Tai et al. BMC Genomics (2018) 19:616
https://doi.org/10.1186/s12864-018-4999-9

BMC Genomics

RESEARCH ARTICLE Open Access

Gene co-expression network analysis reveals coordinated regulation of three characteristic secondary biosynthetic pathways in tea plant (*Camellia sinensis*)

Yuling Tai[1,2†], Chun Liu[3†], Shuwei Yu[2], Hua Yang[2], Jiameng Sun[1], Chunxiao Guo[1], Bei Huang[2], Zhaoye Liu[2], Yi Yuan[1], Enhua Xia[2], Chaoling Wei[2*] and Xiaochun Wan[2*]

Abstract

Background: The leaves of tea plants (*Camellia sinensis*) are used to produce tea, which is one of the most popular beverages consumed worldwide. The nutritional value and health benefits of tea are mainly related to three abundant characteristic metabolites; catechins, theanine and caffeine. Weighted gene co-expression network analysis (WGCNA) is a powerful system for investigating correlations between genes, identifying modules among highly correlated genes, and relating modules to phenotypic traits based on gene expression profiling. Currently, relatively little is known about the regulatory mechanisms and correlations between these three secondary metabolic pathways at the omics level in tea.

Results: In this study, levels of the three secondary metabolites in ten different tissues of tea plants were determined, 87,319 high-quality unigenes were assembled, and 55,607 differentially expressed genes (DEGs) were identified by pairwise comparison. The resultant co-expression network included 35 co-expression modules, of which 20 modules were significantly associated with the biosynthesis of catechins, theanine and caffeine. Furthermore, we identified several hub genes related to these three metabolic pathways, and analysed their regulatory relationships using RNA-Seq data. The results showed that these hub genes are regulated by genes involved in all three metabolic pathways, and they regulate the biosynthesis of all three metabolites. It is notable that light was identified as an important regulator for the biosynthesis of catechins.

Conclusion: Our integrated omics-level WGCNA analysis provides novel insights into the potential regulatory mechanisms of catechins, theanine and caffeine metabolism, and the identified hub genes provide an important reference for further research on the molecular biology of tea plants.

Keywords: Tea plant, Co-expression network analysis, Characteristic metabolites, RNA-seq, Secondary metabolic pathway

Background

Tea, produced from the leaves of the tea plant, *Camellia sinensis* (L.), belonging to family *Theaceae*, is one of the most popular natural non-alcoholic beverages consumed worldwide. To date, nearly 4000 bioactive compounds have been identified in tea [1] including catechins, caffeine, theanine and volatile oils [2]. Catechins generally contain six different monomers, namely catechin (C), gallocatechin (GC), epicatechin (EC), epigallocatechin (EGC), epicatechin gallate (ECG) and epigallocatechin gallate (EGCG) [3]. Catechins, caffeine and theanine are the main three characteristic biologically active compounds in tea [4]. They are not only important contributors to flavour, but also have beneficial effects on human health due to their autoxidation and anticancer activity [5] and their ability to lower blood pressure [6], prevent cardiovascular diseases [7], and assist weight loss [8].

Gene co-expression network analysis (GCNA) is a genetic approach for analysing correlations between genes using large-scale gene expression profiling data that is

* Correspondence: weichl@ahau.edu.cn; xcwan@ahau.edu.cn
†Yuling Tai and Chun Liu contributed equally to this work.
[2]State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei 230036, China
Full list of author information is available at the end of the article

© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

especially useful for investigating relationships between functional modules and phenotypic traits [9, 10]. Weighted GCNA (WGCNA) is one of the most popular GCNA-based approaches, and this correlation-based technique describes and visualises co-expression networks between genes using transcriptomic data [11]. This technique has been successfully utilized to identify the gene modules in *Arabidopsis* and rice that are related to drought and bacterial stress [12]. Module assignment in WGCNA is a flexible process which reduces the complexity of a dataset from hundreds of genes to a smaller number of modules.

Researchers have focused on the molecular mechanisms involved in plant growth, development [13, 14] and the production of secondary metabolites [15] in tea plants. Regulatory mechanisms underlying secondary metabolite biosynthesis, particularly those related to catechins, theanine and caffeine, have been explored at the molecular level. Recent advances in next-generation sequencing of RNA [16] have been accompanied by an increase in the amount of available transcriptomic data from different tissues of tea plants [17], from different species of the genus *Camellia* [18], and from plants grown under different stress conditions [19, 20]. Most research has focused on using RNA-Seq data from tea plants to reveal the regulatory mechanisms and relationships between gene expression and production of the characteristic secondary metabolites. Li et al. constructed a possible transcription factor regulation network of flavonoid, caffeine and theanine biosynthesis using 13 different samples from tea plants (various organs and developmental stages) through correlation analysis [21]. However, relatively few studies have investigated co-expression networks in tea plants using RNA-Seq data. In the present study, WGCNA was performed using RNA-Seq data from ten tissues, and modules significantly correlated with the three characteristic compounds were identified and analysed. Furthermore, highly-connected hub genes related to these modules were identified. This novel approach revealed the regulatory mechanisms of characteristic metabolic pathways in tea plants, and highlighted the important role of light in the biosynthesis of catechins, theanine and caffeine.

Results

Determination of catechins, theanine and caffeine content

High-performance liquid chromatograph (HPLC) analysis was used to determine the content of catechins (C, EC, GC, EGC, ECG and EGCG), theanine and caffeine in ten different tissues from *C. sinensis* cv. *Shuchazao*. (Fig. 1). The results indicated that accumulation of individual catechin compounds and caffeine varied between different tissues and seasons. The total amounts of catechins and caffeine were greater in tender shoots than other tissues, especially in buds and first leaves (> 200 mg/g). Galloylated catechins such as ECG and EGCG were the predominant characteristic phenolic compounds, and reached 48.3 and 122.5 mg/g, respectively. Similarly, the caffeine content was significantly higher in tender shoots, but variation between new and old shoots was less than for catechins. By contrast, total catechins and caffeine were much less abundant in roots than in other tissues (0.67 and 0.0029 mg/g, respectively). However, theanine, an important compound in tea, reached 40.8 mg/g in roots, which was 6-fold higher than in buds.

De novo assembly and functional annotation of the *C. sinensis* transcriptome

We obtained 111 gigabases of sequencing data (average 11 gigabase/sample) from ten samples. De novo assembly of the *C. sinensis* transcriptome was performed using the Trinity package [22], and 91,494 unigenes were obtained after removing short contigs (≤200 bp) and redundancy by TGICL [23]. In order to reduce potential assembly errors, we filtered unigenes with a Fragments Per Kilobase per Million mapped fragments (FPKM) value less than 0.3 in at least eight of the ten tissues [24]. Finally, a high-quality transcriptomic library of 87,319 unigenes with an N50 of 1406 bp and an average length of 874 bp was obtained. Functional annotation successfully aligned 54,827 (62.79%), 58,770 (67.30%), 40,700 (46.61%), 43,910 (50.29%), 25,172 (28.83%) unigenes to the NT (Non-redundant nucleotide database), NR (Non-redundant protein database), Swiss-Prot (Annotated protein sequence database), KEGG (Kyoto Encyclopaedia of Genes and Genomes), and GO (Gene Ontology) databases, respectively. Overall, 72.68% (63,464 of 87,319) of unigenes were annotated (Table 1).

The E-value distribution of the top hits in the NR, NT, KEGG and Swiss-Prot databases shows that an average of 51% of mapped sequences shared significant homology (< $1.0E^{-50}$), and nearly 41.99% of sequences shared greater than 80% similarity. These results confirmed the high quality of the assembled unigenes in the transcriptomic library (Fig. 2).

Analysis of gene expression and identification of differentially expressed genes (DEGs)

Clean reads from ten different *C. sinensis* tissues were mapped to the high-quality unigene sets using Bowtie2 [25]. Expression levels of unigenes were calculated in each tissue using FPKM values, and unigenes with an FPKM ≥0.3 [26] were defined as expressed. The number of genes expressed in each tissue ranged from 65,799 to 74,258, with an average FPKM of 16.47 (Fig. 3). A total of 55,607 DEGs were identified by in pairwise comparison for further analysis.

