum lycopersicum* fruits. J Agric Food Chem. 2023;71(48):18758–68.
29. Mo R, Han G, Zhu Z, Essemine J, Dong Z, Li Y, Yu C. The ethylene response factor ERF5 regulates anthocyanin biosynthesis in 'Zijin'mulberry fruits by interacting with *MYBA* and *F3H* genes. Int J Mol Sci. 2022;23(14):7615.
30. Quan W, Chan Z, Wei P, Mao Y, Bartels D, Liu X. PHD finger proteins function in plant development and abiotic stress responses: an overview. Front Plant Sci. 2023;14:1297607.
31. Lee WY, Lee D, Chung WI, Kwon CS. Arabidopsis ING and Alfin1-like protein families localize to the nucleus and bind to H3K4me3/2 via plant homeodomain fingers. Plant J. 2009;58(3):511–24.
32. Hao S, Ma Y, Zhao S, Ji Q, Zhang K, Yang M, Yao Y. McWRI1, a transcription factor of the AP2/SHEN family, regulates the biosynthesis of the cuticular waxes on the Apple fruit surface under low temperature. PLoS ONE. 2017;12(10):e0186996.
33. Lin YX, Jiang HY, Chu ZX, Tang XL, Zhu SW, Cheng BJ. Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. BMC Genomics. 2011;12(1):76.
34. Huang M, Liu W, Su D, Abid G, Huang B, Li Z. The BES1/BZR1 transcriptional factor SlBES2 regulates photosynthetic apparatus in tomato fruit. BMC Plant Biol. 2025;25(1):70.
35. Zeng T, Yu Q, Shang J, Xu Z, Zhou L, Li W, Wang C. TcbHLH14, a jasmonate associated MYC2-like transcription factor positively regulates pyrethrin biosynthesis in *Tanacetum cinerariifolium*. Int J Mol Sci. 2023;24(8):7379.
36. Gai S, Du B, Xiao Y, Zhang X, Turupu M, Yao Q, Li T. bZIP transcription factor PavbZIP6 regulates anthocyanin accumulation by increasing abscisic acid in sweet Cherry. Int J Mol Sci. 2024;25(18):10207.
37. Wang Y, Li X, Qiu H, Chen R, Xiong A, Xu Z, Ou L. The MADS-RIPENING INHIBITOR–DIVARICATA1 module regulates carotenoid biosynthesis in nonclimacteric capsicum fruits. Plant Physiol. 2025;197(2):kiaf013.
38. Li H, Tian W, Fan Y. Transcriptome and proteome analysis identified Genes/Proteins involved in the regulation of leaf color in *Paulownia Fortunei*. Horticulturae. 2025;11(4):441.
39. Rivero-Manzanilla G, Narváez-Zapata JA, Aguilar-Espinosa M, Carballo-Uicab VM, Rivera-Madrid R. Gene structure and potential regulation of the lycopene cyclase genes in Bixa Orellana L. Physiol Mol Biology Plants. 2023;29(10):1423–35.
40. Apel W, Bock R. Enhancement of carotenoid biosynthesis in transplastomic tomatoes by induced lycopene-to-provitamin A conversion. Plant Physiol. 2009;151(1):59–66.
41. Okamoto S, Yu F, Harada H, Okajima T, Hattan JI, Misawa N, Utsumi R. A short-chain dehydrogenase involved in terpene metabolism from *Zingiber Zerumbet*. FEBS J. 2011;278(16):2892–900.
42. Rajagopal R. Metabolism of indole-3-acetaldehyde. III. Some characteristics of the aldehyde oxidase of Avena coleoptiles. Physiol Plant. 1971;24(2):272–81.
43. Moummou H, Kallberg Y, Tonfack LB, Persson B, van Der Rest B. The plant short-chain dehydrogenase (SDR) superfamily: genome-wide inventory and diversification patterns. BMC Plant Biol. 2012;12(1):219.
44. Rivero-Manzanilla G, Narváez-Zapata JA, Aguilar-Espinosa M, Carballo-Uicab VM, Rivera-Madrid R. Gene structure and potential regulation of the lycopene cyclase genes in *Bixa Orellana* L. Physiol Mol Biology Plants. 2023;29(10):1423–35.
45. Zhao S, Tuan PA, Kim JK, Park WT, Kim YB, Arasu MV, Park S. U. Molecular characterization of carotenoid biosynthetic genes and carotenoid accumulation in *Lycium Chinense*. Molecules. 2014;19(8):11250–62.
46. Namitha KK, Archana SN, Negi PS. Expression of carotenoid biosynthetic pathway genes and changes in carotenoids during ripening in tomato (*Lycopersicon esculentum*). Food Function. 2011;2(3-4):168–173.
47. Moreno JC, Mi J, Alagoz Y, Al-Babili S. Plant apocarotenoids: from retrograde signaling to interspecific communication. Plant J. 2021;105(2):351–75.
48. Yazaki K. ABC transporters involved in the transport of plant secondary metabolites. FEBS Lett. 2006;580(4):1183–91.
49. Dean JV, Willis M, Shaban L. Transport of acylated anthocyanins by the Arabidopsis ATP-binding cassette transporters AtABCC1, AtABCC2, and AtABCC14. Physiol Plant. 2022;174(5):e13780.
