

## **Meta-analysis of RNA-seq data of the medicinal plant *Catharanthus roseus* under some biotic and abiotic stresses**

**Seyede Nasim Tabatabaeipour** 

Ph.D. Student, Department of Plant Breeding and Biotechnology, Faculty of Agriculture, Shahrekord University, Shahrekord, Iran. E-mail address: sntabatabaeipour@gmail.com

**Behrouz Shiran** 

\*Corresponding author. Professor, Department of Plant Breeding and Biotechnology, Faculty of Agriculture, Shahrekord University, Shahrekord, Iran. E-mail address: beshiran45@gmail.com

**Rudabeh Ravash** 

\*Corresponding author. Assistant Professor, Department of Plant Breeding and Biotechnology, Faculty of Agriculture, Shahrekord University, Shahrekord, Iran. E-mail address: r.ravash@gmail.com

**Ali Niazi** 

Professor, Department of Biotechnology, Faculty of Agriculture, University of Shiraz, Shiraz, Iran. E-mail address: niazi@shirazu.ac.ir

**Esmaeil Ebrahimie**

Associate Professor, Dame Roma Mitchell Cancer Research Laboratories, Adelaide Medical School, University of Adelaide, Adelaide, South Australia, Australia. E-mail address: ebrahimiet@gmail.com

---

### ***Abstract***

#### **Objective**

The objectives of genetic and metabolic engineering can be greatly advanced through comprehending the molecular processes of the stress response. Researchers may access an abundance of data through transcriptome studies. Using advanced statistical methods such as meta-analysis to integrate data from many sources offers a new way to find differentially expressed genes (DEGs), get beyond biological complexity, and provide more accurate results. The current work used transcriptome data from RNA-seq meta-analysis to uncover *Catharanthus roseus* DEGs in response to stressors and the production of important secondary metabolites.

#### **Materials and methods**

The EMBL-EBI database provided the RNA-seq data, which was then pre-processed and mapped quality reads using STAR software to the reference genome of *C. roseus*. Each data set's changes

in gene expression were examined independently using the edgeR package, and the output from these analyses was then utilized for a meta-analysis with the use of the metaRNAseq program. Key genes in reaction to stress and pathways were finally examined, along with the biological function and biological pathways involved in the different and significant expression of these genes in response to stress. Co-expression analysis was carried out using the WGCNA software and hub genes associated with stress response and the production of important plant-related metabolites were found.

## Results

The findings of the meta-analysis show that 772 genes with average expression  $\log_2FC \geq |1$  and  $FDR \leq 0.05$  were found, and that 305 and 467 of these genes, respectively, had up- and downregulated expression under stress. Of these, only a meta-analysis enables the identification of 27 genes that were not identified as DEGs in individual analyses. The functional enrichment results of DEGs demonstrate their crucial importance in metabolic processes, response to stimuli, and cellular processes. They also lead multiple pathways, such as the MAPK signaling pathway, the biosynthesis of secondary metabolites, the transduction of plant hormones, and metabolic pathways under stress. Through co-expression network analysis, stress-responsive hubs genes and associated with the TIA biosynthesis pathways were found that may be potential genes involved in the synthesis of therapeutic compounds exclusive to plants.

## Conclusions

In this study, we identified key and novel genes by investigating the molecular responses of *C. roseus* plant under various stresses using meta-analysis and systems biology approaches. These genes have the potential to be employed as candidate genes in genetic and metabolic engineering projects aimed at enhancing the plant's capability to produce therapeutic metabolites in the future.

**Keywords:** *Catharanthus roseus*, RNA-seq, Meta-analysis, Stress, Terpenoid Indole Alkaloids

**Paper Type:** Research Paper.

**Citation:** Tabatabeipour S.N, Shiran B, Ravash R, Niazi A, Ebrahimie E (2024) Meta-analysis of RNA-seq data of the medicinal plant *Catharanthus roseus* under some biotic and abiotic stresses. *Agricultural Biotechnology Journal* 16 (1), 209-234.

---

*Agricultural Biotechnology Journal* 16 (1), 209-234. DOI: 10.22103/jab.2024.22497.1523

Received: November 12, 2023.

Received in revised form: January 12, 2024.

Accepted: January 13, 2024.

Published online: February 20, 2024.




Publisher: Faculty of Agriculture and Technology Institute of Plant Production, Shahid Bahonar University of Kerman-Iranian Biotechnology Society.

© the authors

## فراواکای داده‌های RNA-seq گیاه دارویی پروانش (*Catharanthus roseus*) تحت برخی

### تنش‌های زیستی و غیرزیستی

سیده نسیم طباطبایی پور 


دانشجوی دکتری بیوتکنولوژی کشاورزی، گروه اصلاح نباتات و بیوتکنولوژی، دانشکده کشاورزی، دانشگاه شهرکرد، شهرکرد، ایران. رایانامه: [sntabatabaeipour@gmail.com](mailto:sntabatabaeipour@gmail.com)

بهروز شیران 

\*نویسنده مسئول: استاد، گروه بیوتکنولوژی و اصلاح نباتات، دانشکده کشاورزی، دانشگاه شهرکرد، شهرکرد، ایران. رایانامه: [beshiran45@gmail.com](mailto:beshiran45@gmail.com)

رودابه راوش 

\*نویسنده مسئول: استادیار، گروه بیوتکنولوژی و اصلاح نباتات، دانشکده کشاورزی، دانشگاه شهرکرد، شهرکرد، ایران. رایانامه: [r.ravash@gmail.com](mailto:r.ravash@gmail.com)

علی نیازی 

استاد، گروه بیوتکنولوژی، دانشکده کشاورزی، دانشگاه شیراز، شیراز، ایران. رایانامه: [niazi@shirazu.ac.ir](mailto:niazi@shirazu.ac.ir)

اسماعیل ابراهیمی

دانشیار، آزمایشگاه تحقیقات سرطان، دانشکده پزشکی آدلاید، دانشگاه آدلاید، استرالیای جنوبی، استرالیا. رایانامه: [ebrahimiet@gmail.com](mailto:ebrahimiet@gmail.com)

تاریخ دریافت: ۱۴۰۲/۰۸/۲۲ تاریخ دریافت فایل اصلاح شده نهایی: ۱۴۰۲/۱۰/۲۲ تاریخ پذیرش: ۱۴۰۲/۱۰/۲۳

### چکیده

**هدف:** اهداف مهندسی ژنتیک و متابولیک را می‌توان با درک فرآیندهای مولکولی پاسخ به تنش به میزان زیادی محقق کرد. پژوهشگران می‌توانند از طریق مطالعات ترنسکریپتوم به داده‌های فراوانی در این خصوص دسترسی پیدا کنند. استفاده از روش‌های آماری پیشرفته مانند فراواکای برای ادغام داده‌های بسیاری از منابع، روش جدیدی برای یافتن ژن‌های با بیان متفاوت (DEG) (Differentially Expressed Genes)، برای غلبه بر پیچیدگی زیستی و ارائه نتایج دقیق‌تر ارائه می‌دهد. مطالعه فعلی از فراواکای داده‌های بیانی RNA-seq گیاه دارویی پروانش *Catharanthus roseus* برای کشف DEG‌های غنی‌سازی شده (enrichmented DEG) در پاسخ به عوامل تنش‌زا و تولید متابولیت‌های ثانویه مهم استفاده کرد.

**مواد و روش‌ها:** داده‌های RNA-seq از پایگاه داده EMBL-EBI دریافت و پس از پیش پردازش، نقشه‌یابی خوانش‌های با کیفیت بر روی ژنوم مرجع پروانش با نرم‌افزار STAR صورت گرفت. تغییرات بیان ژن‌ها با استفاده از بسته edgeR به صورت جداگانه برای هر مجموعه داده، بررسی و سپس از خروجی حاصله برای فراواکاوی با استفاده از بسته metaRNASeq استفاده شد. ژن‌های با بیان متفاوت و معنی‌دار در پاسخ به تنش‌های زیستی (تیمارهای حشره *Manduca sexta* سم باکتری Coronatine، باکتری *Phytoplasma* و باکتری Huanglongbing) و غیرزیستی (تیمارهای اتیلن و متیل جاسمونات)، از نظر عملکرد زیستی و مسیرهای زیستی درگیر بررسی شدند و تجزیه و تحلیل هم‌بیانی با بسته WGCNA انجام شد و در نهایت ژن‌های کلیدی و هاب در پاسخ به تنش و مسیر بیوسنتزی تولید متابولیت‌های مهم ترپنوئید ایندول آلکالوئیدها (Terpenoid Indole Alkaloids (TIAs)) تعیین گردید.

**نتایج:** یافته‌های فراواکاوی، ۷۷۲ ژن با بیان متفاوت دارای  $(\text{Log}_2 \text{FC} \geq 1)$  و  $\text{FDR} \leq 0.05$  (False Discovery Rate) را شناسایی کرد که ۳۰۵ و ۴۶۷ ژن از این ژن‌ها به ترتیب دارای بیان بالا و پایین تحت تنش بودند. در این میان، فراواکاوی امکان شناسایی ۲۷ ژن را فراهم نمود که در آنالیزهای انفرادی به عنوان DEG شناسایی نشده بودند. نتایج غنی‌سازی عملکردی DEG‌ها اهمیت حیاتی آنها را در فرآیندهای متابولیک، پاسخ به محرک‌ها و فرآیندهای سلولی نشان داد. آنها همچنین در مسیرهای متعددی مانند مسیر سیگنالینگ، پروتئین‌کینازهای فعال شده با میتوزن (Mitogen (MAPK)) (activated protein kinases)، بیوسنتز متابولیت‌های ثانویه، انتقال هورمون‌های گیاهی و مسیرهای متابولیک تحت تنش اهمیت معنی‌داری داشتند. از طریق تجزیه و تحلیل شبکه هم‌بیانی، ژن‌های هاب پاسخ‌دهنده به تنش و مرتبط با مسیرهای بیوسنتز ترپنوئید ایندول آلکالوئیدها (TIA) یافت شدند که می‌توانند ژن‌های بالقوه‌ای در سنتز ترکیبات دارویی منحصر به فرد گیاه باشند.

**نتیجه‌گیری:** در این تحقیق، با بررسی پاسخ‌های مولکولی گیاه پروانش تحت تنش‌های مختلف با استفاده از رویکردهای فراواکاوی و زیست‌شناسی سامانه‌ای، ژن‌های کلیدی و جدید شناسایی شدند. این ژن‌ها دارای توانایی بالقوه جهت استفاده به عنوان ژن‌های کاندید در پروژه‌های مهندسی ژنتیک و متابولیک بوده که به منظور افزایش توانایی گیاه در تولید متابولیت‌های درمانی در آینده به کار گرفته شوند.

**کلیدواژه‌ها:** پروانش، RNA-seq، فراواکاوی، تنش، ترپنوئید ایندول آلکالوئیدها.

**نوع مقاله:** پژوهشی.

**استناد:** طباطبائی پور سیده نسیم، شیران بهروز، راوش رودابه، نیازی علی، ابراهیمی اسماعیل (۱۴۰۳) فراواکاوی داده‌های RNA-seq گیاه دارویی پروانش (*Catharanthus roseus*) تحت برخی تنش‌های زیستی و غیرزیستی. *مجله بیوتکنولوژی کشاورزی* ۱۶(۱)، ۲۰۹-۲۳۴.



Publisher: Faculty of Agriculture and Technology Institute of Plant Production, Shahid Bahonar University of Kerman-Iranian Biotechnology Society.