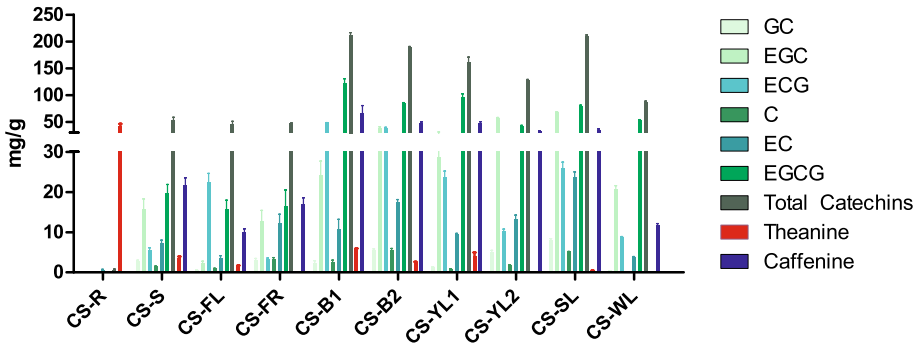


Fig. 1 Content of catechins, theanine and caffeine in different tissues of *Camellia sinensis*. Tissues include apical buds in May (CS-B1) and June (CS-B2), first young leaves (CS-YL1), second young leaves (CS-YL2) and mature leaves in summer (CS-SL), mature leaves in winter (CS-WL), stems (CS-S), flowers (CS-FL), fruits (CS_FR) and roots (CS-R)

Construction of the gene co-expression network

We constructed an unassigned network using DEGs identified by pairwise comparison from the ten tissues [27, 28]. All tissues were clustered initially without any outlier tissues (Additional file 1). A scale-free topology model with a soft threshold of 30 was used to ensure that the network was biologically relevant, resulting in 53,279 (95.81%) out of 55,607 unigenes parsed into 35 co-expression modules, with the module size ranging from 55 to 9041.

Identification of content-related modules

We investigated correlations between the characteristic components (C, GC, EC, EGC, ECG, EGCG, theanine and caffeine) of tea and the 35 co-expression modules. We identified 20 modules that were significantly ($p < 0.05$) correlated with characteristic components (content-related modules), with highly positive coefficients for modules associated with C (0.91), GC (0.87), EC (0.88), EGC (0.72), ECG (0.76), EGCG (0.85), theanine (0.97) and caffeine (0.9) coloured blue, blue, blue, blue, green, white, yellow and green, respectively. In addition, EGCG was significantly negatively correlated with the dark red module (– 0.68), while total catechins were negatively correlated with the light green module (– 0.88), turquoise module (– 0.71) and pale turquoise module (– 0.61). Importantly, there was only two modules (yellow and light cyan) significantly associated with theanine (Fig. 4 and Additional file 2).

Functional annotation and enrichment of content-related modules

To identify the biological roles of modules associated with catechins, theanine and caffeine, functional annotation and enrichment of these modules were analysed by KEGG pathway analysis. Detailed functional enrichment information from KEGG pathway annotation related to these modules (Q-value < 0.05) is shown in Fig. 5. According to KEGG pathway enrichment analysis, unigenes in content-related modules were enriched in different metabolic pathways, especially those related to characteristic metabolites in tea. For example, 'Biosynthesis of amino acids' was enriched in red modules with EGC, and 'Isoflavonoid biosynthesis', 'Anthocyanin biosynthesis', 'Flavonoid biosynthesis', 'Phenylpropanoid biosynthesis'

Table 1 Summary of sequence assembly and functional annotation

Level	Item	No. of sequences	Percentage (%)
Assembly	Total number of Unigenes	87,319	–
	Total bases (Mb)	76,328,365	–
	N50 (bp)	1406	–
	Average Unigene length (bp)	874	–
Annotation	NT-annotated	54,827	62.79
	NR-annotated	58,770	67.30
	SwissProt-annotated	40,700	46.61
	KEGG-annotated	43,910	50.29
	COG-annotated	22,252	25.48
	GO-annotated	25,172	28.83
	All annotated Unigenes	63,464	72.68

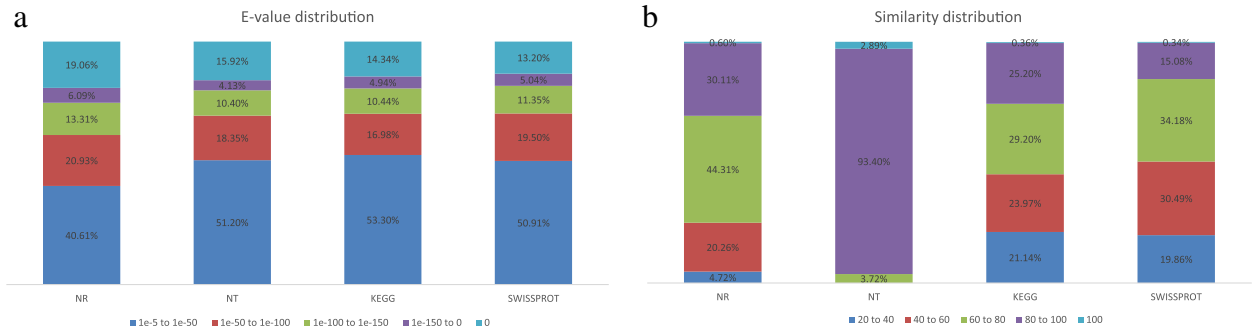


Fig. 2 E-value and sequence similarity analysis of unigenes against NR (Non-redundant protein), Nt (Non-redundant nucleotide), KEGG (Kyoto Encyclopedia of Genes and Genomes) and Swiss-Prot (Annotated protein sequence) databases. **a** E-value distribution of BLAST hits for each unigene with a cutoff E-value of 1.0E-5. **b** Similarity distribution of the top BLAST hits for each unigene

and 'Degradation of aromatic compounds' were enriched in the yellow and light cyan modules associated with theanine. 'Purine metabolism' and 'Tropane, piperidine and pyridine alkaloid biosynthesis' were enriched in the green and magenta modules associated with EGCG and caffeine. 'Nitrogen metabolism', 'Arginine biosynthesis' and 'Alanine, aspartate and glutamate metabolism' were enriched in the turquoise module associated with total catechins.

Functional analysis of unigenes related to photosynthesis

Unigenes in blue, red and magenta modules were significantly enriched in 'Photosynthesis' according to KEGG pathway enrichment analysis. Further functional analysis indicated that these unigenes were associated with ferredoxin, photosystem II, photosystem I, ferredoxin--NADP + reductase, F-type H + −transporting ATPase subunit delta, and some other categories (Additional file 3). Light is an important environmental parameter that drives photosynthesis, and photosynthesis might influence catechin biosynthesis via the provision of carbon sources.

Hub gene identification and visualisation

Hub genes in modules may be more important than other genes in the network, and they can be considered representative of the module in the biology network. Detailed information on all hub genes of each content-related module is listed in Additional file 4. Hub gene analysis identified WD40 repeat, ethylene-responsive transcription factor, MYB, WRKY and bHLH in the blue module. Heat shock proteins in the blue module were also identified, as were ABC transporters in yellow and light cyan modules. The green module contains flavonoid 3',5'-hydroxylase (*F3'5'H*), flavonol synthase (*FLS*) and beta-glucosidase (*βG*). The yellow module includes two glutamine synthetase (*GS*) genes,

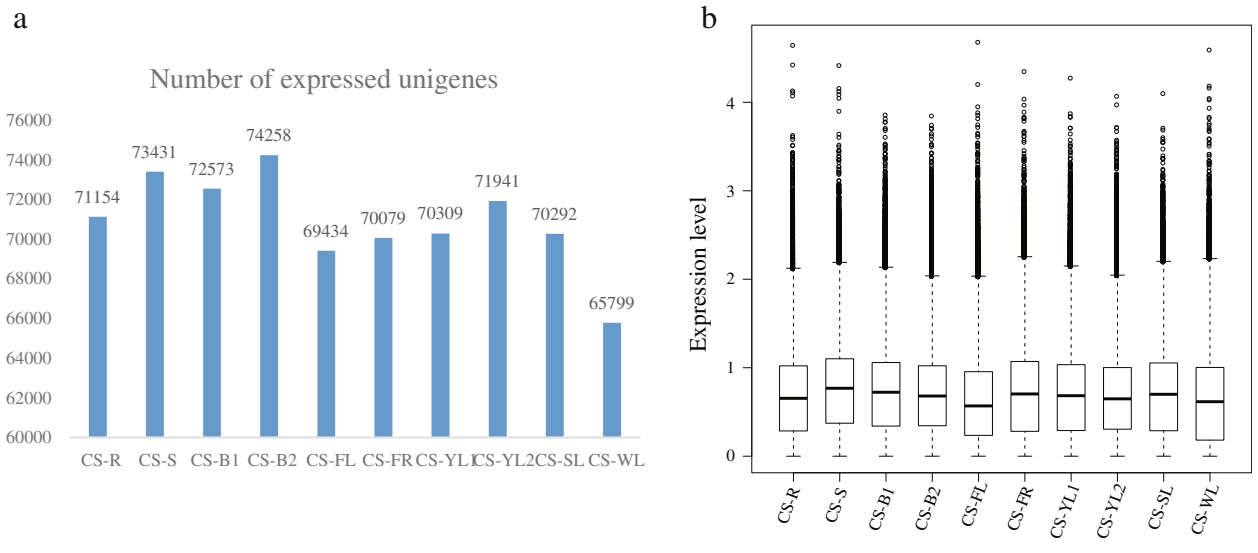


Fig. 3 Number of expressed unigenes and their expression levels in different tissues of *C. sinensis*. **a** Number of unigenes expressed in each tissue (given above bars). **b** Expression levels of unigenes in the ten tissues. Expression levels of all genes were evaluated using log10-transformed FPKM values