50. Li JW, Zhou P, Hu ZH, Xiong AS, Li XH, Chen X, Zhuang J. Carotenoids accumulation and transcript profiling of carotenoid biosynthesis and degradation genes in new shoots and mature leaves of tea plants. Plant Growth Regul. 2025;105:403–416 .
51. Camara B. Plant phytoene synthase complex: component enzymes, immunology, and biosynthesis. Methods Enzymol. 1993;214:352–65.
52. Cunningham FX, Gantt E. Genes and enzymes of carotenoid biosynthesis in plants. Annu. Rev. Plant physiol. Plant Mol Biol. 1998;49:557–83.
53. Chen P, Wei X, Qi Q, Jia W, Zhao M, Wang H, Duan H. Study of terpenoid synthesis and prenyltransferase in roots of *Rehmannia glutinosa* based on iTRAQ quantitative proteomics. Front Plant Sci. 2021;12:693758.
54. Leng X, Wang P, Wang C, Zhu X, Li X, Li H, Fang J. Genome-wide identification and characterization of genes involved in carotenoid metabolism in three stages of grapevine fruit development. Sci Rep. 2017;7(1):4216.

55. Cheniclet C, Rafia F, Saint-Guily A, Verna A, Carde JP. Localization of the enzyme geranyl-geranylpyrophosphate synthase in capsicum fruits by Immunogold cytochemistry after conventional chemical fixation or quick freezing followed by freeze-substitution. Labelling evolution during fruit ripening. Biol Cell. 1992;75:145–54.
56. Walter MH, Fester T, Strack D. Arbuscular mycorrhizal fungi induce the non-mevalonate Methylerythritol phosphate pathway of isoprenoid biosynthesis correlated with accumulation of the yellow pigment and other apo-carotenoids. Plant J. 2000;21:571–8.
57. Walter MH, Hans J, Strack D. Two distantly related genes encoding 1-deoxy-D-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. Plant J. 2002;31:243–54.
58. Córdoba E, et al. Functional characterization of the three genes encoding 1-deoxy-D-xylulose 5-phosphate synthase in maize. J Exp Bot. 2011;62:2023–38.
59. Deng Z, Hu Y, Wang Y. The role of phenylpropanoid metabolism in the response of *Rehmannia glutinosa* to environmental stress. Plant Physiol Biochem. 2017;115:145–52.
60. Hernández I, Ferrer JL, de la Torre AM. The phenylpropanoid pathway in plants: overview and applications. Plant Physiol Biochem. 2014;85:159–71.
61. Chagné D, Lin-Wang K, Volz RK, How NM, Rouse S, Brebdolise S, Carlisle CM, Kumar S, De Silva N, Micheletti D, McGhie T, Crowhurst RN, Storey RD, Velasco R, Hellens RP, Gardiner SE, Allan AC. An ancient duplication of Apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. Plant Physiol. 2013;161:225–39.
62. Song X, Yang Q, Liu Y, Li J, Chang X, Xian L, Zhang J. Genome-wide identification of pistacia R2R3-MYB gene family and function characterization of PcMYB113 during autumn leaf coloration in pistacia chinensis. Int J Biol Macromol. 2021;192:16–27.
63. Zhang X, Wu Q, Li X. Genetic analysis of the phenylpropanoid pathway and its role in pigmentation of *Rehmannia glutinosa*. Plant Biotechnol J. 2020;18(2):348–62.
64. Li X, Li H, Jiang M. Sugar metabolism and its impact on anthocyanin biosynthesis in plants. Front Plant Sci. 2014;5:310.
65. Dixon RA, Paiva NL. Stress-induced phenylpropanoid metabolism. Plant Cell. 1995;7(7):1085–97.
66. Zhao X, Zhang Y, Long T, Wang S, Yang J. Regulation mechanism of plant pigments biosynthesis: anthocyanins, carotenoids, and betalains. Metabolites. 2022;12(9):871.
67. Rodriguez-Amaya DB. A guide to carotenoid analysis in foods. Volume 71. Washington, DC, USA: ILSI; 2001.
68. Qiao H, Zhao W, Tian S, Wang D, Wu H, Wang C, Zhu J, Li N, Zhu X, Mu S, Zhang J, He H. Exploring the mechanisms underlying petal pigmentation differences in two cultivars of physalis Philadelphica based on HPLC and NGS. Horticulturae. 2024;10(5):507.
69. Chen S, Zhou Y, Chen Y, Gu J, Fastp. An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90. https://doi.org/10.1093/bioinformatics/bty560.
70. Varet H, Brillet-Guéguen L, Coppée J-Y, Dillies M-A, SARTools. A DESeq2- and EdgeR-based R pipeline for comprehensive differential analysis of RNA-seq data. PLoS ONE. 2016;11:e0157022.
71. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh. M. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008;36:D480–4.
72. Ashburner M, Ball CA, Blake JA, Botstein D, Cherry JM. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet. 2000;25:25–9.
73. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods. 2001;25(4):402–8.
74. Deng L, Han Y, Tang C, Liao Q, Li Z. Label-free mass spectrometry-based quantitative proteomics analysis of serum proteins during early pregnancy in jennies (Equus asinus). Front Veterinary Sci. 2020;7:569587.

Publisher's Note