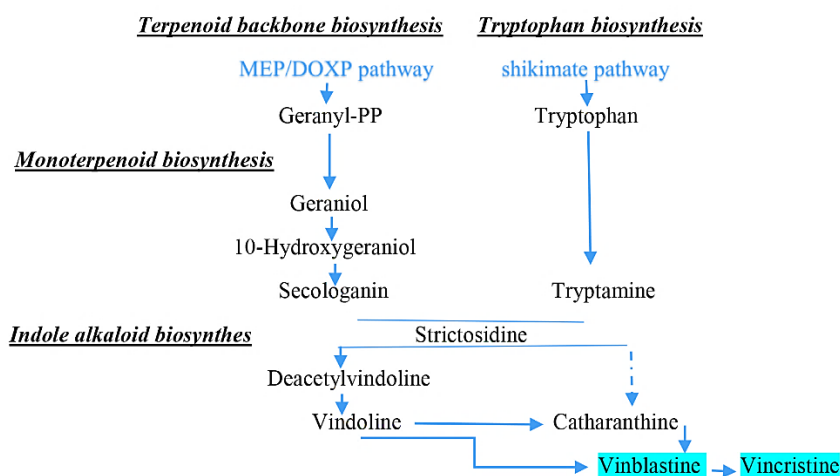
© the authors

## مقدمه

سرطان یکی از مهمترین علل مرگ و میر در سطح جهان است و منجر به حدود ۱۰ میلیون مرگ و میر در سال ۲۰۲۰ شده است (Dhyani et al. 2022). امروزه از داروهای تولید شده شیمیایی و شیمی درمانی برای درمان سرطان استفاده می‌شود (Buga et al. 2019). گیاهان منابع جذابی از مواد شیمیایی درمانی، به ویژه مواد ضد سرطان هستند. این مواد، متابولیت‌های ثانویه‌ای هستند که گیاهان به طور طبیعی برای محافظت از خود در برابر تنش‌های محیطی، عوامل زیستی و غیرزیستی تولید می‌کنند (Paul et al. 2022). این عوامل با تحریک تولید متابولیت، واکنش‌های دفاعی گیاه را برمی‌انگیزد (Al-Khayri et al. 2023). به طور کلی، فنل‌ها، ترپنوئیدها و آلکالوئیدها سه نوع اصلی متابولیت‌های ثانویه را تشکیل می‌دهند (Liu et al. 2017). آلکالوئیدها از انواع اصلی این متابولیت‌ها بوده و ترکیبات نیتروژن‌دار با وزن مولکولی پایین می‌باشند (Pan et al. 2018). ترپنوئید ایندول آلکالوئیدها (TIAs)، به عنوان یکی از مهم‌ترین انواع آن‌ها با خواص دارویی مفید در درمان بیماری‌ها به ویژه سرطان در جهان شناخته شده‌اند. در سلسله گیاهی خانواده‌های Apocynaceae، Rubiaceae و Loganiaceae سه خانواده‌ای هستند که اکثر TIAها در آن‌ها یافت شده است (Soltani et al. 2022). گیاه پروانش به عنوان یکی از مهمترین گیاهان دارویی متعلق به خانواده Apocynaceae با حدود ۱۳۰ نوع TIA می‌باشد. وین‌بلاستین و وین‌کریستین دو TIA مهم در این گیاه بوده که به عنوان مؤثرترین داروهای ضدسرطان شناخته شده‌اند و اغلب در شیمی‌درمانی برای درمان بسیاری از سرطان‌ها در سراسر جهان استفاده می‌شوند (Dhyani et al. 2022). تنها منبع این مواد شیمیایی پروانش است. با این حال، به دلیل تجمع نسبتاً کم و محدود این مواد در گیاه، برای تقاضای جهانی کافی نمی‌باشد (Pan et al. 2018).

با توجه به پیچیدگی بالای مسیر TIAها (شکل ۱) و شناسایی تعداد محدودی از ژن‌های کاندید دخیل در تجمع TIAها، اخیراً از مهندسی ژنتیک و متابولیک برای افزایش تولید آن‌ها استفاده شده است، اما پیشرفت کمی داشته است. در نتیجه، تحقیق بر روی شناخت ژن‌های بالقوه جدید درگیر در تولید و تنظیم TIAها بسیار مهم است (Liu et al. 2017). تنها یک مجموعه نسبتاً کوچک از تمام ژنوم در شرایط خاص بیان می‌شود (Mohammadabadi et al. 2018; Arabpoor Raghbadi et al. 2022). همچنین مقدار محصولات ژن که در همان شرایط ساخته شده سبب تنظیم بیان آن ژن می‌شود (Mohammadabadi et al. 2021; Shahsavari et al. 2021). بهره‌گیری از فناوری RNA-seq، بیان و شناسایی ژن‌ها و بررسی الگوی افتراقی بیان ژن‌ها، یکی از رویکردهای معرفی ژن‌های کاندید مفید در تولید متابولیت‌ها است و مطالعات روی گیاهان دارویی با استفاده از این روش با موفقیت انجام شده است (Guo et al. 2021; Liu et al. 2021). تعداد کم آزمایش‌ها و تکرارها در هر پروژه به دلیل هزینه‌های بالای آن‌ها و داده‌های جمع‌آوری شده در شرایط محدود تنها بخشی از محدودیت‌های این روش است (Bairakdar et al. 2023).

فراواکای یک رویکرد قدرتمند برای یافتن ژن‌های با بیان متفاوت معنی‌دار (Differentially Expressed (DEG) Genes) جدید است، زیرا حجم نمونه و قدرت آماری را افزایش می‌دهد. این راهبرد، قابلیت ترکیب داده‌های رونویسی بسیاری از مطالعات را برای شناسایی مجموعه ژن‌های کنترل‌کننده صفات پیچیده، دارا بوده و از این‌رو فرصتی برای استفاده از داده‌های موجود و شناسایی جنبه‌های مهم فرآیندهای زیستی مرتبط با متابولیت‌های ثانویه فراهم می‌سازد (Tahmasebi et al. 2019). این روش قابلیت اطمینان داده‌های بیانی را بهبود بخشیده و اثرات تغییرپذیری محیطی به واسطه تعداد محدود تکرارهای زیستی و تجربی در مطالعات فردی را کاهش می‌دهد (Rau et al. 2014; De Abreu Neto and Frei 2016; Balan et al. 2018). از آنجا که تنش‌ها باعث تحریک تولید متابولیت‌های ثانویه می‌شوند، در این مطالعه از فراواکای داده‌های RNA-seq تحت تنش‌های مختلف به عنوان محرک در گیاه پروانش به منظور شناسایی DEG‌های جدید و سپس تجزیه و تحلیل غنی‌سازی عملکردی و مسیرهای بیوسنتزی این ژن‌ها استفاده شد. علاوه بر این، تجزیه و تحلیل زیست‌شناسی سیستم‌ها برای یافتن ژن‌های هاب و شبکه‌های ژنی انجام شد که می‌تواند به شناخت بیشتر مسیر بیوسنتزی TIA در جهت تولید متابولیت‌های ثانویه به ویژه داروهای ضد سرطان در آینده کمک نماید. این مطالعه اولین فراواکای روی چندین مجموعه داده RNA-seq مربوط به گیاه پروانش تحت برخی تنش‌ها را ارائه می‌کند.



شکل ۱. مسیرهای بیوسنتزی شماتیک برای ترپنوئید ایندول آلکالوئیدهای (TIAs) پروانش (برگرفته از Liu et al. 2017)

Figure 1. Schematic biosynthetic pathways for terpenoid indole alkaloids (TIAs) in *C. roseus* (Figure adopted from Liu et al. 2017)

مواد و روش‌ها

**جمع‌آوری داده‌های RNA-seq:** برای یافتن داده‌های RNA-seq مرتبط با مطالعات پاسخ به تنش‌های زیستی و غیرزیستی در پروانش، مخزن SRA پایگاه NCBI در زمستان ۱۴۰۰ جستجو و شش مطالعه که با معیارهای مورد نظر در این پژوهش مطابقت داشتند شناسایی شدند. معیارهایی چون داشتن الگوی بیانی RNA-seq، استفاده از پلتفرم Illumina HiSeq برای تولید داده‌ها، کتابخانه‌های از دو انتها خوانش شده و لزوم داشتن هر دو گروه شاهد و تیمار مد نظر قرار داده شد. مطالعات شامل تیمارهای غیرزیستی اتیلن و متیل جاسمونات و عوامل زیستی شامل تیمارهای حشره *Manduca sexta* کروناتین (Coronatine) (سم تولید شده توسط باکتری *Pseudomonas syringae*)، باکتری‌های phytoplasma و *Candidatus Liberibacter asiaticus* (Huanglongbing) بودند. فایل‌های Fastq نمونه‌های انتخاب شده از پایگاه داده نوکلئوتید EMBL-EBI دسترسی آزاد دریافت شد. در نهایت ۴۰ نمونه از مجموع ۷ مطالعه با ۶ تیمار مختلف برای فراواکاوی انتخاب شدند. در هر مطالعه، نمونه‌ها به منظور شناسایی ژن‌های با بیان متفاوت (DEG)، به دو گروه کنترل و تیمار تقسیم شدند.

**پیش‌پردازش داده‌های RNA-seq:** قبل از نقشه‌یابی خوانش‌ها، داده‌های RNA-seq پیش‌پردازش و یکپارچگی مجموعه داده بررسی شد. ابزار FastQC برای ارزیابی کیفیت هر مجموعه داده استفاده شد (Andrews 2010). در صورت آلودگی آدپتور از Trimmomatic برای تمیز کردن فایل‌های fastq با تنظیمات پیش‌فرض استفاده شد (Bolger et al. 2014).

**نقشه‌یابی خوانش‌ها و کمی‌سازی بیان:** خوانش‌های نقشه‌یابی شده روی ژنوم مرجع گیاه پروانش که از طریق فایل‌های اسمبلی ژنوم و حاشیه‌نویسی در قالب GFF3 با استفاده از (Dobin et al. 2013) STAR (v 2.7.0) ایجاد شد، با پارامترهای پیش‌فرض شمارش شد. تمام تحلیل‌های پیش‌پردازش و نقشه‌یابی با استفاده از ترمینال لینوکس، اوبونتو ۲۰،۰۴ انجام شد. ماتریس شمارش خوانش هر ژن برای مراحل بعدی مورد استفاده قرار گرفت.

**تصحیح اثرات دسته‌ای (Batch-effect correction):** این تصحیح با استفاده از نرم‌افزار ComBat-seq روی ماتریس داده‌های شمارش خام به عنوان ورودی با پارامترهای پیش‌فرض اعمال شد (Zhang et al. 2020).

**حاشیه‌نویسی عملکردی ژن‌ها (Functional annotation of genes):** تمام توالی‌های ژن بر اساس پایگاه‌های داده Nr، KOG/COG، Swiss-Prot protein، InterPro و TAIR با استفاده از BLASTX ( $E\text{-value} \leq 10^{-5}$ ) حاشیه‌نویسی شدند. برای حاشیه‌نویسی بیشتر، برنامه Blast2GO نسخه ۳،۰،۶ (Conesa et al. 2005) برای تخصیص اصطلاحات GO، شماره EC و اطلاعات KEGG (Kanehisa and Goto 2000) با توجه به نتایج BLAST به کار رفت.

**شناسایی ژن‌های دخیل در بیوسنتز TIA:** ژن‌های مرتبط با مسیر TIA‌ها از طریق انواع مختلف بلاست‌ها شناسایی شدند. مسیر TIA‌ها منجر به تولید وین‌بلاستین و وین‌کریستین می‌شود. مسیرهای بیوسنتز پایه ترپنوئیدی Terpenoid backbone biosynthesis [KO00900]، مونوترپنوئیدی Monoterpenoid biosynthesis [KO00902]، فنیل‌آلنین، تیروزین، و تریپتوفان Phenylalanine, tyrosine and tryptophan biosynthesis [KO00400] و ایندول آلکالوئید

[KO00901Indole alkaloid biosynthesis] همه درگیر در تولید TIAها می‌باشند. توالی‌های ژنی پروانش با ژن‌های مربوط به تولید TIAها با استفاده از پایگاه‌های داده KEGG حاشیه نویسی شدند و با استفاده از پایگاه‌های مختلف بلاست شدند.