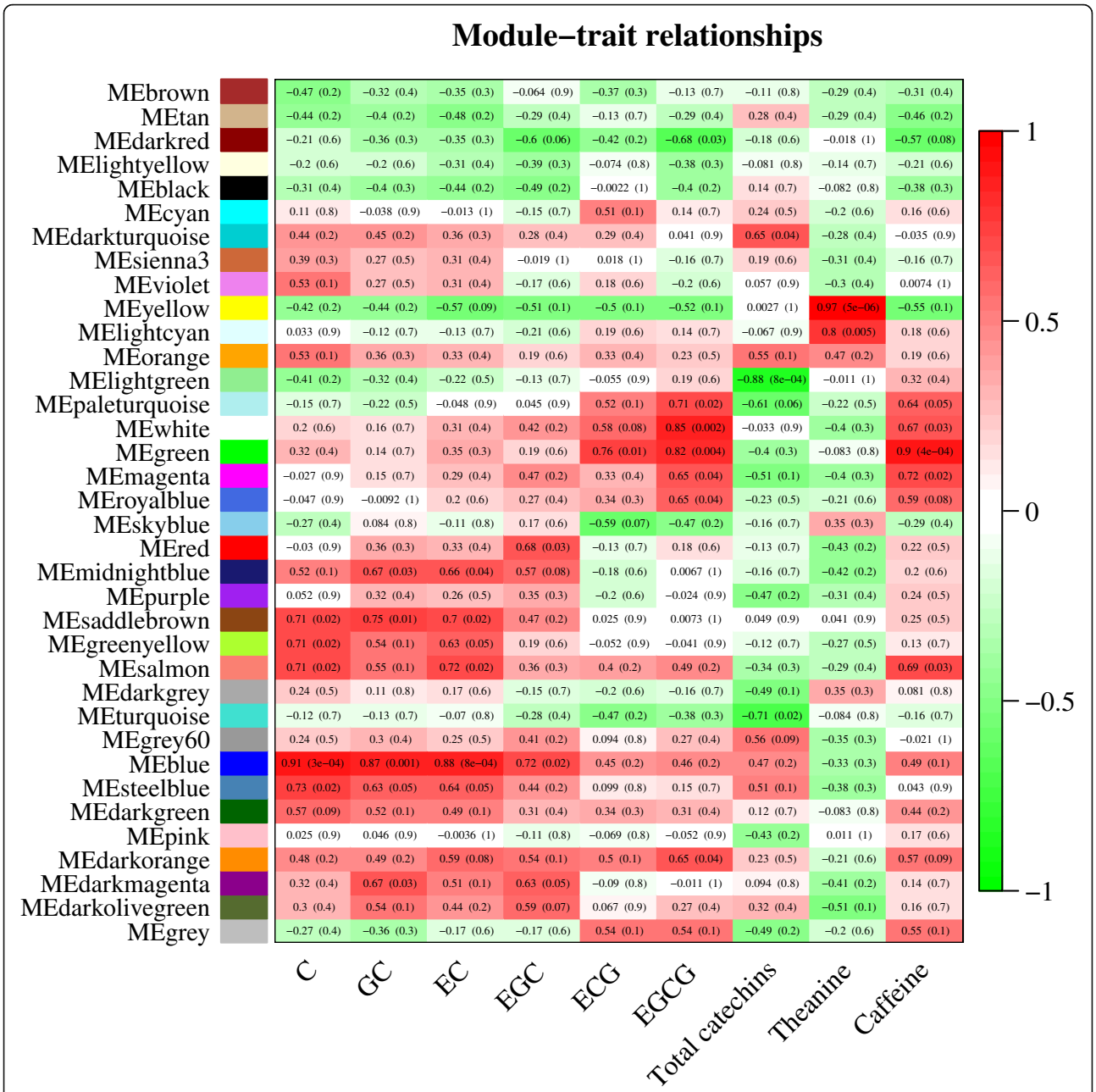
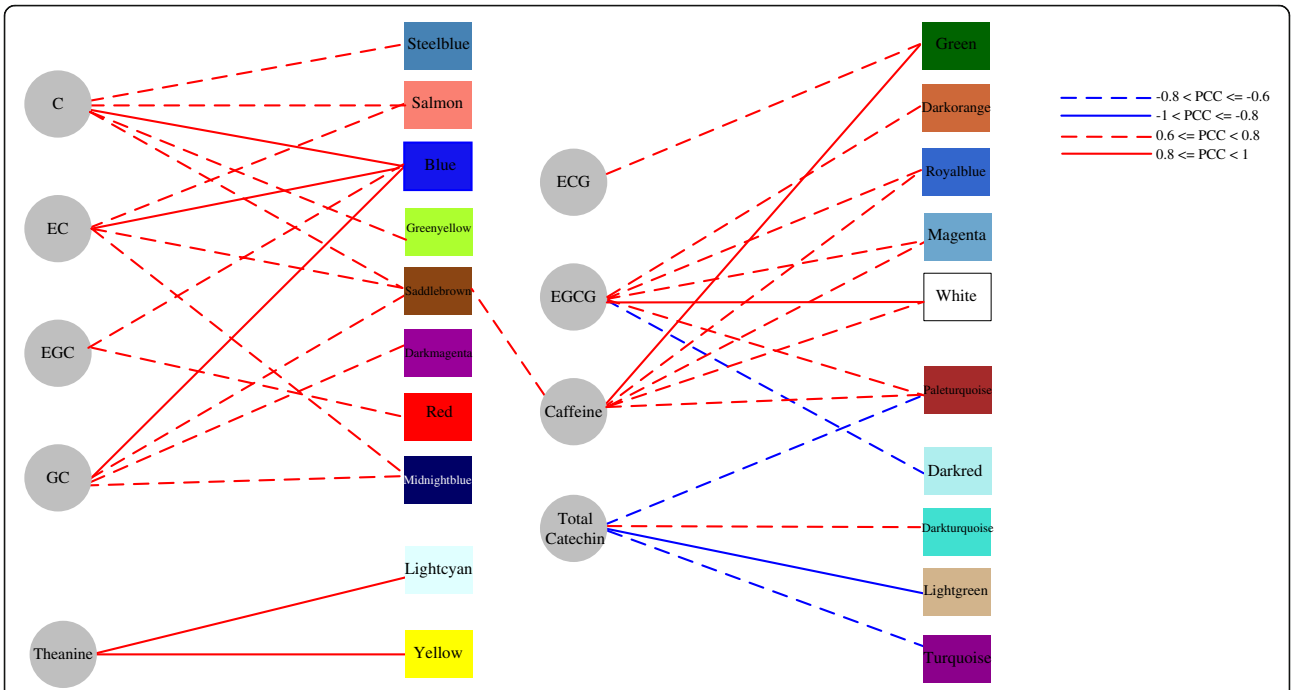


Fig. 4 Module-trait relationships. Each row corresponds to a module eigengene (correlation between a column and a trait). Each cell contains the corresponding correlation and *p*-value. The table is colour-coded by correlation in accordance with the figure

as well as shikimate O-hydroxycinnamoyltransferase (*HCT*) and UDP-glycosyltransferase. The correlation coefficient between F3'5'H and ECG, and EGCG and caffeine was more than 0.76 (p-value < 0.05), and the correlation coefficient between *βG* and ECG, and EGCG and caffeine was more than 0.67 (p-value < 0.05). However, two GS genes were not significantly correlated with catechins and caffeine, although they were significantly correlated with theanine (correlation coefficient > 0.98, p-value < 0.001; Fig. 6 and Table 2).