**بیان افتراقی ژن‌ها:** با استفاده از بسته نرم‌افزاری edgeR (v3.16.5) (Robinson and McCarthy 2010)

ماتریس تعداد خوانش‌ها، نرمال‌سازی و ژن‌های با بیان متفاوت (DEG) بین نمونه‌های کنترل و تیمار در هر مطالعه RNA-seq شناسایی شدند. ژن‌های با  $P\text{-value} < 0.05$  و  $\log_2FC \leq -1$  یا  $\log_2FC \geq 1$  به عنوان ژن‌های با بیان متفاوت قلمداد شدند.

**فراواکاوی (متاآنالیز):** فراواکاوی داده‌های RNA-seq حاصل از مطالعات تیمارهای تنشی در این تحقیق انجام شد

(نوع تیمارها در قسمت جمع‌آوری داده‌های RNA-seq گفته شده است). شش مطالعه با شناسه‌های دسترسی PRJEB14626، PRJEB25987، PRJNA236160، PRJNA431141، PRJNA185483 و PRJNA358259 شامل ۱۸ نمونه شاهد و ۲۲ نمونه تیمار در فراواکاوی مورد تجزیه و تحلیل قرار گرفت. شماره دسترسی PRJNA358259 شامل اعمال دو نوع تیمار متفاوت در دو بخش مجزا و در قالب دو مطالعه آنالیز شد. فراواکاوی با استفاده از بسته metaRNASeq انجام شد (Rau et al. 2014). آزمون احتمال ترکیبی فیشر توسط metaRNASeq پیاده‌سازی و در مطالعات متعدد استفاده می‌شود. این رویکرد، همانطور که از نام آن پیداست، مقادیر  $p$  (p-values) برای هر ژن از مطالعات مختلف را با استفاده از رابطه ۱ ترکیب می‌کند.

$$F_g = -2 \sum_{s=1}^S \ln(p_{gs}) \quad (\text{رابطه ۱})$$

که در این رابطه مقدار  $p$  خام برای یک ژن ( $g$ ) در تجزیه و تحلیل افتراقی برای مطالعه ( $s$ ) در اینجا با  $p_{gs}$  نشان داده شده است.  $F_g$  دارای توزیع مربع کای (Chi-squared distribution) با درجه آزادی  $2S$  با مقادیر  $p$  مستقل است. از آنجایی که مقادیر  $p$  کوچکتر منجر به  $F_g$  بزرگتر می‌شود، فرضیه صفر رد می‌شود. نرخ کشف کاذب بنجامینی-هوکرگ-Benjamini Hochberg false discovery rate (FDR) برای اصلاح مقادیر  $p$  استفاده شد و مقادیر  $p$  تعدیل شده  $0.05$  از نظر آماری معنی‌دار در نظر گرفته شد. در این مطالعه، ژن‌هایی با FDR برابر با  $0.05$  یا کوچکتر از آن و تغییرات فولد چنج (Fold Change) بزرگتر یا مساوی  $2$  ( $FC \geq 2$ ) یا لگاریتم فولد چنج بزرگتر یا کوچکتر از  $1 \pm \log_2FC \leq -1$  یا  $\log_2FC \geq 1$  به عنوان ژن‌های با بیان متفاوت معنی‌دار (DEG) در نظر گرفته شدند. سپس این ژن‌ها برای تجزیه و تحلیل‌های پایین‌دستی استفاده شدند.

**هستی‌شناسی ژن (انتولوژی ژن (Gene Ontology)) و حاشیه‌نویسی مسیر DEGها:** هستی‌شناسی ژنی

DEGها (GO) با استفاده از برنامه Blast2GO نسخه ۳،۰،۶ (Conesa et al. 2005) و ابزار تحت وب DAVID (<http://david.abcc.ncifcrf.gov>) برای طبقه‌بندی‌های عملکردی GO با توجه به فرآیند زیستی، عملکرد مولکولی، و اجزاء سلولی در آستانه معنی‌دار  $p\text{-value} < 0.05$  انجام و با استفاده از ابزار Blast2GO، مسیرهای کلیدی پایگاه داده KEGG شناسایی شدند.



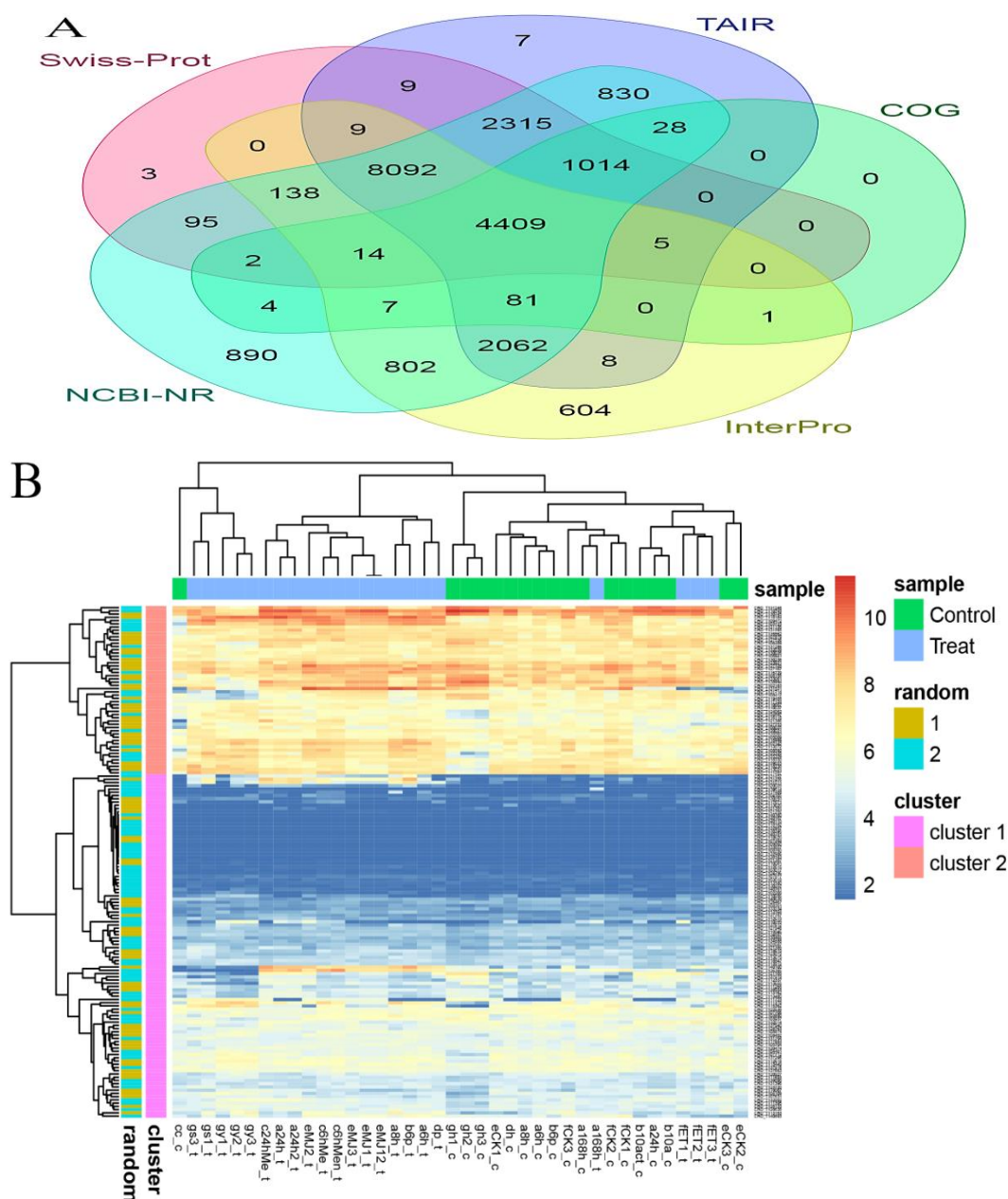
**تجزیه و تحلیل شبکه هم‌بیان:** تجزیه و تحلیل شبکه هم‌بیان ژنی وزن دار (Langfelder and Horvath 2008) بر روی ماتریس مقادیر بیان نرمال شده DEGها برای همه نمونه‌ها به منظور شناسایی گروه‌هایی از DEGها با الگوهای بیان مشابه انجام شد. بر اساس همبستگی پیرسون بین هر جفت DEG، یک ماتریس شباهت تولید شد که با استفاده از یک تابع توان به یک ماتریس مجاورتی تبدیل شد. سپس از تحلیل خوشه‌بندی سلسله مراتبی برای محاسبه ماتریس همپوشانی توپولوژیکی (TOM) استفاده شد. برای کشف ماژول‌های هم‌بیان ژن، در نهایت یک الگوریتم برشی درختی پویا پیاده‌سازی شد. حداقل اندازه ماژول ۳۰ ژن و ارتفاع برش ۰/۹۷۵ برای تعریف ماژول‌ها به کار رفت. امتیازات ارتباطی محاسبه شد و ژن‌های یک ماژول با بالاترین ارتباط اتصالی به عنوان ژن هاب انتخاب شدند. با استفاده از ابزار CytoHubba (Chin et al. 2014) از پلتفرم نرم‌افزار Maximal Clique Centrality (MCC)، ژن‌های هاب ماژول‌های انتخاب شده شناسایی شد. تابع Cytoscape v3.9.1 برای تحلیل محاسباتی ارتباط بین گره‌ها استفاده شد. ابزار آنالین DAVID برای انجام تحلیل عملکردی GO ماژول‌های مهم استفاده شد.

## نتایج و بحث

در این بررسی، ۴۰ نمونه از داده‌های بیان ژن گیاه پروانش تحت تنش‌های مختلف از مجموعه داده‌های در دسترس عموم، جمع‌آوری و با ترکیب داده‌های بیانی تنش‌های مختلف، از فراواکوی مجموعه داده‌های RNA-seq برای شناسایی ژن‌های کاندید احتمالی در پاسخ به عوامل تنش و تولید متابولیت‌های مهم به ویژه متابولیت‌های ثانویه وین کریستین و وین بلاستین با عملکرد ضدسرطان استفاده شد.

### حاشیه نویسی عملکردی کل یونی ژن‌ها: با استفاده از جستجوی بلاست ( $E\text{-value} \leq 10^{-5}$ )، حاشیه‌نویسی

یونی ژن‌ها (یک یونی ژن رونوشت منحصر به فرد با شناسه خاص موجود است که از یک ژن رونویسی می‌شود) بر اساس شباهت توالی در برابر پایگاه‌های داده عمومی، از جمله پایگاه داده پروتئین Nr، Swiss-Prot، TAIR، InterPro، و COG در نهایت ۲۲۸۱۴ یونی ژن از کل ۳۴۳۶۳ یونی ژن موجود در نمونه‌ها بیان شد، در حالی که یونی ژن‌های دیگر در هیچ نمونه‌ای بیان نشدند. در مجموع، ۴۴۰۹ یونی ژن (۱۹/۳۲ درصد) به طور قابل توجهی در همه پایگاه‌های داده مطابقت داشتند (شکل ۲A). علاوه بر این، تمام یونی ژن‌ها با ژن‌های مربوط به بیوسنتز TIA با استفاده از پایگاه‌های داده KEGG حاشیه‌نویسی شدند و با پایگاه‌های اطلاعاتی متعدد بلاست شدند. کل بلاست‌ها ۱۵۹ ژن دارای شناسه ژنی اختصاصی پروانش [شناسه‌های ژنی که با CRO (برگرفته از نام علمی گیاه) آغاز می‌شوند و به آن‌ها ژن‌های CRO (*Catharanthus roseus* genes) اطلاق می‌شود]، مرتبط با مسیر TIA را نشان دادند و نقشه حرارتی (Heatmap) این ژن‌ها در شکل ۲B نشان داده شده است.



شکل ۲. (A) ون دیاگرام کل یونیزن ها که با استفاده از پنج پایگاه داده متفاوت حاشیه نویسی شده است. (B) نقشه حرارتی ژن های CRO مربوط به مسیر TIA که محور عمودی مربوط به ژن ها و محور افقی نشان دهنده اسامی نمونه ها می باشد

Figure 2. (A) Venn diagram of the unigenes annotated using the five-different database. (B) The heat map of CRO genes related to the TIA pathway, where the vertical axis corresponds to the genes and the horizontal axis indicates the names of the samples

شناسایی DEGها با استفاده از فراواکاو: ۷۷۲ ژن با بیان متفاوت بین شرایط تیمار و کنترل توسط فراواکاو شناسایی شد که بیان آن‌ها در پاسخ به شرایط تنش تغییر کرد که به ترتیب ۳۰۵ (۴۵ درصد) و ۴۶۷ (۵۵ درصد) متاژن افزایش و کاهش بیان داشتند. در میان DEGهای شناسایی شده، ژن‌هایی چون ژن *ATHB-7* (خانواده ژنی HD-ZIP) کدکننده پروتئین هومئوباکس-زیپ لوسین (Homeobox-leucine zipper protein *ATHB-7*)، آلفاگلوکان فسفریلاز (Alpha-glucan phosphorylase) یا آیزوایزیم H (H isozyme) (نام جایگزین: فسفریلاز H نشاسته (Starch phosphorylase H))، فسفولیپاز D، *CRO\_T108501* و *CRO\_T110204* بیشترین افزایش بیان را داشتند. در بین DEGها ۲۷ متاژن در هیچ یک از مطالعات تکی یافت نشد که به ترتیب شامل ۹ و ۱۸ ژن افزایش و کاهش بیان یافته بود. از جمله این ژن‌های جدید شناسایی شده می‌توان به ژن کدکننده پروتئین ATP سنتاز فرضی (*YMF19*)، ناحیه انگشت روی (*DOF5.3*) (Dof zinc finger protein) و برخی ژن‌های *DOF5.3*، ژن شبه پروتئین *HSPRO2* مقاومت به نماتد (Nematode resistance protein-like *HSPRO2*) و برخی ژن‌های *CRO* شناخته نشده اشاره کرد که در پاسخ به انواع تنش‌ها و محرک‌ها نقش دارند. فاکتورهای رونویسی خانواده HD-ZIP درگیر در تنظیم تنش غیرزیستی گیاه می‌باشند. برخی از این فاکتورها تحمل به خشکی را در گیاهان از طریق مسیر انتقال سیگنال ABA تنظیم می‌کنند. *ATHB6* یک ژن خانواده HD-ZIP I است که انتقال سیگنال ABA را به طور منفی تنظیم می‌کند و هدف پروتئین فسفاتاز ABI1 است که در پایین دست ABI1 عمل می‌کند. ABI1، یک پروتئین فسفاتاز C2، جزء کلیدی در سیگنال دهی ABA است (Li et al. 2022). نتایج نشان می‌دهد که سلول‌ها ایزوآنزیم‌های مختلف را در شرایط تنش بیان می‌کنند، فسفریلاز نشاسته PHS1 (ایزوآنزیم H) با آزادسازی گلوکز-۱-فسفات از انتهای غیر احیاکننده زنجیره‌های گلوکان به تجزیه نشاسته تحت تنش خشکی کمک می‌کند و مشخص نیست که دیگر آنزیم‌های تجزیه کننده نشاسته به طور بالقوه در این فرآیند دخالت دارند یا خیر (Thalmann and Santelia 2017; Zhang et al. 2018). فسفولیپاز D (*PLD*) در واکنش به تنش‌های متعدد در گیاه نقش دارد. رونویسی و فعالیت ژن آن با قرارگرفتن در معرض این شرایط افزایش می‌یابد و تغییر سطح پروتئین *PLD* منجر به تغییر تحمل تنش می‌شود. خانواده *PLD*های گیاهی نسبتاً بزرگ و ناهمگن است و ایزوفرم‌های مختلف آن در پاسخ به تنش مجزا نقش دارند. *PLD* و محصول آن، اسید فسفاتیدیک، اثرات خود را با عملکرد در آبشارهای انتقال سیگنال و با تأثیرگذاری بر وضعیت بیوفیزیکی غشاهای لیپیدی اعمال می‌کنند (Distéfano et al. 2015). بسیاری از پروتئین‌های مرتبط با بیوسنتز ATP از جمله پروتئین ATP سنتاز *YMF19* تحت تنش افزایش بیان داشته، و نقش این پروتئین در مطالعه‌ای پیرامون پاسخ به تنش فلز سنگین سرب تأیید شده است (Xia et al. 2019). *DOF*ها گروه بزرگی از TFهای گیاهان هستند، که شامل پروتئین متصل به DNA با یک انگشت روی می‌باشند. ثابت شده که در عملکردهای متنوعی مانند پاسخ به تنش گیاهی و شبکه‌های تنظیمی متابولیک گیاه نقش دارند (Zou and Sun 2023). اعضای خانواده *DOF* همچنین به متابولیسم ثانویه گیاه کمک می‌کنند. برای مثال *AtDOF4.2* در آراییدوپسیس متابولیسم فنیل پروپان را تحت تنش تنظیم می‌کند (Skirycz et al. 2007)، در حالی که *FaDOF2* در توت فرنگی مسیر فرار فنیل پروپانوئید را تنظیم می‌کند (Molina-Hidalgo et al. 2017).

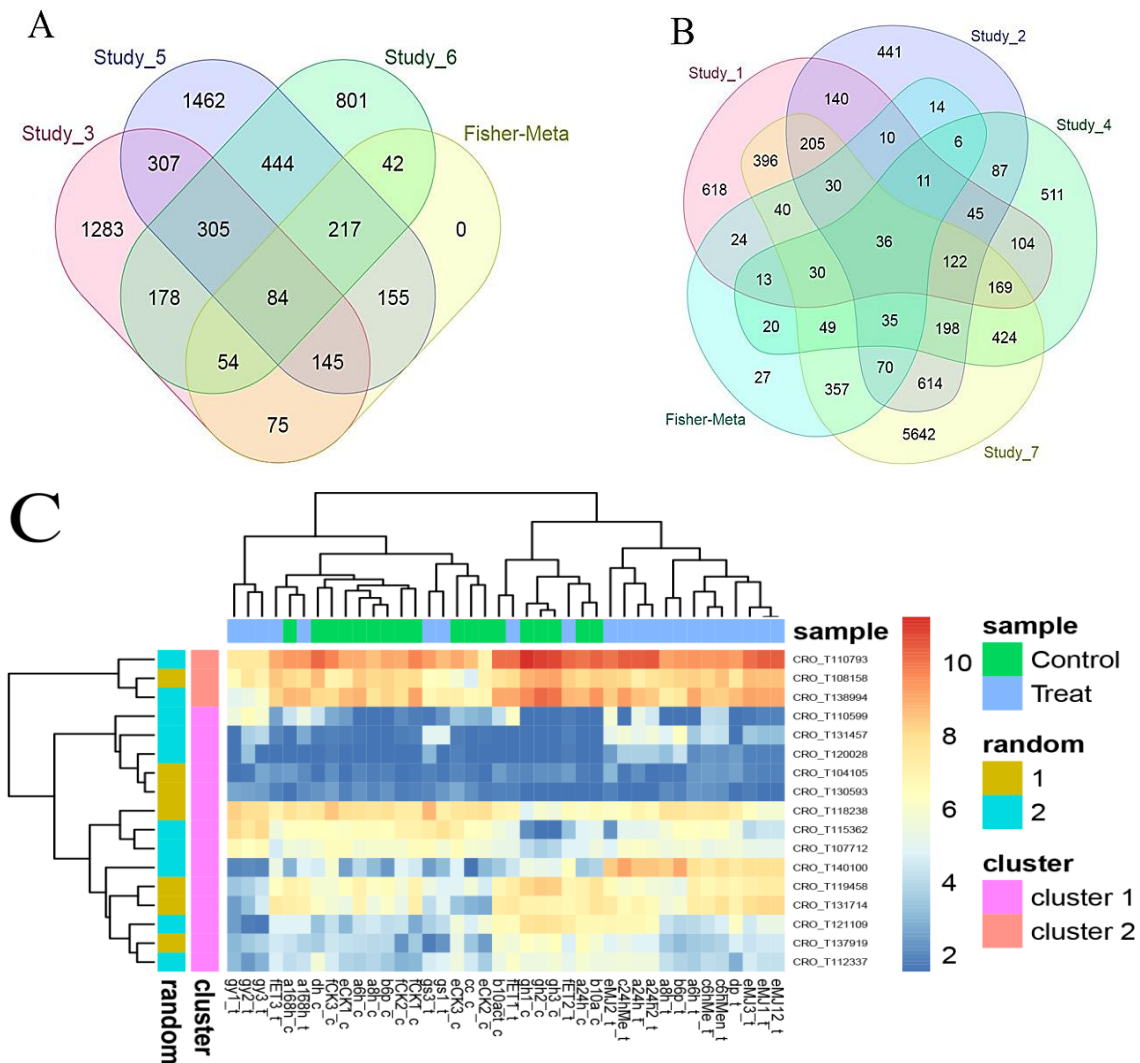
*HSPRO2* یک پروتئین حاوی LRR شبه *HsIpro-I* از آراییدوپسیس است، پروتئینی که مقاومت به نماتد سیست چغندر را ایجاد می‌کند. این پروتئین آراییدوپسیس در مقاومت پایه و در پاسخ دفاعی به باکتری (*Syringae pseudomonas*) نقش مهمی دارد و در برابر تنش اکسیداتیو و اسید سالیسیلیک پاسخ نشان می‌دهد (Murray et al. 2007).

علاوه بر ژن‌های مرتبط با تنش‌ها، ۱۷ ژن با بیان‌های مختلف در بین این DEGها در مسیرهای مربوط به بیوستنز TIA یافت شد که در نقشه حرارتی نشان داده شده است (شکل ۳C) و در بین آن‌ها ۴ ژن کدکننده آنزیم‌های مسیر ایندول آلکالوئید که مسیر انتهایی تولید وین‌کریستین و وین‌بلاستین است، شناسایی شد. این ژن‌ها *CRO\_T115362*، *CRO\_T120028*، *CRO\_T110599*، *CRO\_T104105* بودند که به ترتیب کدکننده آنزیم‌های 16-O-ec:2.1.1.94 - 16-hydroxylase.methyltransferase ec:1.14.14.103 - 16-O-acetyltransferase ec:4.3.3.2 - synthase بودند (شکل ۶).