Genes highly co-expressed with *F3'5'H*, *FLS*, *βG* and *GSs* are shown in Fig. 7. Many more genes were co-expressed with *F3'5'H* and *βG* than with *FLS*. We also found five genes involved in 'Purine metabolism' that were co-expressed with *F3'5'H* and *βG*, and 11 genes involved in 'Biosynthesis of amino acids' that were co-expressed with *F3'5'H*, *βG* and *FLS* (Fig. 7a). A number of genes were co-expressed with *GSs*, along with ten, one and three genes involved in 'Phenylpropanoid biosynthesis, 'Flavonoid biosynthesis' and



Module	KEGG Pathway	Genes in module	All annotated genes	Pvalue	Qvalue
Blue	Photosynthesis - antenna proteins	28	46	1.53E-09	2.04E-06
	Ribosome	142	1018	1.78E-06	1.19E-04
	Photosynthesis	36	171	3.74E-06	1.66E-04
	Ubiquitin mediated proteolysis	103	767	1.89E-04	6.28E-02
Greenyellow	Spliceosome	36	1732	1.94E-04	2.00E-02
Red	Photosynthesis	44	171	3.74E-21	4.86E-19
	Aminoacyl-tRNA biosynthesis	48	434	1.93E-08	1.26E-06
	Porphyrin and chlorophyll metabolism	32	246	1.14E-07	3.84E-06
	Metabolic pathways	538	9649	1.18E-07	3.84E-06
	Vitamin B6 metabolism	12	52	3.02E-06	7.86E-05
	Biosynthesis of secondary metabolites	317	5504	7.47E-06	1.62E-04
	Glyoxylate and dicarboxylate metabolism	39	419	2.49E-05	4.63E-04
	Carbon metabolism	92	1318	4.36E-05	7.09E-04
	Glycine, serine and threonine metabolism	31	316	6.14E-05	8.88E-04
	Biosynthesis of amino acids	84	1297	9.43E-04	1.23E-02
	One carbon pool by folate	12	93	1.09E-03	1.28E-02
	Glutathione metabolism	25	305	3.75E-03	4.06E-02
	Degradation of aromatic compounds	7	45	4.07E-03	4.07E-02
Lightcyan	Flavonoid biosynthesis	13	307	1.03E-06	1.12E-04
	Plant-pathogen interaction	37	2223	1.25E-05	6.80E-04
	Phenylpropanoid biosynthesis	19	937	1.60E-04	5.80E-03
	Stilbenoid, diarylheptanoid and gingerol biosynthesis	9	280	3.76E-04	1.02E-02
	Anthocyanin biosynthesis	5	102	1.23E-03	2.68E-02
Yellow	Plant-pathogen interaction	246	2223	4.70E-23	6.20E-21
	Isoflavonoid biosynthesis	25	117	1.16E-08	7.65E-07
	Anthocyanin biosynthesis	21	102	3.25E-07	1.43E-05
	Biosynthesis of secondary metabolites	398	5504	1.12E-06	3.69E-05
	Stilbenoid, diarylheptanoid and gingerol biosynthesis	32	280	2.00E-04	5.27E-03
	Regulation of autophagy	45	477	9.47E-04	1.79E-02
	Degradation of aromatic compounds	9	45	9.49E-04	1.79E-02
	Metabolic pathways	616	9649	2.63E-03	4.34E-02
Magenta	Porphyrin and chlorophyll metabolism	17	246	9.60E-06	1.12E-03
	Ribosome	39	1018	8.26E-05	4.83E-03
	Photosynthesis	11	171	6.36E-04	2.48E-02
Darkorange	ABC transporters	11	604	1.91E-06	1.38E-04
	Stilbenoid, diarylheptanoid and gingerol biosynthesis	6	280	1.89E-04	6.82E-03

Module	KEGG Pathway	Genes in module	All annotated genes	Pvalue	Qvalue
Green	DNA replication	52	396	6.78E-10	8.48E-08
	Base excision repair	39	282	2.13E-08	1.33E-06
	Homologous recombination	44	350	4.93E-08	2.05E-06
	Nucleotide excision repair	39	426	3.99E-04	1.25E-02
	Mismatch repair	34	377	1.17E-03	2.88E-02
	Phagosome	38	440	1.38E-03	2.88E-02
	Purine metabolism	70	963	2.48E-03	4.43E-02
Royalblue	Fatty acid elongation	7	133	6.21E-07	4.66E-05
	Fatty acid metabolism	7	176	4.77E-04	1.79E-02
	Biosynthesis of unsaturated fatty acids	4	122	1.06E-03	2.64E-02
	Biotin metabolism	3	66	1.82E-03	3.41E-02
	RNA degradation	10	978	3.20E-03	4.51E-02
	Monoterpenoid biosynthesis	3	84	3.61E-03	4.51E-02
Paleturquoise	Tropane, piperidine and pyridine alkaloid biosynthesis	3	188	1.12E-03	3.62E-02
	Phenylalanine, tyrosine and tryptophan biosynthesis	3	222	1.81E-03	3.62E-02
	Base excision repair	3	282	3.55E-03	4.74E-02
Turquoise	Spliceosome	267	1732	4.51E-17	5.96E-15
	Ribosome biogenesis in eukaryotes	105	740	7.44E-06	4.91E-04
	Arginine biosynthesis	41	235	5.77E-05	1.94E-03
	Glucosinolate biosynthesis	31	160	6.08E-05	1.94E-03
	Protein processing in endoplasmic reticulum	166	1342	7.34E-05	1.94E-03
	Cyanoamino acid metabolism	71	489	1.01E-04	2.23E-03
	Alanine, aspartate and glutamate metabolism	48	299	1.23E-04	2.31E-03
	RNA transport	200	1686	1.58E-04	2.60E-03
	Mismatch repair	56	377	2.89E-04	4.24E-03
	Plant hormone signal transduction	185	1589	6.78E-04	8.88E-03
	ABC transporters	80	604	7.40E-04	8.88E-03
	Nitrogen metabolism	39	251	9.50E-04	1.04E-02
	Glyoxylate and dicarboxylate metabolism	58	419	1.33E-03	1.35E-02
	Ubiquitin mediated proteolysis	96	767	1.53E-03	1.44E-02
	Basal transcription factors	39	265	2.60E-03	2.29E-02
	Nucleotide excision repair	57	426	3.12E-03	2.58E-02
	Linoleic acid metabolism	23	142	5.76E-03	4.48E-02
	DNA replication	52	396	6.62E-03	4.61E-02
	RNA polymerase	51	387	6.64E-03	4.61E-02

Fig. 5 KEGG functional enrichment analysis of co-expression modules associated with phenotypic traits. The top panel shows correlations between modules and traits, while the bottom panel indicates the KEGG functional enrichment analysis of each trait-related module

'Anthocyanin biosynthesis', respectively. We also identified six genes involved in 'Purine metabolism' (Fig. 7b).

Validation of unigenes by qPCR

In order to confirm the accuracy of unigene expression levels, eight unigenes from 18 content-related modules were selected for qPCR analysis, and their relative expression levels were compared with FPKM values from RNA-Seq data. The results showed that expression of all eight unigenes measured by qPCR was consistent with the RNA-Seq data. Of the eight unigenes, correlation analysis between FPKM values and qPCR data showed that three had a correlation coefficient > 0.9, and three had a correlation coefficient > 0.7 (Fig. 8).

Discussion

We constructed a *C. sinensis* gene co-expression network using a WGCNA approach and identified co-expression modules using transcriptome data from ten tissues. Correlation analysis between co-expression modules and three characteristic metabolites (catechins, theanine and caffeine) was carried out, and 20 significant content-related modules (p-value < 0.05) and 6 highly significant content-related modules (correlation coefficient ≥ 0.8 and p-value < 0.05) were identified. These modules consist of highly connected

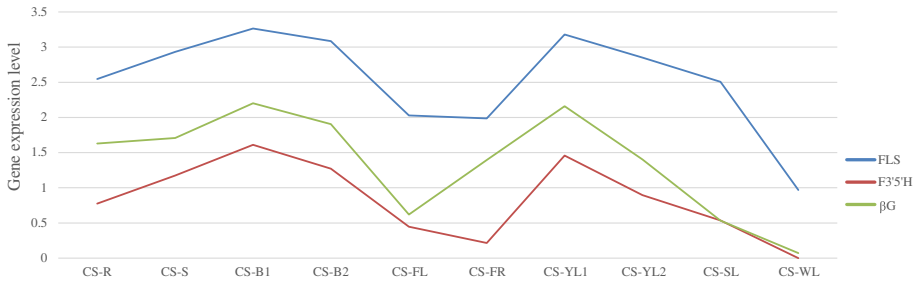


Fig. 6 Gene expression patterns of F3′5′H, FLS and βG in the green module. Expression levels of F3′5′H, FLS and βG genes are shown using log10-transformed FPKM values

functional genes, and different modules appear to be involved in individual functions [29]. Meanwhile, KEGG pathway enrichment analysis of modules associated with catechins, theanine and caffeine indicated that the three characteristic secondary pathways in tea plants are related to each other at the transcriptomic level. The results also indicated that one component can be regulated by multiple modules, and one module can simultaneously be associated with multiple components (Fig. 5).

Amino acid metabolism-related pathways were also found to be enriched in modules associated with catechins. For instance, 'Glycine, serine and threonine metabolism', and 'Biosynthesis of amino acids' were significantly enriched in the red module, which was highly positively correlated with EGC. 'Arginine biosynthesis', 'Alanine, aspartate and glutamate metabolism', and 'Nitrogen metabolism', were significantly enriched in the turquoise module, which was highly negatively correlated with total catechins. 'Carbon metabolism' was significantly enriched in the red module. Pathways enriched in the red and turquoise modules suggest that carbon and amino acid metabolism may have an important influence on EGC, EGCG and total catechins. Unigenes in the yellow and light cyan modules, which were significantly associated with theanine, were enriched in pathways involved in catechin biosynthesis such as 'Flavonoid biosynthesis', 'Anthocyanin biosynthesis' and 'Isoflavonoid biosynthesis'. These results indicate that biosynthesis of theanine is highly correlated with catechins, which suggests that theanine might play a vital role in the biosynthesis of catechins as unknown precursors. This result is consistent with the work of Feldheim et al. [30] who monitored the turnover of theanine in tea by investigating the distribution of isotopically labelled N-ethyl theanine in tea seedlings and young shoots. They found that the N-ethyl group of theanine was incorporated into the phloroglucinol nucleus of catechins. Similarly, Tanaka et al. [31] demonstrated that theanine is degraded to a Strecker aldehyde and conjugated with polyphenol rings to generate a novel polyphenol, ethylpyrrolidinonyl theasinensin, during the production of black tea.