#### هستی‌شناسی ژن (GO) و تجزیه و تحلیل KEGG متاژن‌ها: هستی‌شناسی ژن برای اختصاص اصطلاحات

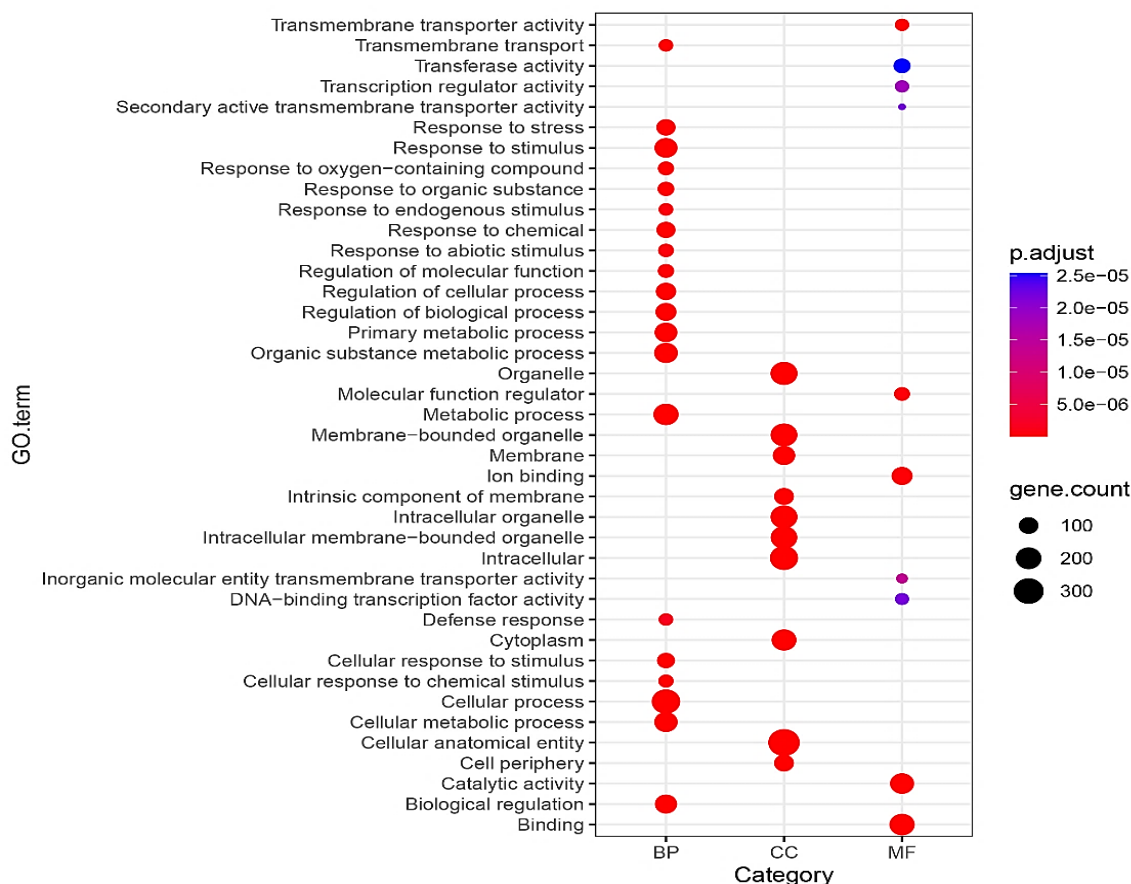
ویژه مربوط به آن (GO Term) روی کل DEGها به منظور شناسایی و طبقه‌بندی عملکردهای آن‌ها انجام شد، که در دسته فرآیندهای زیستی، فرآیند سلولی ((Cellular process (GO:0009987)، پاسخ به محرک‌ها (Response to stimulus (GO:0050896))، فرآیند متابولیکی ((Metabolic process (GO:0008152) و در دسته اجزای سلولی، موجودیت آناتومیک سلولی ((Cellular anatomical entity (GO:0110165)، درون سلولی ((Intracellular (GO:0005622)، اندامک ((Organelle (GO:0043226)، مهم‌ترین اصطلاحات GO غنی‌شده در بین DEGها بودند. در دسته عملکرد مولکولی، فعالیت کاتالیتیک ((Catalytic activity (GO:0003824)، اتصال ((Binding (GO:0005488) و ناقل ترانس ممبران ((Transmembrane transporter activity (GO:0022857) غالب بودند (شکل ۴).

این نتایج نشان می‌دهد که تعداد قابل توجهی از ژن‌ها و آنزیم‌ها در جریان تولید متابولیت‌ها شرکت می‌کنند. گیاهان در پاسخ به تنش متابولیسم خود را در سطح ژنتیکی تغییر می‌دهند که بر سنتز متابولیت‌ها تأثیر می‌گذارد. آن‌ها با تولید متابولیت‌های ثانویه از طریق انواع مسیرهای متابولیک به این شرایط پاسخ می‌دهند (Banothu and Uma 2022). نشان داده شد که برخی از مسیرهای متابولیک مهم، که طی آن پروتئین‌ها و متابولیت‌ها تنظیم می‌شوند، مسئول افزایش مکانیسم‌های دفاعی در هنگام تنش هستند. بهبود تحمل در پروانش و سایر گونه‌های گیاهی می‌تواند تا حدی به افزایش برخی فرآیندهای متابولیک توسط تنش بیشتر نسبت داده شود. تجزیه و تحلیل بیان ژن نشان داده که مجموعه‌ای از DEGها که در پاسخ‌های تنظیمی و دفاعی دخیل هستند، می‌توانند به پاسخ به تنش و تنظیم متابولیسم ثانویه کمک کنند (Verma et al. 2014; Gao et al. 2016; Li et al. 2019; Jia et al. 2020).



شکل ۳. (A و B) نمودار ون تعداد ژن‌های بیان شده افتراقی در مطالعات تکی و فراواکاوی (به دلیل پیچیدگی در قالب دو ون دیاگرام نشان داده شده است). (C) نقشه حرارتی DEGs فراواکاوی شناسایی شده در مسیرهای TIA

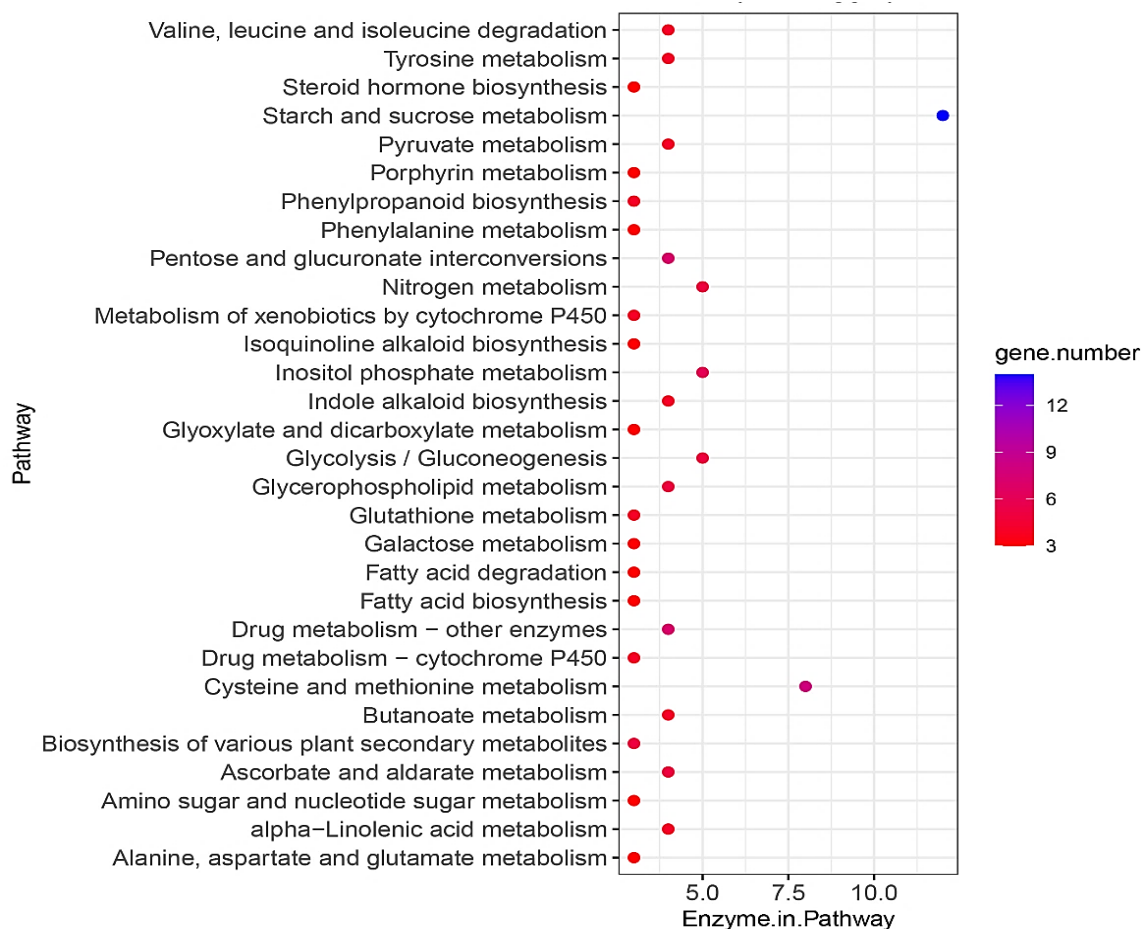
Figure 3. (A, B) Venn diagram showing the number of differential expressed genes from individual studies and meta-analysis. (C) Heatmap of the meta-analysis DEGs identified in the TIA pathways



شکل ۴. تجزیه و تحلیل هستی شناسی ژن (GO) مرتبط با DEGs های پروانش تحت برخی تنش‌ها

Figure 4. Gene ontology (GO) analysis associated with DEGs in *C. roseus* under some stresses

تجزیه و تحلیل مسیر KEGG برای DEGs ها، غنی‌سازی قابل توجهی را در مسیرها نشان داد. مهم‌ترین مسیرهای به دست آمده از پایگاه دیوید بر اساس تعداد ژن‌ها عبارت بودند از: «مسیر سیگنالینگ MAPK»، «بیوستز متابولیت‌های ثانویه»، «انتقال سیگنال هورمون گیاهی»، «مسیرهای متابولیک»، «متابولیسم نشاسته و ساکارز» و پنج مسیر مهمتر حاصل از Blast2GO بر اساس تعداد آنزیم‌های موجود در مسیر متابولیسم نشاسته و ساکارز، متابولیسم سیستین و متیونین، متابولیسم اینوزیتول فسفات، متابولیسم گلیکولیز / گلوکونوزن و متابولیسم نیتروژن بودند (شکل ۵).



شکل ۵. مسیرهای KEGG مرتبط با DEG ها در پروانش تحت برخی از تنش‌ها

Figure 5. KEGG pathways associated with DEGs in *C. roseus* under some of stresses

تعداد زیادی از ژن‌ها به مسیرهای متابولیک و به طور خاص بیوسنتز متابولیت‌های ثانویه اختصاص داده شدند که نشان‌دهنده بازآرایی قابل توجهی در متابولیسم گیاه به عنوان بخشی از سازگاری با تنش است. متابولیت‌های ثانویه در غلبه بر تنش‌های محیطی و عملکردهای دفاعی در پاسخ به تنش مهم هستند (Yeshi et al. 2022). مسیر متابولیسم نشاسته و ساکارز نیز قابل توجه بوده است. نشاسته منبع اولیه انرژی برای گیاهان است و همچنین در سیستم پاسخ به تنش نقش دارد. محصولات میانی، مانند مواد تنظیم کننده اسمزی و مولکول‌های سیگنال دهنده، ممکن است با سایر مسیرهای سیگنالینگ برای اصلاح پاسخ‌های پایین دست تعامل داشته باشند (Chen et al. 2020). همچنین، در غنی‌سازی، مسیر گلیکولیز/گلوکونئوزن شناسایی شد که نشان داده شده در تنش نقش دارد. ثابت شده است که تنش بیان برخی از آنزیم‌های مهم درگیر در متابولیسم گلیکولیز/گلوکونئوزن را تغییر می‌دهد (Zhu et al. 2019). فرآیند گلیکولیز شامل اکسیداسیون گلوکز (تولید شده از نشاسته و ساکارز) به پیروات است. سپس چرخه TCA این پیروات را برای تولید احیاکننده‌ها دکربوکسیله می‌کند (Bandehagh and Taylor 2020). چندین واسطه چرخه TCA در تولید

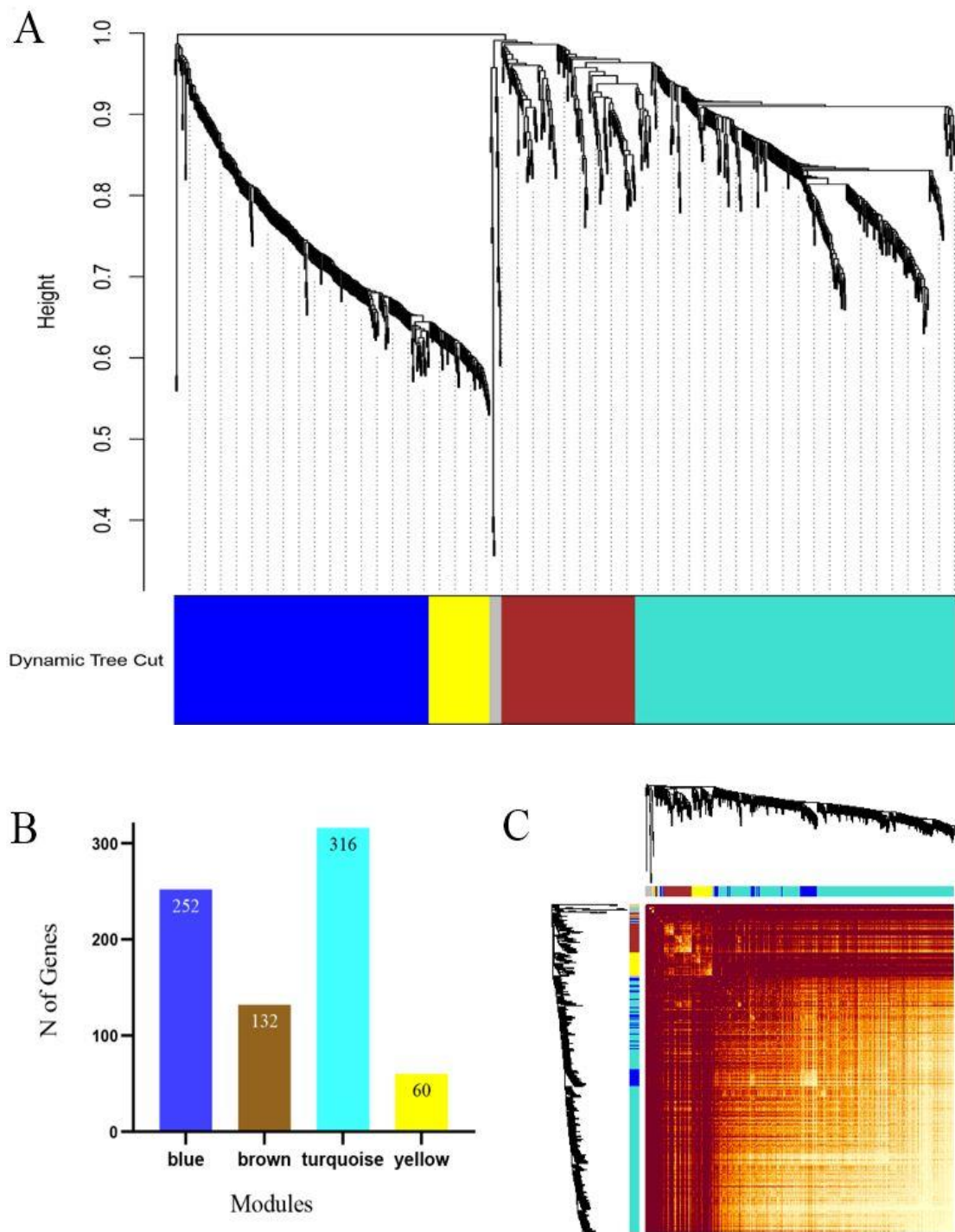
برخی آلكالوئیدها نقش دارند (Araújo et al. 2014). علاوه بر این به نظر می‌رسد مسیرهای دیگری که غنی‌سازی شده‌اند، در پاسخ به تنش مهم باشند. از جمله این مسیرها متابولیسم سیستئین و متیونین، متابولیسم اینوزیتول فسفات، و متابولیسم نیتروژن بوده که به عنوان پیش‌ساز چندین متابولیت ثانویه مهم برای رشد گیاه عمل می‌کنند.

در بین DEG‌های حاصل از فراواکاوی، ۹ ژن مرتبط با مسیرهای بیوسنتزی TIA‌ها شناسایی شد که سهم مسیرهای بیوسنتز ستون اصلی ترپنوئیدی [KO00900] Terpenoid backbone biosynthesis، ژن *CRO\_T108158* کدکننده آنزیم دلتاایزومراز (Delta-isomerase - ec:5.3.3.2)، بیوسنتز مونوترپنوئیدی Monoterpenoid biosynthesis [KO00902]، ژن‌های *CRO\_T119458* کدکننده آنزیم برنیل پیروفسفات هیدرولاز (bornyl - ec:3.1.7.3) و *CRO\_T131714* کدکننده آنزیم اسید گلوکزیل ترنسفرز (pyrophosphate hydrolase - ec:2.4.1.323) و *CRO\_T112337* کدکننده آنزیم ترنس‌آمیناز (transaminase - ec:2.6.1.5) و *CRO\_T137919* کدکننده آنزیم اکسیداز (oxidase - ec:1.4.3.2) و بیوسنتز ایندول آلكالوئید [KO00901] Indole، ژن‌های *CRO\_T104105* کدکننده آنزیم ۱۶-متیل ترنسفرز (16-O- - ec:2.1.1.94) alkaloid biosynthesis، *CRO\_T115362* کدکننده آنزیم سینتاز (synthase - ec:4.3.3.2)، *CRO\_T120028* کدکننده آنزیم اُ-استیل ترنسفرز (O-acetyltransferase - ec:2.3.1.107) و *CRO\_T110599* کدکننده آنزیم ۱۶-هیدروکسیلاز (16-hydroxylase - ec:1.14.14.103)، بود. این آنزیم‌ها توسط ژن‌های متعدد کدگذاری می‌شوند و نشان داده شده که فراواکاوی ژن‌های بیشتری را نسبت به مطالعات فردی شناسایی می‌کند (Deelen et al. 2021; Wu et al. 2021). مسیر بیوسنتز ایندول آلكالوئیدها به عنوان مسیر نهایی درگیر در تولید وین کریستین و وین بلاستین در شکل ۶ نشان داده شده است.

**تجزیه و تحلیل هم‌بیانی:** به منظور شناسایی ژن‌هایی که بیشترین همبستگی را در بین تمام DEG‌ها داشتند، WGCNA با رویکرد زیست‌شناسی سیستمی در جهت شناسایی فرآیندها و عملکردهای دخیل در تنش‌ها و سنتز متابولیت‌ها، استفاده شد. با توجه به الگوریتم برش درخت پویا، DEG‌ها به ۴ ماژول گروه‌بندی شدند که هر کدام دارای ۶۰ تا ۳۱۶ ژن بودند (شکل ۷A, B). این ماژول‌ها شامل فیروزه‌ای (۳۱۶ ژن)، آبی (۲۵۲ ژن)، و قهوه‌ای (۱۳۲ ژن) و زرد (۶۰ ژن) بودند. شکل ۷C خوشه‌بندی سلسله مراتبی ماتریس هم‌پوشانی توپولوژیکی (TOM) DEG‌ها را نشان می‌دهد. تجزیه و تحلیل غنی‌سازی برای دسته فرآیند زیستی به منظور درک فعالیت‌های زیستی مرتبط با ماژول‌ها انجام شد. ماژول‌های فیروزه‌ای و قهوه‌ای به ترتیب با ۸۹ و ۶۳ ژن ارتباط معنی‌داری با پاسخ به تنش‌های مختلف و پاسخ‌های دفاعی گیاه داشتند. از جمله این فرآیندهای غنی شده در ماژول فیروزه‌ای پاسخ دفاعی به باکتری (GO:0042742)، پاسخ به اوومیسیت‌ها (GO:GO:0002239)، پاسخ به اسیدسالیسیلیک (GO:0009751)، تنظیم پاسخ دفاعی (GO:0031347)، پاسخ دفاعی به قارچ (GO:0050832) و در ماژول قهوه‌ای تنظیم







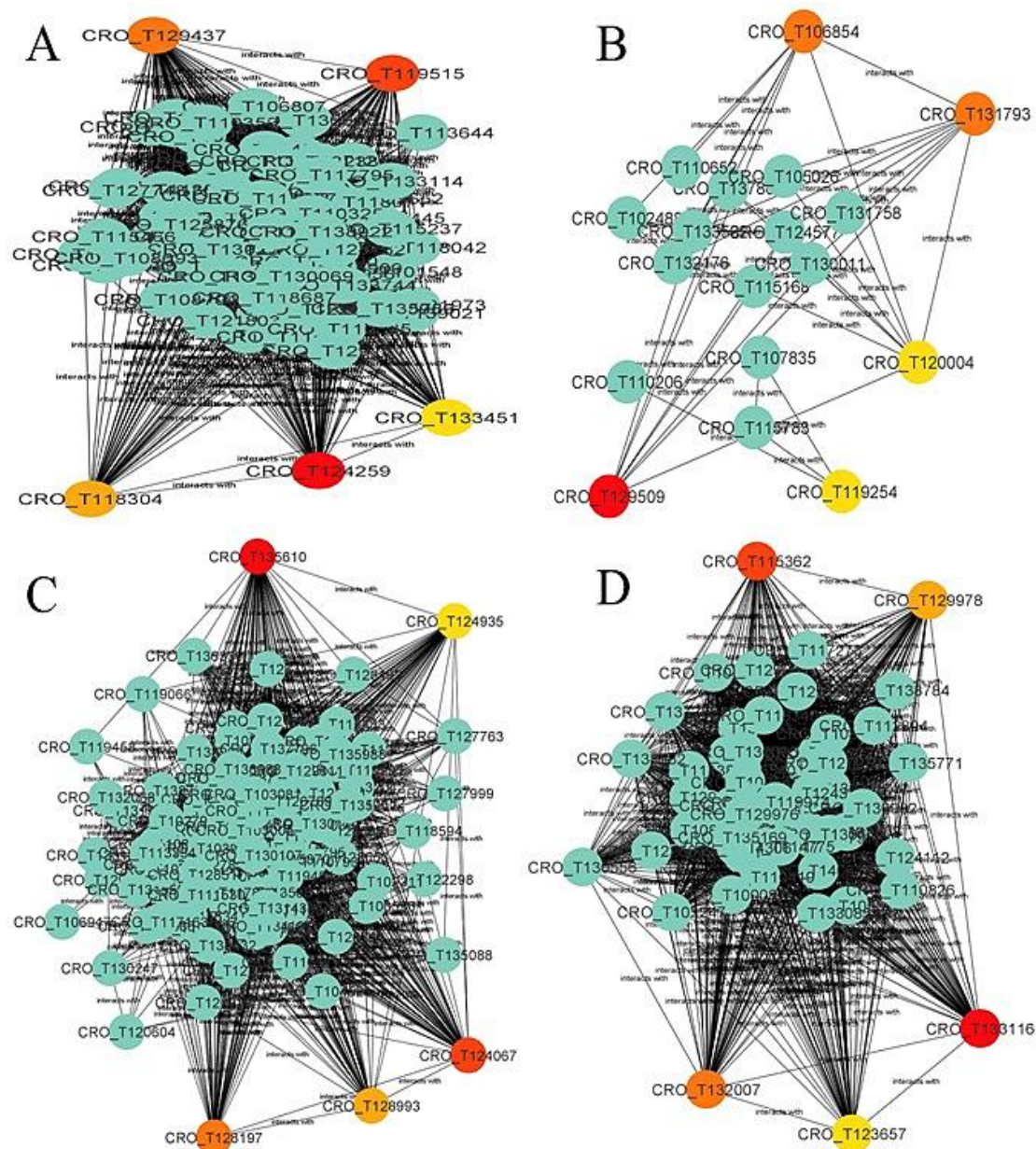
شکل ۷. تجزیه و تحلیل شبکه هم‌بیان وزنی ژن (A) درخت خوشه‌ای سلسله مراتبی DEGs و (B) تعداد ژن‌ها در هر ماژول (C) طرح نقشه حرارتی TOM شبکه

**Figure 7. Weighted gene co-expression network analysis (A) Hierarchical cluster tree of the DEGs and (B) number of genes in each module. (C) Network TOM heatmap plot**

ژن‌هایی با همبستگی‌های قوی در هر ماژول، که به عنوان ژن هاب شناخته می‌شوند، ژن‌هایی هستند که نقش‌های اساسی در شبکه ایفا می‌کنند. برای هر ماژول، پنج ژن هاب شناسایی شد. ژن‌های هاب ماژول‌های اصلی یعنی فیروزه‌ای شامل *CRO\_T128993*، *(LNK2) CRO\_T128197*، *(SMUBP2) CRO\_T124067*، *(PPDK) CRO\_T135610* شامل *(ALP1) CRO\_T129509* و ماژول قهوه‌ای شامل *(Pyruvate kinase isozyme G) CRO\_T124935*، *(LNK1) CRO\_T129509* و ماژول قهوه‌ای شامل *(ALP1) CRO\_T129509* و *(At1g07160) CRO\_T120004*، *CRO\_T106854*، *(MZN1.16) RO\_T131793* و *CRO\_T119254* بودند. شبکه ماژول‌ها و ژن‌های هاب در شکل ۸ نشان داده شده است.