In plants, nucleotides can be derived de novo from 5-phosphoribosyl-1-pyrophosphate and various simple molecules, but they can also be synthesised from preformed nucleosides and nucleobases via salvage reactions [32]. The de novo pathway of purine and pyrimidine biosynthesis is constitutive, but salvage enzymes may perform a special role in the activation of resting cells and in the response to environmental changes [33]. The green module was positively correlated with caffeine, ECG and EGCG. Based on KEGG pathway enrichment analysis, 'Purine metabolism', 'Nucleotide excision repair', 'DNA replication', 'Homologous recombination' and 'Base excision repair' were significantly enriched, suggesting salvage reactions, rather than the de novo pathway, may play a more important role in the biosynthesis of caffeine.

Module hub genes are generally considered representative of a given module in a biological network. Previous studies reported that MYB-bHLH-WDR (MBW) ternary complexes comprise the essential regulatory machinery for catechin and anthocyanin biosynthesis [34, 35]. In the present study, transcription factors MYB, bHLH, WD40,

Table 2 Correlation analysis of gene expression related to caffeine and catechins

Gene ID	Correlation coefficient				*p*-value			
	Caffeine	ECG	EGCG	Theanine	Caffeine	ECG	EGCG	Theanine
F3′5′H_Unigene20838	0.86	0.76	0.81	–	0.001483	0.01045	0.004592	–
FLS_Unigene16195	0.88	0.76	0.83	–	0.000645	0.01005	0.003184	–
βG_CL6189.Contig3	0.79	0.67	0.76	–	0.006371	0.03351	0.0110702	–
GS_Unigene22464	–	–	–	0.98275	–	–	–	0.00000038
GS_CL1326.Contig1	–	–	–	0.98167	–	–	–	0.000000483

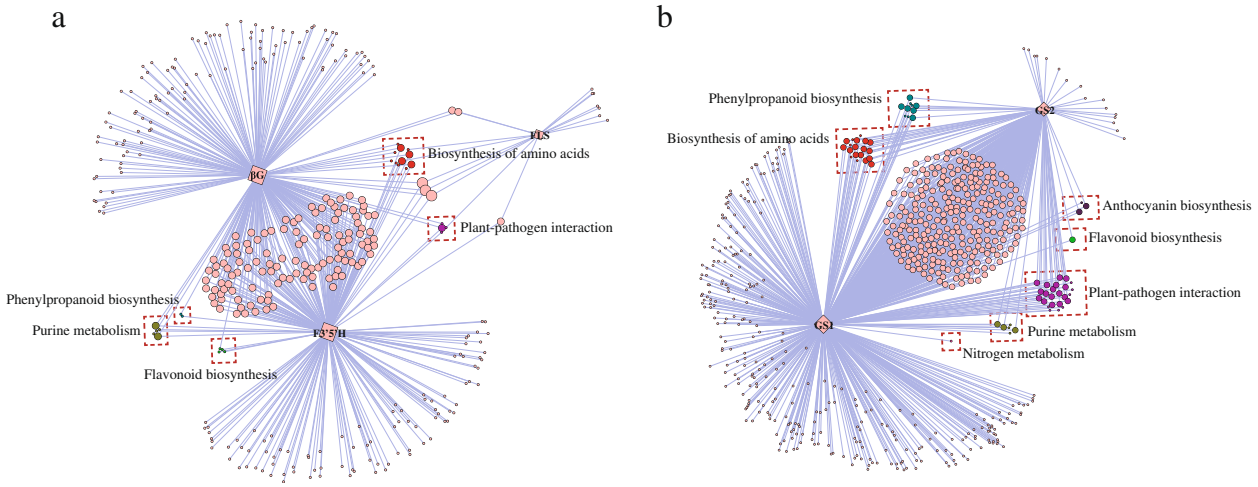


Fig. 7 Genes highly co-expressed with genes involved in catechin and theanine pathways. **a** Genes highly co-expressed with F3'5'H, FLS and βG in the catechin pathway. **b** Genes highly co-expressed with GS in the theanine pathway. Coloured circles represent genes, and edges represent correlations among genes

WRKY, and zinc finger were identified as hub genes in modules related to C, GC, EC and EGC. In addition, three genes involved in flavonoid biosynthesis (*F3'5'H, FLS* and *βG*) were identified in modules related to ECG, ECGC and caffeine, along with two genes involved in theanine biosynthesis (*GS*) in a module related to theanine. Flavonoid-3'5'-hydroxylase (*F3'5'H*), which belongs to the cytochrome P450 family, is the key enzyme related to anthocyanin biosynthesis [36]. Galloylated catechins such as ECG or EGCG are produced by ECGT from the

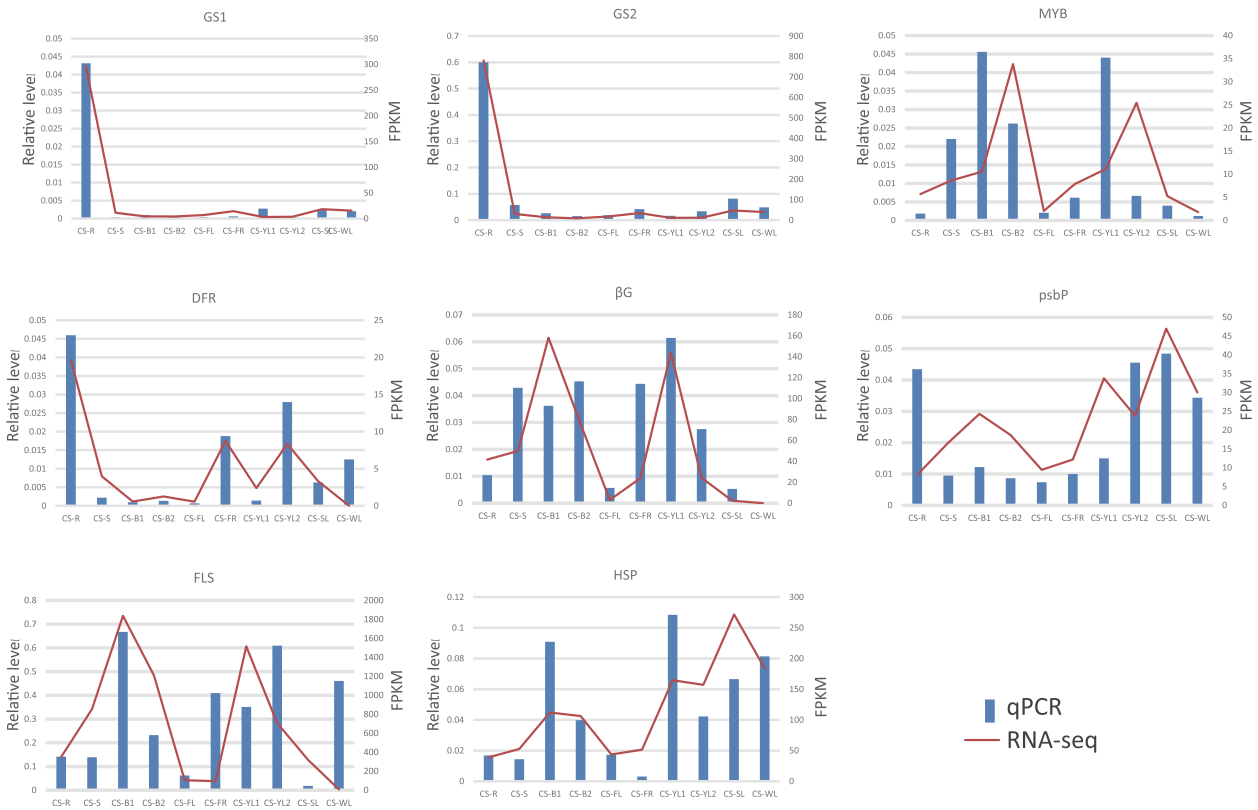


Fig. 8 Validation of the expression of candidate unigenes by qPCR. Gene expression levels were determined by qPCR and are presented as mean ± SD values calculated by the $2^{\Delta Ct}$ method

substrates βG and nongalloylated catechins EC or EGC [37]. Correlation analysis between *F3'5'H*, *FLS*, *βG*, catechins and caffeine showed that *F3'5'H*, *FLS* and *βG* were significantly positively correlated not only with ECG and EGCG, but also caffeine, which indicates the existence of a regulatory relationship between catechin and caffeine pathways. Genes highly co-expressed with *F3'5'H*, *βG* and *FLS* were also involved in purine metabolism and biosynthesis of amino acids, while genes highly co-expressed with *GS*s were also involved in phenylpropanoid biosynthesis, flavonoid biosynthesis, anthocyanin biosynthesis and purine metabolism.