ژن‌های *(PPDK) CRO\_T135610* و *(SMUBP2) CRO\_T124067* در ماژول فیروزه‌ای به ترتیب با امتیاز اتصال ۳۴۳ و ۳۲۴ بالاترین اتصال را در بین ژن‌های هاب این دو ماژول داشتند. پیرووات فسفات دیکناز (*PPDK*) آنزیمی است که تبدیل متقابل آدنوزین تری فسفات (*ATP*)، فسفات (*Pi*) و پیرووات را به آدنین مونوفسفات (*AMP*)، پیروفسفات (*PPi*) و فسفوانول پیرووات (*PEP*) را در حضور یون‌های منیزیم و پتاسیم/سدیم ( $K^+/Na^+$  و  $Mg^{2+}$ ) کاتالیز می‌کند. ثابت شده که *PPDK* نقش کلیدی در سازوکارهای دفاعی در برابر تنش‌های غیرزیستی مختلف، مانند خشکی، شوری، فلزات سنگین، غرقابی، اکسیژن کم و سرما، و تنش زیستی، از جمله آلودگی ویروسی دارد (Wang et al. 2018; Yadav et al. 2020).

پیش‌بینی شده پروتئین اتصال به DNA کدگذاری شده با *SMUBP-2* توسط *novel\_mir\_4* و *novel\_mir\_147* همولوگ فاکتور ۲۹ مرتبط با SAGA را هدف قرار می‌دهد. این همولوگ یک جزء تنظیم‌کننده کروماتین از مجموعه استیل‌اسیون هیستون و مربوط به تنظیم رونویسی است. *SMUBP-2* هم تنظیم‌کننده رونویسی است که دارای فعالیت هلیکازی ۵ تا ۳ است. در چمن *Brachypodium distachyon*، مورد هدف قرار گرفتن *SMUBP-2* توسط *novel-mir\_147* در ارتباط با تنش  $H_2O_2$ ، رخ داد (Lv et al. 2016). در گیاهان این ژن کمتر مورد مطالعه قرار گرفته است اما در جانوران می‌تواند در متابولیسم RNA، بیوزنر ریپوزوم یا شروع ترجمه، تنظیم رونویسی و ایمنی نقش داشته باشد (De Planell-Saguer et al. 2009). در ماژول قهوه‌ای هم *(ALP1) CRO\_T129509* و Zinc finger CCCH domain-containing MZN (Antagonist of Like Heterochromatin Protein1) *ALP1* به ترتیب بالاترین امتیاز اتصال را داشتند. ژن *ALP1* (Antagonist of Like Heterochromatin Protein1) کدکننده پروتئین مشتق شده از ترانسپوزاز (*transposase*) بوده که ممکن است دارای فعالیت نوکلئازی باشد. یکی از ویژگی‌های قابل توجه ترانسپوزون‌ها این است که اغلب در هنگام تنش فعال می‌شوند؛ برای مثال، در آرابیدوپسیس چندین رتروترانسپوزون با تیمار شوک حرارتی فعال می‌شوند (Cavrak et al. 2014) و نشان داده شده اهداف *ALP1* برای ژن‌های دخیل در پاسخ تنش زیستی و غیرزیستی غنی شده‌اند (Lianget al. 2015).



شکل ۸. شبکه اتصال درون ماژولی و ژن‌های هاب ماژول‌های آبی (A)، قهوه‌ای (B)، فیروزه‌ای (C) و زرد (D). ژن‌های هاب با رنگ‌های زرد، نارنجی یا قرمز نشان داده شده‌اند که هر چه امتیاز ارتباط بالاتر باشد رنگ‌ها تیره‌تر و به قرمز نزدیک می‌شود

Figure 8. Intramodular connectivity and hub genes of blue (A), brown (B), turquoise (C) and yellow (D) modules. Hub genes highlighted in yellow, orange, or red colors have been demonstrated, and as the correlation score increases, the colors become darker, leaning more towards red

ژن *MZN* از ژن‌های خانواده *CCCH*، پروتئین‌های انگشت روی حاوی موتیف را با سه سیستین و یک هیستیدین کد می‌کنند. آن‌ها نقش مهمی در پردازش *RNA* به عنوان پروتئین‌های اتصال‌دهنده *RNA* در حیوانات دارند. پروتئین‌های این خانواده می‌تواند در فعالیت‌های مختلف از جمله رشد و نمو گیاه و پاسخ به تنش‌ها و در واقع در مقاومت آن‌ها به تنش مؤثر باشند. مطالعات بیان ژن در برنج تأیید کرده که پروتئین‌های *CCCH* الگوهای بیان مختلفی را داشته که نشان‌دهنده عملکردهای متنوعی است (Wang et al. 2008). بررسی الگوی بیان ثابت کرده که اکثر اعضای این زیرخانواده توسط تنش‌های غیرزیستی یا زیستی تنظیم می‌شوند و بنابراین آن‌ها می‌توانند نقش مؤثری در تحمل تنش داشته باشند (Han et al. 2021). در مطالعات قبلی به ارتباط این ژن‌ها با تولید متابولیت‌های ثانویه گیاهی به ندرت پرداخته شده یا اصلاً مورد بررسی قرار نگرفته‌اند. در بین ژن‌های هاب، برخی همچون *CRO\_T119254* قبلاً شناسایی نشده‌اند و بنابراین عملکرد مشخصی برای آن‌ها یافت نشد. از آنجا که ماژول‌های فیروزه‌ای و قهوه‌ای ارتباط با پاسخ به تنش و عملکرد دفاعی، را نشان دادند و به ترتیب بیشترین ژن مرتبط با مسیر بیوستیزی *TIA*ها را داشتند بنابراین ژن‌های هاب شناسایی شده می‌توانند به‌عنوان کاندیدای مطالعات ژنتیکی به دلیل درجه بالایی از اتصال، درون و در میان ماماژول‌های غنی‌شده برای عملکردهای پاسخ تنش و سنتز متابولیت‌های ثانویه به خصوص مواد ضدسرطان وین کریستین و وین‌بلاستین در پروانش به عنوان تنها منبع این متابولیت‌ها، عمل کنند. ارتباط بیانی مشترک گزارش شده در این‌جا می‌تواند برای استنباط بیشتر عملکردهای بالقوه برای آن ژن‌هایی که پروتئین‌هایی با فعالیت‌های ناشناخته در پروانش را کد می‌کنند مورد استفاده قرار گیرد. برخی از ژن‌های شناسایی شده در این مطالعه هرگز در گیاه پروانش مورد مطالعه قرار نگرفته و حتی در این گونه شناخته شده نیستند، که ممکن است ژن‌های مفیدی در تنظیم و تولید متابولیت‌های ثانویه باشند.

#### نتیجه‌گیری: در این مطالعه، از فراواکای و زیست‌شناسی سامانه‌ای محاسباتی برای کشف ژن‌های پاسخ‌دهنده به تنش‌های

زیستی و غیرزیستی در داده‌های *RNA-seq*، برای درک چگونگی واکنش گیاه پروانش به برخی تنش‌ها و متابولیت‌های ثانویه خاص این گیاه استفاده شد. در این مطالعه در میان ۷۷۲ متاژن شناسایی شده که اکثراً در فرآیندهای سلولی و متابولیکی درگیر بودند، علاوه بر شناسایی ژن‌های کلیدی چون ژن کدکننده پروتئین زیپ لوسین *ATHB-7*، فسفوریلاز *H*، فسفولیپاز *D*، *CRO\_T108501* و *CRO\_T110204* با بیشترین افزایش بیان، ۲۷ متاژن همچون ژن‌های *YMF19*، *DOF5.3* و *HSPRO2* که در آنالیزهای انفرادی به عنوان *DEG* شناسایی نشده بودند، در پاسخ به انواع تنش‌ها و محرک‌ها شناسایی شد. این ژن‌ها به عنوان کدکننده پروتئین‌های مرتبط با ساخت *ATP* و نواحی تنظیم‌کننده رونویسی همچون ناحیه انگشت روی، تحت تنش‌های مختلف از جمله فلزات افزایش بیان یافته و به عنوان عامل رونویسی در پاسخ به تنش‌ها، محرک‌ها، مقاومت پایه و در پاسخ دفاعی به باکتری، تنش اکسیداتیو و اسید سالیسیلیک نقش مهمی دارند. همچنین با تجزیه و تحلیل شبکه هم‌بیان، ژن‌های هاب *PPDK*، *SMUBP2* و *ALP1* و *MZN* در ماژول‌های مهم غنی‌سازی شده در پاسخ به تنش‌ها و مسیرهای *TIA* شناسایی شدند که به عنوان ژن‌های مرکزی در این ماژول‌ها بیشترین ارتباط ممکن را با دیگر ژن‌ها دارند و تغییرات بیان آن‌ها با تأثیر بر دیگر ژن‌های ماژول می‌تواند ارتباط مستقیمی با پاسخ به تنش‌ها و سنتز متابولیت‌ها داشته باشد و از آنجا که این ژن‌های هاب،

ماژول‌های هم‌بیان قرار دارند هرچه بیان بالاتری داشته باشند، دیگر ژن‌های ماژول هم می‌توانند افزایش بیان نشان دهند. با توجه به عملکرد زیستی این ژن‌ها در فعالیت‌هایی چون بیان آنزیم‌های مهم و عوامل رونویسی مؤثر، می‌توانند نقش کلیدی در مکانیسم‌های دفاعی در برابر تنش‌ها، متابولیسم RNA، ترجمه، تنظیم رونویسی و ایمنی داشته و همچنین در تولید متابولیت‌ها مؤثر باشند. بنابراین ژن‌های مذکور به عنوان یک منبع مفید برای تحقیقات مولکولی و همچنین ژن‌های کاندیدای بالقوه برای مهندسی ژنتیک و متابولیک در گیاه پروانش پیشنهاد می‌شوند.

**سپاسگزاری:** از مرکز محاسبات سریع دانشگاه شهرکرد و پژوهشکده بیوتکنولوژی دانشگاه شیراز که ما را در پیشبرد این مطالعه یاری کردند و همچنین از سردبیر و داوران محترم به خاطر مطالعه متن مقاله حاضر و ارائه نظرها و پیشنهادهای ارز شمند سپاسگزاری می‌شود.

## منابع

- عرب پور رق آبادی زهرا، محمدآبادی محمدرضا، خضری امین (۱۴۰۰) الگوی بیانی ژن *p32* در بافت‌های ران، دست، راسته و چربی پشت بره کرمانی. مجله بیوتکنولوژی کشاورزی ۱۳(۴)، ۲۰۰-۱۸۳.
- محمدآبادی محمدرضا، کرد محبوبه، نظری محمود (۱۳۹۷) مطالعه بیان ژن لپتین در بافت‌های مختلف گوسفند کرمانی با استفاده از real time PCR. مجله بیوتکنولوژی کشاورزی ۱۰(۳)، ۱۲۲-۱۱۱.

## References

- Al-Khayri JM, Rashmi R, Toppo V, et al (2023) Plant secondary metabolites: the weapons for biotic stress management. *Metabolites* 13, 716-753.
- Andrews S (2010) Fastqc - a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Babraham Bioinforma <http://www.bioinformatics.babraham.ac.uk/projects>.
- Arabpoor Raghabadi Z, Mohammadabadi M, Khezri A (2022) The expression pattern of *p32* gene in femur, humeral muscle, back muscle and back fat tissues of Kermani lambs. *Agric Biotech J* 13, 183-200 (in Persian).
- Araújo WL, Martins AO, Fernie AR, Tohge T (2014) 2-oxoglutarate: linking TCA cycle function with amino acid, glucosinolate, flavonoid, alkaloid, and gibberellin biosynthesis. *Front Plant Sci* 5, 552-558.
- Bah SY, Forster T, Dickinson P, et al (2018) Meta-analysis identification of highly robust and differential immune-metabolic signatures of systemic host response to acute and latent tuberculosis in children and adults. *Front Genet* 9, 457-473.
- Bairakdar MD, Tewari A, Truttmann MC (2023) A meta-analysis of RNA-seq studies to identify

- novel genes that regulate aging. *Exp Gerontol* 173, 1-34.
- Balan B, Marra FP, Caruso T, Martinelli F (2018) Transcriptomic responses to biotic stresses in *Malus x domestica*: a meta-analysis study. *Sci Rep* 8, 1-12.
- Bandehagh A, Taylor NL (2020) Can alternative metabolic pathways and shunts overcome salinity induced inhibition of central carbon metabolism in crops?. *Front Plant Sci* 11, 1072-1090.
- Banothu V, Uma A (2022) Effect of Biotic and Abiotic Stresses on Plant Metabolic Pathways. In: Phenolic Compounds - Chemistry, Synthesis, Diversity, Non-Conventional Industrial, Pharmaceutical and Therapeutic Applications. Badria FA (ed). IntechOpen, London, UK, pp. 1-11.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinform* 30, 2114-2120.
- Buga AM, Docea AO, Albu C, et al (2019) Molecular and cellular stratagem of brain metastases associated with melanoma (review). *Oncol Lett* 17, 4170-4175.
- Cavrak VV, Lettner N, Jamge S, et al (2014) How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS Genet* 10, 1-12.
- Chen Y, Jiang Y, Chen Y, et al (2020) Uncovering candidate genes responsive to salt stress in *Salix matsudana* (Koidz) by transcriptomic analysis. *PLoS One* 15, 8-31.
- Chin CH, Chen SH, Wu HH, et al (2014) Cytohubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 8, 4-11.
- Conesa A, Götz S, García-Gómez JM, et al (2005) Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinform* 21, 3674-3676.
- De Abreu Neto JB, Frei M (2016) Microarray meta-analysis focused on the response of genes involved in redox homeostasis to diverse abiotic stresses in rice. *Front Plant Sci* 6, 1-14.
- De Planell-Saguer M, Schroeder DG, Rodicio MC, et al. (2009) Biochemical and genetic evidence for a role of IGHMBP2 in the translational machinery. *Hum Mol Genet* 18, 2115-2126.
- Deelen J, Evans DS, Arking DE, et al. (2021) Publisher correction: a meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat Commun* 12, 1-2.
- Dhyani P, Quispe C, Sharma E, et al. (2022) Anticancer potential of alkaloids: a key emphasis to colchicine, vinblastine, vincristine, vindesine, vinorelbine and vincamine. *Cancer Cell Int* 22, 1-20.
- Distéfano AM, Valiñas MA, Scuffi D, et al. (2015) Phospholipase D  $\delta$  knock-out mutants re tolerant to severe drought stress. *Plant Signal Behav* 10, 1-7.
- Dobin A, Davis CA, Schlesinger F, et al. (2013) Star: ultrafast universal RNA-seq aligner. *Bioinform* 29, 15-21.

- Gao C, Yang B, Zhang D, et al. (2016) Enhanced metabolic process to indole alkaloids in *Clematis terniflora* DC. after exposure to high level of UV-B irradiation followed by the dark. BMC Plant Biol 16, 1-15.
- Guo J, Huang Z, Sun J, et al. (2021) Research progress and future development trends in medicinal plant transcriptomics. Front Plant Sci 12, 1-10.
- Han G, Qiao Z, Li Y, et al. (2021) The roles of CCCH zinc-finger proteins in plant abiotic stress tolerance. Int J Mol Sci 22, 15-35.
- Jia Y, Li X, Liu Q, et al. (2020) Physiological and transcriptomic analyses reveal the roles of secondary metabolism in the adaptive responses of stylosanthes to manganese toxicity. BMC Genom 21, 1-17.
- Kanehisa M, Goto S (2000) Kegg: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28, 27-30.
- Langfelder P, Horvath S (2008) Wgcna: an R package for weighted correlation network analysis. BMC Bioinform 9, 1-13.
- Li S, Liu Z, Jia Y, et al. (2019) Analysis of metabolic pathways related to fertility restoration and identification of fertility candidate genes associated with *Aegilops kotschy* cytoplasm in wheat (*Triticum aestivum* L.). BMC Plant Biol 19, 1-17.
- Li Y, Yang Z, Zhang Y, et al. (2022) The roles of HD-ZIP proteins in plant abiotic stress tolerance. Front Plant Sci 13, 1-19.
- Liang SC, Hartwig B, Perera P, et al. (2015) Kicking against the PRCs – a domesticated transposase antagonises silencing mediated by Polycomb group proteins and is an accessory component of Polycomb repressive complex 2. PLoS Genet 11, 1-26.
- Liu H, Shi J, Wu M, Xu D (2021) The application and future prospect of RNA-seq technology in chinese medicinal plants. J Appl Res Med Aromat Plants 24, 1-17.
- Liu J, Gao F, Ren J, et al. (2017) A novel AP2/ERF transcription factor CR1 regulates the accumulation of vindoline and serpentine in *Catharanthus roseus*. Front Plant Sci 8, 1-11.
- Lv DW, Zhen S, Zhu GR, et al. (2016) High-throughput sequencing reveals H<sub>2</sub>O<sub>2</sub> stress-associated micro RNAs and a potential regulatory network in *Brachypodium distachyon* seedlings. Front Plant Sci 7, 1567-1587.
- Mohammadabadi M, Masoudzadeh SH, Khezri A, et al. (2021) Fennel (*Foeniculum vulgare*) seed powder increases delta-like non-canonical notch ligand 1 gene expression in testis, liver, and humeral muscle tissues of growing lambs. Heliyon 7(12), e08542.
- Mohammadabadi MR, Kord M, Nazari M (2018) Studying expression of *leptin* gene in different tissues of Kermani sheep using Real Time PCR. Agric Biotech J 10, 111-



122 (in Persian).

- Molina-Hidalgo FJ, Medina-Puche L, Cañete-Gómez C, et al. (2017) The fruit-specific transcription factor FaDOF2 regulates the production of eugenol in ripe fruit receptacles. *J Exp Bot* 68, 4529-4543.
- Murray SL, Ingle RA, Petersen LN, Denby KJ (2007) Basal resistance against *Pseudomonas syringae* in arabidopsis involves WRKY53 and a protein with homology to a nematode resistance protein. *Mol Plant-Microbe Interact* 20, 1431-1438.
- Pan Y jie, Lin Y chao, Yu B fan, et al. (2018) Transcriptomics comparison reveals the diversity of ethylene and methyl-jasmonate in roles of TIA metabolism in *Catharanthus roseus*. *BMC Genom* 19, 1-14.
- Pathania S, Bagler G, Ahuja PS (2016) Differential network analysis reveals evolutionary complexity in secondary metabolism of *Rauwolfia serpentina* over *Catharanthus roseus*. *Front Plant Sci* 7, 1-17.
- Paul I, Poddar Sarkar M, Bhadoria PBS (2022) Floral secondary metabolites in context of biotic and abiotic stress factors. *Chemoecology* 32, 49-68.
- Rau A, Marot G, Jaffrézic F (2014) Differential meta-analysis of RNA-seq data from multiple studies. *BMC Bioinf* 15, 1-14.
- Robinson M, McCarth D (2010) Edger: differential expression analysis of digital gene expression data. *Bioconductor Fhrc Org* 1-76.
- Shahsavari M, Mohammadabadi M, Khezri A, et al. (2021) Correlation between insulin-like growth factor 1 gene expression and fennel (*Foeniculum vulgare*) seed powder consumption in muscle of sheep. *Anim Biotechnol* 34(4), 882-892.
- Skiryecz A, Jozefczuk S, Stobiecki M, et al. (2007) Transcription factor AtDOF4;2 affects phenylpropanoid metabolism in *Arabidopsis thaliana*. *New Phytol* 175, 425-438.
- Soltani N, Firouzabadi FN, Shafeinia A, et al. (2022) De novo transcriptome assembly and differential expression analysis of *Catharanthus roseus* in response to salicylic acid. *Sci Rep* 12, 1-13.
- Tahmasebi A, Ashrafi-Dehkordi E, Shahriari AG, et al. (2019) Integrative meta-analysis of transcriptomic responses to abiotic stress in cotton. *Prog Biophys Mol Biol* 146, 112-122.
- Thalman M, Santelia D (2017) Starch as a determinant of plant fitness under abiotic stress. *New Phytol* 214, 943-951.
- Verma M, Ghangal R, Sharma R, et al. (2014) Transcriptome analysis of *Catharanthus roseus* for gene discovery and expression profiling. *PLoS One* 9, 7-18.
- Wang D, Guo Y, Wu C, et al. (2008) Genome-wide analysis of CCCH zinc finger family in *Arabidopsis* and rice. *BMC Genom* 9, 44-64.

- Wang H, Liu C, Ma P, et al. (2018) Functional characterization of cytosolic pyruvate phosphate dikinase gene (*MecyPPDK*) and promoter (*MecyPPDKP*) of cassava in response to abiotic stress in transgenic tobacco. *Crop Sci* 58, 2002-2009.
- Wu W, Howard D, Sibille E, French L (2021) Differential and spatial expression meta-analysis of genes identified in genome-wide association studies of depression. *Transl Psychiatry* 11, 1-12.
- Xia C, Hong L, Yang Y, et al. (2019) Protein changes in response to lead stress of lead-tolerant and lead-sensitive industrial hemp using swath technology. *Genes* 10, 1-16.
- Yadav S, Rathore MS, Mishra A (2020) The pyruvate-phosphate dikinase (*C<sub>4</sub>-smppdk*) gene from *Suaeda monoica* enhances photosynthesis, carbon assimilation, and abiotic stress tolerance in a *C<sub>3</sub>* plant under elevated CO<sub>2</sub> conditions. *Front Plant Sci* 11, 345-367.
- Yeshi K, Crayn D, Ritmejerytė E, Wangchuk P (2022) Plant secondary metabolites produced in response to abiotic stresses has potential application in pharmaceutical product development. *Mol* 27, 750-757.
- Zhang Y, Lin Z, Wang M, Lin H (2018) Selective usage of isozymes for stress response. *ACS Chem Biol* 13, 3059-3064.
- Zhang Y, Parmigiani G, Johnson WE (2020) Combat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* 2, 1-21.
- Zhu Y, Liu Q, Xu W, et al. (2019) De novo assembly and discovery of genes that involved in drought tolerance in the common vetch. *Int J Mol Sci* 20, 1-17.
- Zou X, Sun H (2023) Dof transcription factors: specific regulators of plant biological processes. *Front Plant Sci* 14, 1-13.