Tea is a sciophilous plant adapted to the understorey of tropical rainforests that possesses numerous inducible physiological adaptations protecting against light-associated damage. In previous studies, researchers found that shade treatment can effectively improve the quality of tea beverages [38], and the leaves of tea plants grown in the shade contain higher amino acid levels and a lower catechin content [39]. Furthermore, shade treatment can effectively reduce the biosynthesis of flavonoids and lignins by reducing the expression of genes in the flavonoid pathway [40]. Researchers [41] cloned a novel CsDFR gene that actively responds to light treatment, and showed that light might be effective for activating the biosynthesis of phenylpropanoids that protect against light stimuli. Recently, Tai et al. [42] analysed the promoters of *LAR, TCS* and *TS* in a tea BAC library, and identified numerous light-responsive cis-acting elements in *LAR, TCS* and *TS* genes. Light is an important environmental parameter that drives photosynthesis, and it might regulate genes related to the catechin biosynthesis pathway [40, 41]. In the present study, we found that photosynthesis-related unigenes were significantly enriched in modules positively associated with C, GC, EC, EGC and EGCG based on KEGG pathway enrichment analysis. These findings strongly indicate that photosynthesis might influence catechin synthesis via provision of carbon sources. Further investigations are clearly required to uncover the relationship between light and the biosynthesis of catechins.

Conclusion

We analysed the content of catechins, theanine and caffeine in ten different tissues from tea plants, and constructed a co-expression network to investigate relationships between genes and these three characteristic metabolites. The results indicated that genes related to catechins, theanine and caffeine were influenced by each other, especially key genes associated with the metabolic pathways of these characteristic compounds. Furthermore, light was identified as an important factor in the biosynthesis of catechins. WGCNA proved to be a novel method for analysing the connection between metabolites and gene expression. This method holds potential for further exploration of large-scale transcriptomic data.

Methods

Plant material

Six-year-old tea plants (*C. sinensis* L. O. Kuntze cv. *Shuchazao*) were used in this study. The field experiment was performed in a typical tea-producing garden at De Chang Fabrication Base in Anhui Province, China (Shucheng, latitude 31.3 N, longitude 117.2E above sea level) under natural conditions. Tea plants were grown in an experimental plot with 150 cm between rows and 40 cm between plants within a row, and yellow brown acidic soil was employed. Tea plants were divided into three groups, each group consisted three rows, with at least 100 plants pooled per group, and samples were randomly selected from these three groups, with no fewer than 6–10 samples obtained from ten different tissues of tea plants. The maximum air temperature in the tea garden was about 27°C in the daytime and a minimum of 21°C at night during May and June, and ~ 27°C in the daytime and a minimum of 21°C at night in December. Apical buds (CS-B1) in May, apical buds (CS-B2) in June, first young leaves (CS-YL1) in June, second young leaves (CS-YL2) in June, mature leaves in summer (CS-SL) in June, stem (CS-S), mature leaves in winter (CS-WL) in December, flowers (CS-FL), fruits (CS-FR) in June, and roots (CS-R) in June comprised the ten different tissue samples studied. Tea plants were watered and fertilised equivalently, and tea plants with uniform height and crown breadth, and without signs of disease and insects, were selected for experiments. Three biological replicates were performed for each sample. All samples were immediately frozen in liquid nitrogen, and stored at − 80°C until RNA extraction.

Extraction and HPLC determination of catechins, theanine and caffeine

Catechins, theanine and caffeine were extracted from samples as described previously [18] with some modifications. Catechins and caffeine were extracted with 80% methanol, while theanine was extracted with hot water as previously described [43]. The obtained supernatants (catechins, theanine and caffeine) were filtered through a 0.22 μm membrane prior to HPLC analysis. All samples were analysed using three biological repeats. The catechin and caffeine content was determined using a Waters 2695 HPLC system (Waters, USA). The column temperature was set to 25°C, and the detection wavelength was 278 nm. The theanine content was measured using a Waters 600E series HPLC system (Waters, USA) at a detection wavelength of 199 nm [44]. The mobile phase ratio and injection approach were as described previously [18]. C, GC, EC, EGC, ECG, EGCG, theanine and caffeine

standards were purchased from Shanghai Winherb Medical Technology, Ltd., China.

RNA isolation, transcriptomic library construction and RNA-Seq

Total RNA was extracted separately from the ten tissues using a modified CTAB (cetyltrimethyl ammonium bromide) method with three biological replicates [45]. The yield and quality of RNA were determined by agarose gel electrophoresis (AGE) and a Nanodrop 2000 instrument. RNA samples with an A260/A280 > 1.8, A260/A230 > 1.8, and RNA integrity number (RIN) > 8 were considered acceptable for library construction. Equal amounts of RNA from three different samples were pooled before cDNA library preparation, and mRNA enrichment, cDNA synthesis, fragmentation, adapter addition, selection of fragment size, PCR amplification, and transcriptomic sequencing were performed by staff at the Beijing Genome Institute (BGI; Shenzhen, China) as previously described by Liu et al. [46] and Gu et al. [47]. Briefly, mRNAs were purified from total RNA using magnetic beads with Oligo (dT) and cleaved into short sequences. First-strand cDNA synthesis was then performed with random primers (TaKaRa, Japan), and double-stranded cDNAs were then prepared using these short fragments as templates. Adapters were ligated to the short fragments using T4 DNA ligase (Invitrogen, USA), and after end repair and ligation of adapters, products were enriched by PCR to generate the cDNA library. The cDNA library was examined using an Agilent 2100 Bioanalyzer prior to sequencing on an Illumina HiSeq 2000 sequencing platform [48]. Paired-end reads were generated with a length of 90 bp for each read.

Data preprocessing and de novo assembly

Raw sequencing reads were subjected to preliminary screening to remove low-quality reads and reads with adaptor sequences using the filter command in SOAPnuke (version 1.5.6) with the low-quality threshold set to 10 [49]. We obtained 111 gigabases (average 11 gigabase/sample) from ten samples, and in order to reduce the assembly error, remaining reads were then de novo assembled using the Trinity package (release-20,130,225) with parameters '--min_glue 3 and --min_kmer_cov 3'. These assembly parameters were used previously by Shao et al. [50] and Li et al. [51]. Assembled unigenes were defined after removing short and redundant contigs (≤200 bp) with TGICL software (version 2.1) [23] from the Trinity assembly. Final assembled datasets were defined by removing unigenes with FPKM values less than 0.3 based on the work of Ramskold et al. [26], as applied in several other studies including Łabaj et al. [52] and Sam et al. [53]

Annotation and classification of unigenes

All assembled non-redundant and filtered unigenes were annotated by alignment to the NR database, the Swiss-Prot protein database, and the COG database using BLASTX with an E-value cut-off of 1×10^{-5} [54]. Meanwhile, these unigenes were also annotated using the NT database and BLASTN. KEGG metabolic pathway annotation of unigenes was carried out after mapping to the KEGG database [55], which helped to elucidate the complex biological functions of genes. Based on the BLAST results from the NR database, GO annotation was carried out using the Blast 2 GO program (version 2.3.4) [56].

Identification of gene expression and DEGs

Expression levels of unigenes were calculated using the FPKM method. Firstly, reads were mapped to unigene datasets by Bowtie2 (version 2.1.0, http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) at a sensitive setting. Based on the Bowtie results, FPKM values for each unigene were subsequently calculated by RESM (version 1.2.29) [57] with default parameters. DEGs were identified based on the method described by Audic et al. [58]. Genes with $|\log2ratio| \geq 1$ and false discovery rate (FDR) < 0.05 were identified as DEGs.

Construction of gene co-expression networks

Gene co-expression networks were constructed using the WGCNA approach with R packages (version 3.2.2). DEGs expressed in at least one pairwise comparison in ten tissues were retained for co-expression network construction by WGCNA analysis [11]. All tissues were initially clustered to analyse the sample height. Following application of the scale-free topology criterion described previously, a soft threshold of 30 was chosen. Based on the topological overlap-based dissimilarity measure [59], unigenes were first hierarchically clustered, and the gene dendrogram was used for module detection by the dynamic tree cut method (mergeCutHeight = 0.25, minModuleSize = 30). In the weighted gene co-expression network, gene connectivity was based on the edge weight (ranging from 0 to 1) determined by the topology overlap measure, which reflects the strength of the communication between two genes. The weights across all edges of a node were summed and used to define the level of connectivity, and nodes with high connectivity were considered hub genes.

Identification of content-related modules

To identify modules associated with catechins, theanine and caffeine, we first calculated the module eigengenes of each module, then correlated these with the catechin, theanine and caffeine content using Pearson's correlation coefficients and an asymptotic confidence interval based on Fisher's Z transformation. Modules with

p-values < 0.05 were identified as content-related modules. To further characterise these modules, enrichment of annotated unigenes in each content-related module was investigated using the phyper function within the R platform based on KEGG pathway annotation, and q-value or FDR corrections were applied by multiple testing [60]. We defined KEGG pathways with a q-value or FDR < 0.05 as significantly enriched [61].

Module hub gene selection and visualisation

The most central and connected genes, involved in numerous interactions, were considered hub genes [62], which are likely to play a more important role in a given module than other genes in the overall co-expression network. In this study, we categorised the top 2% of the most highly connected genes in a module as hub genes based on the size of the module. Co-expression interactions and patterns of hub genes were visualised using Cytoscape [63].

qPCR validation of selected unigenes

In order to evaluate the assembly quality of RNA-seq data, the expression patterns of eight selected transcripts were monitored by qPCR. RNA samples were isolated from samples using the CTAB method [45], and total RNA was reverse-transcribed into single-stranded cDNAs using a reverse transcription kit for real-time PCR (TaKaRa). Detailed information (unigene IDs and primer sequences) related to the selected transcripts used for qPCR is listed in Additional file 5. PCR amplification was performed according to the manufacturer's instructions using a CFX96TM real-time PCR system (Bio-Rad) with an annealing temperature of 60 °C. The housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) was used as an internal reference gene, and relative expression levels of target genes were calculated using the $2^{\Delta Ct}$ method [64]. All qPCRs were analysed using three technical and three biological replicates.

Additional files

Additional file 1: Clustering dendrogram of samples based on gene expression. (PDF 99 kb)

Additional file 2: Modules significantly ($p < 0.05$) correlated with characteristic components in tea. (XLSX 11 kb)

Additional file 3: Functional analysis of unigenes in blue, red and magenta modules enriched in the photosynthesis pathway. (XLSX 12 kb)

Additional file 4: Hub genes for each highly significant content-related module. (XLSX 54 kb)

Additional file 5: Unigene IDs and primer sequences for selected transcripts used for qPCR validation. (XLSX 9 kb)

Abbreviations

AGE: Agarose gel electrophoresis; C: Catechin; COG: Clusters of Orthologous Groups of Proteins; CS_FR: FruitCS-RRoot; CS-B1: Apical bud; CS-B2: Apical bud; CS-FL: Flower; CS-S: Stem; CS-SL: Mature leaf in summer; CS-WL: Mature leaf in winter; CS-YL1: First young leaf; CS-YL2: Second young leaf; EC: Epicatechin; ECG: Epicatechin gallate; EGC: epigallocatechin; EGCG: Epigallocatechin gallate; F3′5′H: Flavonoid 3′,5′-hydroxylase; FLS: Flavonol synthase; FPKM: Fragment Per Kilobase of exon model per Million mapped reads; GAPDH: Glyceraldehyde-3-phosphate dehydrogenase; GC: Gallocatechin; GO: Gene Ontology; GS: Glutamine synthetase; HCT: Shikimate O-hydroxycinnamoyl transferase; HPLC: High-performance liquid chromatography; KEGG: Kyoto Encyclopedia of Genes and Genomes; Nr: Non-redundant protein database; Nt: Non-redundant nucleotide database; qPCR: Quantitative real-time polymerase chain reaction; Swiss-Prot: Annotated protein sequence database; WGCNA: Weighted gene co-expression network analysis; βG: Beta-glucosidase

Acknowledgements

We would like to thank the native English speaking scientists of Elixigen Company (Huntington Beach, California) for editing our manuscript.

Funding

This work was supported by the Special Innovative Province Construction in Anhui Province [grant number 15czs08032], the Special Project for Central Guiding Science and Technology Innovation of Region in Anhui Province [grant number 2016080503B024], and the Natural Science Foundation of Anhui Province [grant number 31170283], and the Youth Foundation of Anhui Agricultural University [grant number 2016ZR012]. The design of this study and the manuscript writing were supported by the Science and Technology Project of Anhui Province, China [grant number 13Z03012]. And the collection, analysis, and interpretation of data used in this study were supported by youth Foundation of Anhui Agricultural University [grant number 2016ZR012].

Availability of data and materials

The Illumina RNA-seq data generated from *C. sinensis* are available from the NCBI SRA database (http://trace.ncbi.nlm.nih.gov/Traces/sra) under accessions SRR1928149 and SRP056466.

Authors' contributions

XCW and CLW participated in the research design. YLT and CL analysed data and drafted the manuscript, SWY and JMS performed most of the experiments, and CXG, YY, BH and ZYL provided technical assistance. HY participated in material collection and manuscript revision. EHX contributed to revising the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Six-year-old tea plants (*C. sinensis* cv. *Shuchazao*) were used in this study. The field experiment was performed in a typical tea-producing garden at De Chang Fabrication Base in Anhui Province, China (Shucheng, latitude 31.3 N, longitude 117.2E) under natural conditions. Collection of plant materials complied with the institutional, national and international guidelines. No specific permits were required.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]School of Life Science, Anhui Agricultural University, Hefei 230036, China. [2]State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei 230036, China. [3]BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China.

Received: 29 September 2017 Accepted: 8 August 2018
Published online: 15 August 2018

References

1. Mahmood T, Akhtar N, Khan BA. The morphology, characteristics, and medicinal properties of *Camellia sinensis*' tea. J Med Plant Res. 2010; 4(19):2028–33.
2. Mamati GE, Liang Y, Lu J. Expression of basic genes involved in tea polyphenol synthesis in relation to accumulation of catechins and total tea polyphenols. J Sci Food Agric. 2006;86(3):459–64.
3. Punyasiri P, Abeysinghe I, Kumar V, Treutter D, Duy D, Gosch C, Martens S, Forkmann G, Fischer T. Flavonoid biosynthesis in the tea plant *Camellia sinensis*: properties of enzymes of the prominent epicatechin and catechin pathways. Arch Biochem Biophys. 2004;431(1):22–30.
4. Liang YR, Ma WY, Lu JL, Wu Y. Comparison of chemical compositions of *Ilex latifolia Thumb* and *Camellia sinensis* L. Food Chem. 2001;75(3):339–43.
5. Sasazuki S, Tamakoshi A, Matsuo K, Ito H, Wakai K, Nagata C, Mizoue T, Tanaka K, Tsuji I, Inoue M. Green tea consumption and gastric cancer risk: an evaluation based on a systematic review of epidemiologic evidence among the Japanese population. Jpn J Clin Oncol. 2012;42(4):335.
6. Hodgson JM, Croft KD, Woodman RJ, Puddey IB, Fuchs D, Draijer R, Lukoshkova E, Head GA. Black tea lowers the rate of blood pressure variation: a randomized controlled trial. Am J Clin Nutr. 2013;97(5): 943–50.
7. Hollman PC, Feskens EJ, Katan MB. Tea flavonols in cardiovascular disease and cancer epidemiology. Proc Soc Exp Biol Med. 1999;220(4):198–202.
8. Auvichayapat P, Prapochanung M, Tunkamnerdthai O, Sripanidkulchai BO, Auvichayapat N, Thinkhamrop B, Kunhasura S, Wongpratoom S, Sinawat S, Hongprapas P. Effectiveness of green tea on weight reduction in obese Thais: a randomized, controlled trial. Physiol Behav. 2008;93(3):486–91.
9. Fuller T, Langfelder P, Presson A, Horvath S. Review of weighted gene coexpression network analysis. In: Handbook of Statistical Bioinformatics: Springer; 2011. p. 369–88.
10. Ruan J, Dean AK, Zhang W. A general co-expression network-based approach to gene expression analysis: comparison and applications. BMC Syst Biol. 2010;4(1):8.
11. Steve H, Peter L. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9(1):559.
12. Shaik R, Ramakrishna W. Genes and co-expression modules common to drought and bacterial stress responses in Arabidopsis and rice. PLoS One. 2013;8(10):e77261.
13. Obanda M, Owuor PO. Impact of shoot maturity on chlorophyll content, composition of volatile flavour compounds and plain black tea chemical quality parameters of clonal leaf. J Sci Food Agric. 1995;69(4):529–34.
14. Pokinda O, Martin O, Hastingse N, Wilsonl M. Influence of region of production on clonal black tea chemical characteristics. Food Chem. 2008; 108(1):263–71.
15. Nagar PK, Sood S. Changes in endogenous auxins during winter dormancy in tea (*Camellia sinensis* L.) O. Kuntze. Acta Physiol Plant. 2006;28(2):165–9.
16. Wang Z, Gerstein M, Snyder M, Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
17. Shi CY, Hua Y, Wei CL, Yu O, Zhang ZZ, Jiang CJ, Sun J, Li YY, Qi C, Tao X. Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. BMC Genomics. 2011;12(1):131.
18. Tai Y, Wei C, Yang H, Zhang L, Chen Q, Deng W, Wei S, Zhang J, Fang C, Ho C. Transcriptomic and phytochemical analysis of the biosynthesis of characteristic constituents in tea (*Camellia sinensis*) compared with oil tea (*Camellia oleifera*). BMC Plant Biol. 2015;15(1):190.
19. Xin-Chao Wang Q-YZ, Ma C-l, Zhang Z-H, Cao H-l, Kong Y-M, Yue C, Hao X-Y, Chen L, Ma J-Q, Jin J-Q, li X, Yang Y-J. Global transcriptome profiles of Camellia sinensis during cold acclimation. BMC Genomics. 2013;14(1):1–15.
20. Thirugnanasambantham K, Prabu G, Palanisamy S, Chandrabose SR, Mandal AK. Analysis of dormant bud (Banjhi) specific transcriptome of tea (*Camellia sinensis* (L.) O. Kuntze) from cDNA library revealed dormancy-related genes. Appl Biochem Biotechnol. 2013;169(4):1405–17.
21. Li CF, Zhu Y, Yu Y, Zhao QY, Wang SJ, Wang XC, Yao MZ, Luo D, Li X, Chen L. Global transcriptome and gene regulation network for secondary metabolite biosynthesis of tea plant (*Camellia sinensis*). BMC Genomics. 2015;16(1):560.
22. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52.
23. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B. TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics. 2003;19(5):651.
24. Rau A, Gallopin M, Celeux G, Jaffrézic F. Data-based filtering for replicated high-throughput transcriptome sequencing experiments. Bioinformatics. 2013;29(17):2146–52.
25. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9(4):357.
26. Ramskold D, Wang ET, Burge CB, Sandberg R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol. 2009;5(12):e1000598.
27. Bai S, Sun Y, Qian M, Yang F, Ni J, Tao R, Lin L, Shu Q, Dong Z, Teng Y. Transcriptome analysis of bagging-treated red Chinese sand pear peels reveals light-responsive pathway functions in anthocyanin accumulation. Sci Rep. 2017;7(1):63.
28. Wu S, Lei J, Chen G, Chen H, Cao B, Chen C. De novo Transcriptome assembly of Chinese kale and global expression analysis of genes involved in Glucosinolate metabolism in multiple tissues. Front Plant Sci. 2017;8:92.
29. Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH. Functional organization of the transcriptome in human brain. Nat Neurosci. 2008;11(11):1271.
30. Feldheim W, Yongvanit P, Cummings PH. Investigation of the presence and significance of theanine in the tea plant. J Sci Food Agric. 1986;37(6):527–34.
31. Tanaka T, Watarumi S, Fujieda M, Kouno I. New black tea polyphenol having N-ethyl-2-pyrrolidinone moiety derived from tea amino acid theanine: isolation, characterization and partial synthesis. Food Chem. 2005;93(1):81–7.
32. Emmersen J. Generating unigene collections of expressed sequence tag sequences for use in mass spectrometry identification. Methods Mole Biol (Clifton, NJ). 2007;367:77.
33. Stasolla C, Katahira R, Thorpe TA, Ashihara H. Purine and pyrimidine nucleotide metabolism in higher plants. J Plant Physiol. 2003;160(11):1271–95.
34. Loïc L, Debeaujon I, Routaboul J-M, Baudry A, Pourcel L, Nesi N, Caboche M. Genetics and biochemistry of seed flavonoids. Plant Biol. 2006;57(57):405–30.
35. Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends Plant Sci. 2015;20(3):176–85.
36. Shimada Y, Nakano-Shimada R, Ohbayashi M, Okinaka Y, Kiyokawa S, Kikuchi Y. Expression of chimeric P450 genes encoding flavonoid-3′,5′-hydroxylase in transgenic tobacco and petunia plants 1. FEBS Lett. 1999;461(3):241–5.
37. Liu Y, Gao L, Liu L, Yang Q, Lu Z, Nie Z, Wang Y, Xia T. Purification and characterization of a novel Galloyltransferase involved in Catechin Galloylation in the tea plant (*Camellia sinensis*). J Biol Chem. 2012;287(53):44406–17.
38. Zhang WJ, Liang YR, Zhang FZ, Chen CS, Zhang YG, Chen RB, Weng BQ. Effects on the yield and quality of oolong tea by covering with shading net. J Tea Sci. 2004;24(4):276–82.
39. Kangmo K, Jungnam C, Jiyoung K, Jeongkee K, Langgook Y, Sangjun L, Hong YS, Choonghwan L. Metabolomics analysis reveals the compositional differences of shade grown tea (*Camellia sinensis* L.). J Agric Food Chem. 2009;58(1):418–26.
40. Wang YS, Gao LP, Shan Y, Liu YJ, Tian YW, Xia T. Influence of shade on flavonoid biosynthesis in tea (*Camellia sinensis* (L.) O. Kuntze). Sci Hortic. 2012;141(3):7–16.
41. Gedda L. Light-induced expression of genes involved in phenylpropanoid biosynthetic pathways in callus of tea (*Camellia sinensis* (L.) O. Kuntze). Sci Hortic. 2012;133(1):72–83.
42. Tai Y, Wang H, Wei C, Su L, Li M, Wang L, Dai Z, Shao L, Jiang J, Luo M. Construction and characterization of a bacterial artificial chromosome library for *Camellia sinensis*. Tree Genet Genomes. 2017;13(4):89.
43. Wang L, Gong LH, Chen CJ, Han HB, Li HH. Column-chromatographic extraction and separation of polyphenols, caffeine and theanine from green tea. J Crohns Colitis. 2016;10(8):873.
44. Yin-Hua LI, Liu ZH, Huang JA, Gong YS. Determination of Theanine in tea by reversed-phase high performance liquid chromatography-evaporative light-scattering detector. J Tea Sci. 2005;25(3):225–8.
45. Shi CY, Wan XC, Jiang CJ, Sun J. Method for high-quality total RNA isolation from tea plants (*Camellia sinensis* (L.) O. Kuntze). J Anhui Agric Univ. 2007; 34(3):360–3.

46. Liu S, Shi XX, Jiang YD, Zhu ZJ, Qian P, Zhang MJ, Yu H, Zhu QZ, Gong ZJ, Zhu ZR. De novo analysis of the *Tenebrio molitor* (Coleoptera: Tenebrionidae) transcriptome and identification of putative glutathione S -transferase genes. Appl Entomol Zool. 2014;50(1):1–9.
47. Gu XC, Zhang YN, Kang K, Dong SL, Zhang LW. Antennal Transcriptome Analysis of Odorant Reception Genes in the Red Turpentine Beetle (RTB), *Dendroctonus valens*. PLoS One. 2015;10(5):e0125159.
48. Tsanakas GF, Manioudaki ME, Economou AS, Kalaitzis P. De novo transcriptome analysis of petal senescence in *Gardenia jasminoides* Ellis. BMC Genomics. 2014;15(1):554.
49. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z. SOAPnuke: a MapReduce acceleration supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 2017;7(1):1–6.
50. Shao Y, Wei J, Wu F, Zhang H, Yang D, Liang Z, Jin W. DsTRD: Danshen transcriptional resource database. PLoS One. 2016;11(2):e0149747.
51. Li S, Liu H, Bai J, Zhu X. Transcriptome assembly and identification of genes and SNPs associated with growth traits in largemouth bass (Micropterus salmoides). Genetica. 2017;145(2):175–87.
52. Łabaj PP, Leparc GG, Linggi BE, Markillie LM, Wiley HS, Kreil DP. Characterization and improvement of RNA-Seq precision in quantitative transcript expression profiling. Bioinformatics. 2011;27(13):i383–91.
53. Sam LT, Lipson D, Raz T, Cao X, Thompson J, Milos PM, Robinson D, Chinnaiyan AM, Kumar-Sinha C, Maher CA. A comparison of single molecule and amplification based sequencing of cancer transcriptomes. PLoS One. 2011;6(3):e17305.
54. Altschul S, Gish W, Miller W, Myers E, Lipman D. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
55. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
56. Conesa A, Götz S, García-Gómez JM, Terol J, Manuel N, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674–6.
57. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323. Bmc Bioinform. 2011;12(1):323.
58. Audic S, Claverie JM. The significance of digital gene expression profiles. Genome Res. 1997;7(10):986–95.
59. Zhang B, Horvath S. A General Framework For Weighted Gene Co-Expression Network Analysis. Stat Appl Genet Mol Biol. 2005;4(1) Article17
60. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med. 1990;9(7):811–8.
61. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004; 32(Database issue):277–80.
62. Liu J, Ling J, Tu X. Weighted gene co-expression network analysis identifies specific modules and hub genes related to coronary artery disease. BMC Cardiovasc Disord. 2016;16(1):1–8.
63. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
64. Miracco C, Cosci E, Oliveri G, Luzi P, Pacenti L, Monciatti I, Mannucci S, De Nisi MC, Toscano M, Malagnino V. Protein and mRNA expression of autophagy gene Beclin 1 in human brain tumours. Int J Oncol. 2007; 30(2):429–36.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions