


RESEARCH

Open Access



# Characterization of novel loci controlling seed oil content in *Brassica napus* by marker metabolite-based multi-omics analysis

Long Li<sup>1,2†</sup>, Zhitao Tian<sup>1,2†</sup>, Jie Chen<sup>1,2</sup>, Zengdong Tan<sup>1,2</sup>, Yuting Zhang<sup>1,2</sup>, Hu Zhao<sup>1,2</sup>, Xiaowei Wu<sup>1,2</sup>, Xuan Yao<sup>1,2</sup>, Weiwei Wen<sup>3,4,5</sup>, Wei Chen<sup>1,2,4,5\*</sup> and Liang Guo<sup>1,2,4,5\*</sup> 

<sup>†</sup>Long Li and Zhitao Tian contributed equally to this work.

\*Correspondence: chenwei0609@mail.hzau.edu.cn; guoliang@mail.hzau.edu.cn

<sup>1</sup> National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China

<sup>2</sup> Hubei Hongshan Laboratory, Wuhan, China

<sup>3</sup> Key Laboratory of Horticultural Plant Biology (MOE), College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan, China

<sup>4</sup> Shenzhen Institute of Nutrition and Health, Huazhong Agricultural University, Wuhan, China

<sup>5</sup> Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

## Abstract

**Background:** Seed oil content is an important agronomic trait of *Brassica napus* (*B. napus*), and metabolites are considered as the bridge between genotype and phenotype for physical traits.

**Results:** Using a widely targeted metabolomics analysis in a natural population of 388 *B. napus* inbred lines, we quantify 2172 metabolites in mature seeds by liquid chromatography mass spectrometry, in which 131 marker metabolites are identified to be correlated with seed oil content. These metabolites are then selected for further metabolite genome-wide association study and metabolite transcriptome-wide association study. Combined with weighted correlation network analysis, we construct a triple relationship network, which includes 21,000 edges and 4384 nodes among metabolites, metabolite quantitative trait loci, genes, and co-expression modules. We validate the function of BnaA03.TT4, BnaC02.TT4, and BnaC05.UK, three candidate genes predicted by multi-omics analysis, which show significant impacts on seed oil content through regulating flavonoid metabolism in *B. napus*.

**Conclusions:** This study demonstrates the advantage of utilizing marker metabolites integrated with multi-omics analysis to dissect the genetic basis of agronomic traits in crops.

**Keywords:** *Brassica napus*, Metabolomics, mGWAS, mTWAS, WGCNA, Seed oil content

## Background

The plant kingdom contributes the most metabolite diversity and produces 100,000 to 1 million types of metabolites [1]. Plant metabolites are essential for plant growth and development, and they play key roles in the defense and adaptation to the environment [2, 3]. Metabolites link genotype to phenotype for physical traits [4–6]. Over recent years, more and more studies have discovered the connection between marker metabolites and agronomic traits, such as primary metabolites and starch-



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. 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 in a credit line to the data.

cold sweetening-related traits in potato [7]; fruit load in tomato [8], lignin precursors, and biomass in maize [9]; and secondary metabolites and maize kernel weight [10]. Compared with conventional QTL mapping of agronomic traits which usually discovers a limited quantity of loci [11–14], secondary metabolites have a highly polygenic architecture for large-effect loci [10, 15–18]. A vast array of studies shows that metabolomics is a powerful tool for grasping a better understanding of plant metabolite synthesis and functions [10, 15, 19, 20].

Metabolite genome-wide association studies (mGWAS) have detected lots of loci that are related to corresponding metabolites in several main crops, such as rice [15, 21], maize [10, 18], and wheat [22]. Transcriptome-wide association studies (TWAS) have been developed to interpret the relationship between gene expression and phenotype with whole-genome gene expression levels taken into account [23, 24]. Recent studies indicate that TWAS has been a novel and powerful tool for gene-phenotype prediction [25–27]. Metabolite TWAS (mTWAS) can unearth more metabolite-associated and trait-correlated genes. Expression genome-wide association study (eGWAS) is an approach using gene expression information to perform GWAS [28]. Therefore, gene expression diversity can be explained by genomic variation, and the expression quantitative trait loci (eQTLs) can imply the potential regulatory network of genes for dissecting complex traits [29, 30]. High-throughput metabolomics analysis united with multi-omics analysis is powerful for dissecting the genetic basis of complex traits in crops.

*Brassica napus* is the world's third-largest oil crop with a roughly 13% production of edible oil around the world. Improvement of seed oil content (SOC) has been one of the most important goals for *B. napus* breeding [31–33]. The first step of fatty acid (FA) synthesis in plants occurs in the plastid [34], where the carboxylation of acetyl-CoA is catalyzed by acetyl-CoA carboxylase to produce malonyl-CoA [35–37]. Then, fatty acid synthase uses malonyl-CoA as a substrate, adding two carbons per cycle to synthesize acyl carbon chains to synthesize saturated fatty acids with 16 and 18 carbons [38]. Furthermore, malonyl-CoA is also a substrate for flavonoid biosynthesis. Many genes are responsible for flavonoid biosynthesis and regulation in plants [39–41]. TRANSPARENT TESTA 4 (TT4) catalyzes the conversion of malonyl-CoA and 4-coumaroyl-CoA to chalcone in the first step of the flavonoid biosynthesis pathway [42, 43]. Mutation of *TT4* leads to a dramatic decrease in flavonoids and about 14% increase in SOC in *Arabidopsis* [43]. It has also been reported that knockout of *TT2* or *TT8* in *B. napus* using CRISPR/Cas9 technology results in a yellow-seeded phenotype with a significant increase in oil content and a significant decrease in flavonoids in seeds [44–46]. Yellow-seeded seed has less seed coat content (SCC), higher meal protein content, and higher SOC than black-seeded seed [47, 48], and there is a huge economic value for yellow-seeded *B. napus* breeding.

In this study, a comprehensive metabolome blueprint has been drawn in *B. napus* seed. By integrating multi-omics data including genome, transcriptome, and metabolome, we dissected the genetic basis of SOC in *B. napus* by performing mGWAS, mTWAS, eGWAS, and WGCNA analysis. We identified 131 SOC-correlated marker metabolites and identified 445 metabolite quantitative trait loci (mQTL) of these marker metabolites. SOC-correlated modules were subsequently characterized, and these results indicate the potential regulation relationships between genes and metabolites. Finally,

we experimentally validated three candidate genes (*BnaA03.TT4*, *BnaC02.TT4*, and *BnaC05.UK*), which impact SOC through the regulation of metabolome in the seeds.

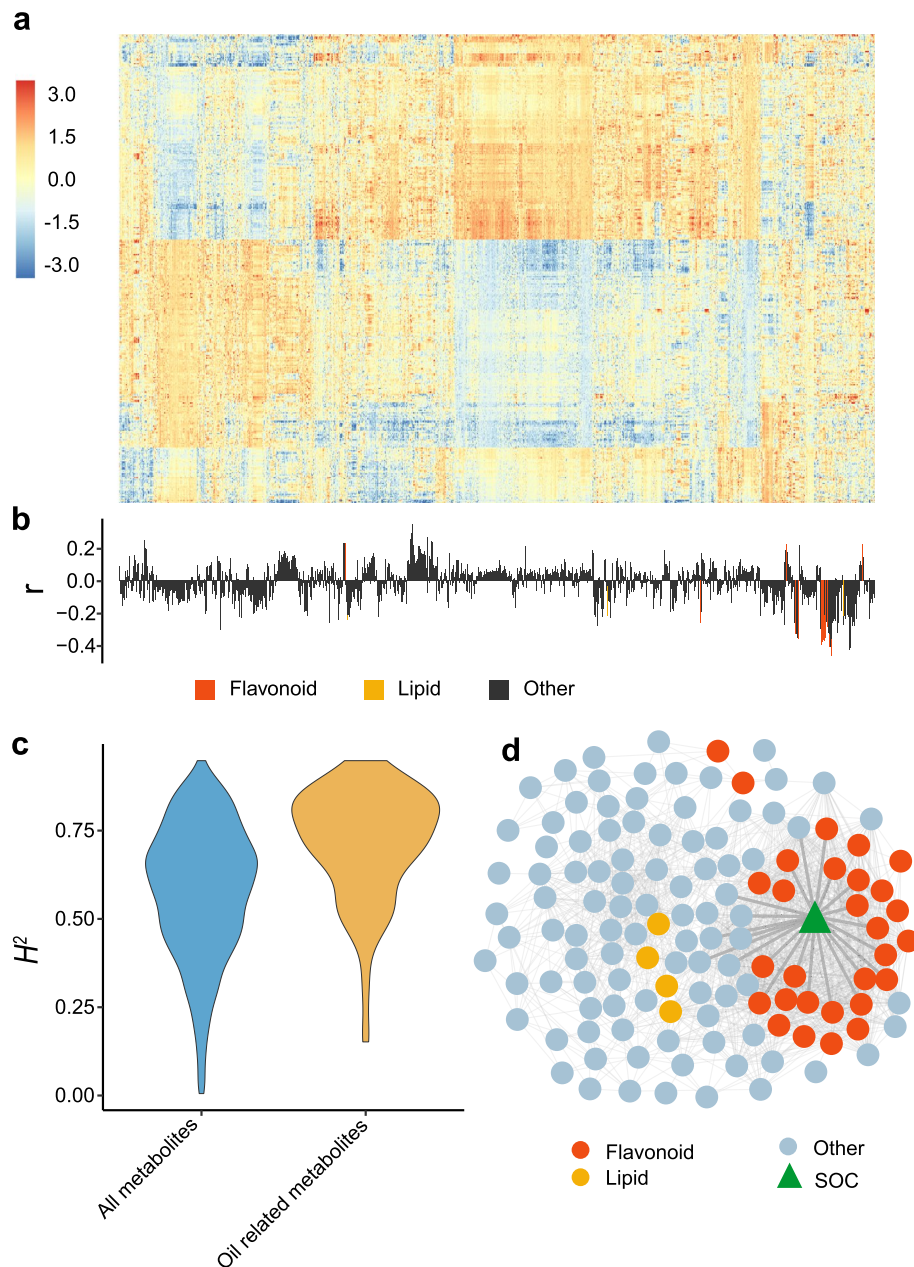
## Results

### Metabolic profiling constructs a metabolome landscape of *B. napus* seed

A total of 2173 metabolites, including 80 amino acid derivatives, 38 nucleic acid derivatives, 60 lipids, 92 flavonoids, 29 phenylethanoids, 42 terpenoids, and 99 other metabolites, have been quantified in the mature seeds of 388 *B. napus* accessions using the widely targeted metabolomics method (Fig. 1a; Additional file 1: Table S1 and S2) [19], and the metabolite content data in 2 years are listed in Additional file 1: Table S3 and Table S4. To investigate the metabolic changes affecting the SOC in *B. napus*, 131 SOC-correlated metabolites ( $|r| > 0.2$ ) are identified as marker metabolites of SOC, including 26 positively correlated metabolites and 105 negatively correlated ones (Fig. 1b), which are selected for further analysis. From the heatmap of the relative content of all metabolites, we find that most of those SOC-correlated metabolites exhibit similar accumulation patterns (Fig. 1a, b). Besides, we have calculated the broad-sense heritability of all metabolites, showing a wide variation range from 0.01 to 0.95 with a mean value of 0.58, and the SOC-correlated metabolites have a higher mean value of 0.71 (Fig. 1c). These results suggest that elevated heritability in SOC correlated metabolites by biasing the samples towards SOC variation. We have annotated 45 SOC-correlated metabolites based on standard and reference spectral library, and 30 metabolites are annotated as flavonoids and 4 metabolites are annotated as lipids (Additional file 1: Table S5). Most (27 out of 30) of the flavonoids and all 4 lipids ( $|r| > 0.20$ ) are negatively correlated with SOC (Additional file 1: Table S5, Additional file 2: Fig. S1). In addition, a hypergeometric test was conducted to rule out the possible bias of significant correlation between the SOC and the flavonoid family metabolites ( $P = 6.75 \times 10^{-16}$ , Additional file 1: Table S6). For better studying more correlated SOC markers, we use a higher threshold ( $|r| > 0.35$ ) and find that flavonoids are more correlated to SOC than other metabolites (Fig. 1d). These results suggest that they can serve as marker metabolites of SOC in *B. napus*.

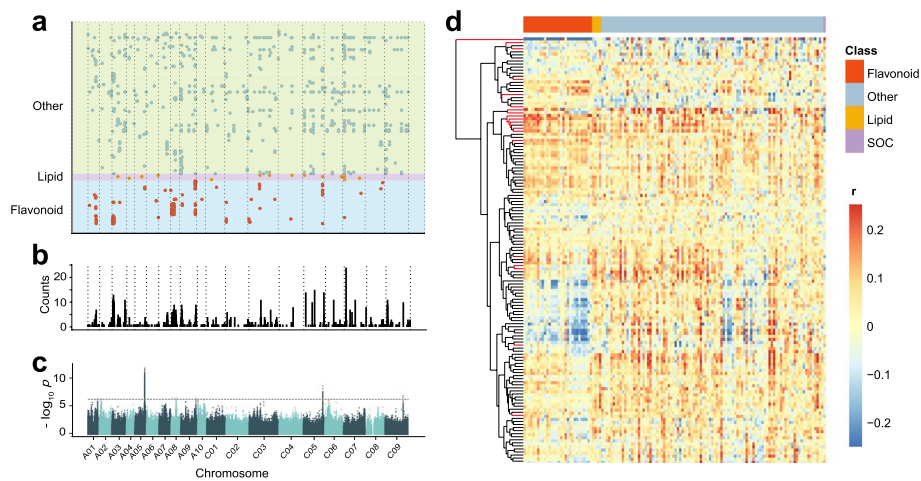
### Association analysis between metabolome, transcriptome, and variome uncovers mQTLs potentially affecting SOC

A total number of 8,274,830 and 8,288,519 variations (minor allele frequency  $> 5\%$ ) in 388 *B. napus* accessions have been selected for conducting mGWAS for SOC-correlated metabolites in 2 years respectively (2017 and 2018) [25]. A Bonferroni correction of  $p = 1.2 \times 10^{-7}$  which is described in the methods has been employed as the genome-wide threshold for 131 SOC-correlated metabolites, and a total of 1318 variation-metabolite associations for 128 metabolites are detected, including 329, 17, and 972 associations corresponding to 29 flavonoids, 4 lipids, and 95 other metabolites, respectively (Fig. 2a; Additional file 1: Table S7). All the signals can be divided into 445 mQTLs according to the LD blocks of this panel (100 kb for each block) [25], and there are obvious hotspots in the association analysis. Twelve mQTLs on chromosomes A03, C03, C05, C06, C07, and C09 are detected by more than 10 SOC-correlated metabolites (Fig. 2b; Additional file 1: Table S7). In total, six mQTLs detected by SOC-correlated metabolites are colocalized with SOC-related loci detected previously [25] (Fig. 2c; Table 1).



**Fig. 1** Profiling of metabolites in the mature seed of a *B. napus* natural population. **a** The heatmap of 2173 detected metabolites in the natural population. The horizontal axis represents the metabolites detected in each individual, and the vertical axis shows the individuals. The color of each cell represents the normalized content of the metabolite in the related individual. **b** The correlation between the detected metabolites and SOC. Each vertical line represents a metabolite that corresponds to the above column in **a**. **c** Distribution of the values of broad-sense heritability ( $H^2$ ) of 2173 metabolites and 131 SOC-correlated metabolites ( $p = 2.2 \times 10^{-16}$ ). **d** Network of metabolites-metabolites and SOC-metabolites ( $|r| > 0.35$ ), the connections between metabolites and SOC are bolded

In our previous study, we obtained the population transcriptome data 40 days after flowering (DAF) developing seeds [25]. To better understand the linkage between transcripts and SOC-correlated metabolites, we have performed mTWAS using the 40 DAF transcriptome dataset with 70,781 genes. mTWAS identifies 7316 genes



**Fig. 2** mGWAS and module-trait association studies of SOC and SOC-correlated metabolites. **a** Chromosomal distribution of all lead SNPs associated with SOC-correlated metabolites (the results of 2017 and 2018 are combined). **b** Chromosomal distribution of all mQTLs associated with SOC-correlated metabolites (the results of 2017 and 2018 are combined). The interval size is 100 kb. **c** Manhattan plot of GWAS results of the BLUP values of SOC across 5 years in WH [25]. **d** Heatmap of correlation among SOC-correlated metabolites, SOC, and eigenvalues of modules obtained by WGCNA. Each eigenvalue represents the profile of the genes in the correlated module. The vertical axis shows the modules, and the branch in red represents the modules significantly correlated with SOC

**Table 1** Summary of the hotspot loci detected by mGWAS co-localized with SOC-related loci [25]

mQTL <sup>a</sup>	Chromosome	Position	Metabolites <sup>b</sup>	Environment (SOC) <sup>c</sup>
182	A01	21,217,005–21,240,857	S21_S-4234	W3/H4/H6/A6
422	A03	2,231,384–2,368,373	S21_S-0148, S21_S-0470, S21_S-1706, S21_S-3662, S21_S-0649, S21_S-0968, S21_S-3600, S21_L-0232, S21_S-1022, S21_S-2354, S21_S-3675	W4/H5/H6/A6
956	A05	20,294,478–20,328,897	mr1077, mr1203, S21-4865	W1/W2/W3/W4/W5/W6/H4/H5/A6/H6
1507	A08	10,480,586–10,489,644	S21-1083, S21-1235	W2/W6/A6
1863	A09	31,177,191–31,257,536	S21_L-0942, mr1103, mr1226, mr1227, mr1228, mr1082, mr1083	W2/W6/A6
3947	C05	39,949,109–40,027,927	S21-1612, wm0013, mr4076, wm0031, mr002, mr1263, S21-4865, mr1204, S21_S-1264, S21-4990, mr4064	W1/W2/W3/W4/W6/C4/A6

<sup>a</sup> The IDs of the QTL detected by SOC-correlated metabolites

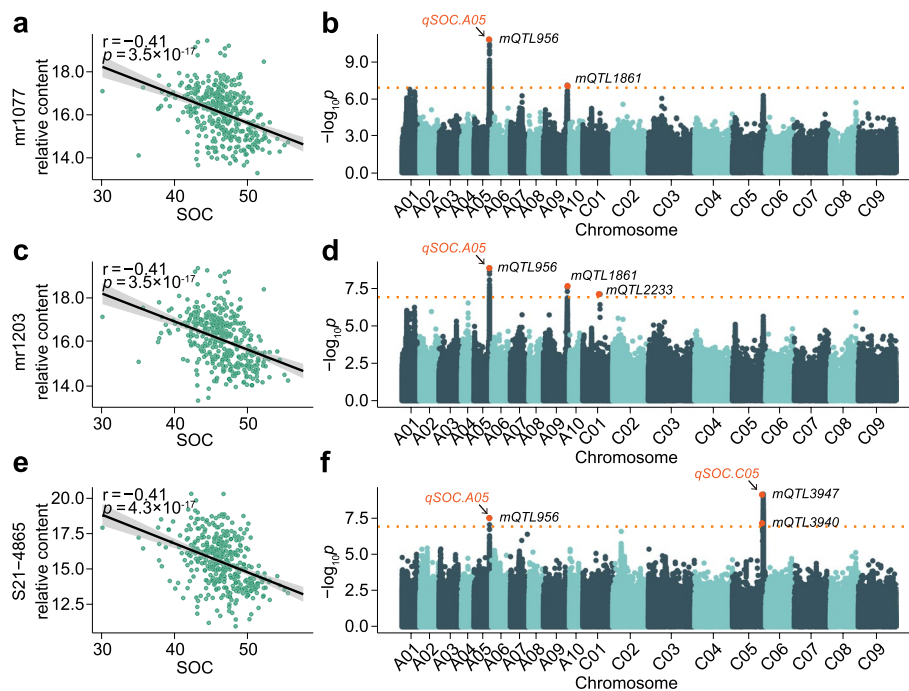
<sup>b</sup> The IDs of the SOC metabolites

<sup>c</sup> Previous GWAS study on SOC of *B. napus* [25]. Environments in this study. WH, W; HF, H; CD, C; 2013, 1; 2014, 2; 2016, 3; 2017, 4; 2018, 5; BLUP-predicted oil content in WH, W6; BLUP-predicted oil content in HF, H6; BLUP-predicted oil content in three locations, A6

having significant association with SOC-correlated metabolites ( $p < 1.41 \times 10^{-5}$ ; Additional file 1: Table S8). WGCNA has been conducted for all these transcripts associated with SOC-correlated metabolites, which are clustered into 139 modules (Additional file 1: Table S9), and the correlation coefficient of module eigengenes to

SOC-correlated metabolites and SOC are then calculated (Fig. 2d; Additional file 1: Table S10). Among all of the modules, 19 modules are significantly correlated with SOC, especially module 15 (size = 946,  $r = 0.41$ ,  $p = 6.58 \times 10^{-11}$ ). The GO enrichment analysis indicates that module 15 is enriched in the flavonoid biosynthetic and long-chain fatty-acyl-CoA metabolic processes (Additional file 2: Fig. S2a). In addition, module 27, module 43, and module 136 are also enriched in the lipid metabolic and seed development processes (Additional file 2: Fig. S2b-S2d).

GWAS for SOC suggests a significant associative peak ( $qSOC.A05$ ) on chromosome A05 (position = 20,328,269,  $p = 2.64 \times 10^{-8}$ ) which is also identified in our previous study [25]. Interestingly, this locus is also detected by Chrysoeriol C-hexoside (lead SNP  $p = 1.83 \times 10^{-11}$ ), methylLuteolin C-hexoside (lead SNP  $p = 1.37 \times 10^{-9}$ ), and an unknown metabolite (lead SNP  $p = 2.85 \times 10^{-8}$ ) (Additional file 1: Table S7). These three metabolites are negatively correlated with SOC (Fig. 3a–f). In addition, it is interesting that a SOC-related locus is detected on chromosome A09 (position = 31,257,536,  $p = 4.04 \times 10^{-7}$ ), and this position is also co-localized with a hot block detected by 7 SOC-correlated flavonoids (Additional file 2: Fig. S3a-3n). According to the above six co-localized loci, six lead SNPs are used to train a random forest model for predicting the SOC levels. The SOC level of each training and testing set is calculated based on the SOC under five environments. The training set including 108 low SOC individuals (range from 29.4 to 37.8%), 189 medium SOC individuals (range from 37.6 to 40.6%), and 55 high SOC individuals (range from 41.0 to 47.6%) is utilized to train the



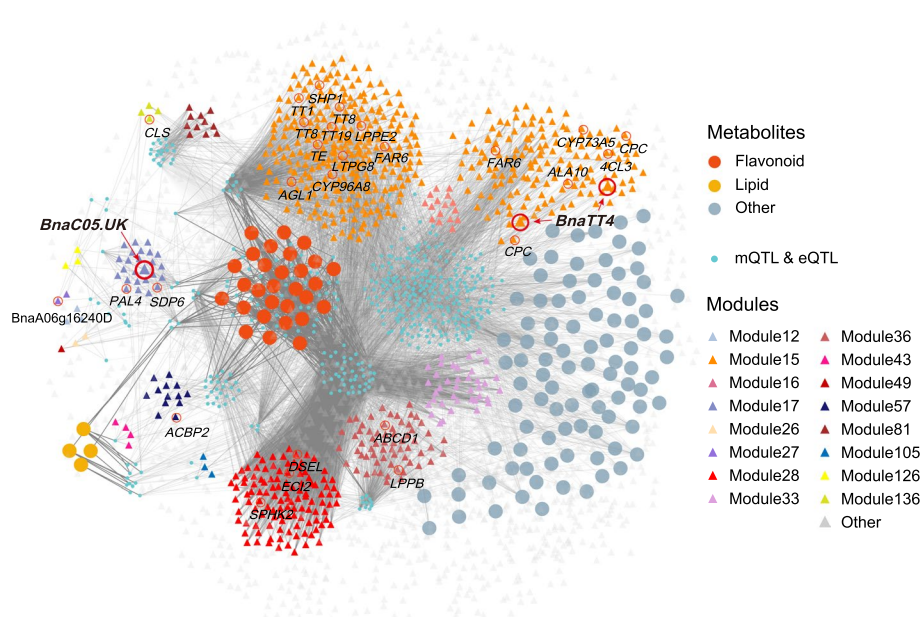
**Fig. 3** mGWAS results related to the loci on chromosome A05 (position = 20,328,269 bp) (2017). **a** Correlation analysis between SOC and mr1077 (Chrysoeriol C-hexoside). **b** Manhattan plot of mGWAS for mr1077. **c** Correlation analysis between SOC and mr1203 (methylLuteolin C-hexoside). **d** Manhattan plot of mGWAS for mr1203. **e** Correlation analysis between SOC and S21-4865. **f** Manhattan plot of mGWAS for S21-4865.  $qSOC.A05$  and  $qSOC.C05$  are two stable SOC-related QTLs identified in our previous study [25]

model (Additional file 2: Fig. S4a-S4c). The test set includes 47 low SOC individuals, 82 medium SOC individuals, and 23 high SOC individuals which are utilized for testing the predictive accuracy of the models trained by the lead SNPs and random SNPs. The test area under curves (AUCs) for the model trained with the peak SNPs are 0.61 for the low SOC, 0.54 for the medium SOC, and 0.72 for the high SOC individuals. The mean AUCs for the model trained with the random SNPs are 0.52 for the low SOC, 0.51 for the medium SOC, and 0.52 for the high SOC individuals (Additional file 2: Fig. S4d-S4f). The results indicate that these six peak SNPs could be used for marker-assisted selection of high SOC individuals, and these findings will facilitate the future *B. napus* breeding.

To uncover the relationship between genomic polymorphism and gene expression, 7316 genes associated with SOC-correlated metabolites are selected for eQTL calculation in this panel (Additional file 1: Tables S8, S11). As a result, 117,782 lead SNPs are significantly associated with the selected genes, and those SNPs are united into 5242 eQTLs, if their physical distance is < 100 kb (Additional file 1: Table S8). Among all the eQTLs, some eQTLs on chromosomes A03, A08, A09, and C09 are associated with more than 200 genes (Additional file 2: Fig. S5). It suggests that there might be master regulators in those eQTLs to regulate a set of genes related to SOC-correlated metabolites, directly or/and indirectly influencing SOC.

### Triple relationship network discovers promising genes for improving SOC

To gain a full comprehension of the genome, transcriptome, and variome, the mGWAS results and transcriptome analysis results have been integrated into a triple relationship network (metabolite-QTL-gene) (Fig. 4; Additional file 1: Table S12), which includes 444 QTLs identified simultaneously by both metabolites and transcripts, 3813 genes, and 127 SOC correlated metabolites. According to the triple relationship network and the results of mTWAS, we have selected 240 candidate genes based on gene functional annotation, which may be related to SOC accumulation (Additional file 1: Table S13). For example, Shatterproof (*SHPI*) is a MADS-BOX transcript factor (TF), along with *SHP2* control the lignification of dehiscence zone cells in plants [49]. *SHPI* and *SHP2* are essential for ovule development [50]. In our study, the transcript level of *BnaA04.SHP1* (BnaA04g01810D) is significantly negatively correlated with SOC ( $r = -0.32$ ,  $p = 2.2 \times 10^{-7}$ , Additional file 2: Fig. S6a). Interestingly, mTWAS results indicate that *BnaA04.SHP1* is significantly associated with 44 metabolites negatively correlated with SOC (Additional file 1: Table S13). eGWAS suggests the expression level of *BnaA04.SHP1* is associated with two *cis*-eQTLs and six *trans*-eQTLs (Additional file 2: Fig. S6b-S6c). Haplotype analysis shows that haplotype N has a higher expression level of *BnaA04.SHP1* ( $p = 2.0 \times 10^{-11}$ , Additional file 2: Fig. S6d). In addition, CAPRICE (*CPC*) is an MYB TF, which positively regulates the root development [51] and represses the biosynthesis of anthocyanin in plants [52]. In our study, the transcript level of *BnaA05.CPC* (BnaA05g01400D) is negatively correlated with SOC ( $r = -0.15$ ,  $p = 0.015$ , Additional file 2: Fig. S7a). mTWAS results showed that *BnaA05.CPC* is significantly associated with 33 SOC-correlated metabolites (Additional file 1: Table S13). eGWAS suggests the expression level of *BnaA05.CPC* is associated with one *cis*-eQTL and five *trans*-eQTLs (Additional file 2: Fig. S7b-S7c). Haplotype analysis shows that haplotype T has a higher expression level of *BnaA05.CPC* ( $p = 8.2 \times 10^{-9}$ , Additional file 2: Fig. S7d). It



**Fig. 4** Network built based on the correlation among SOC-correlated metabolites, genes, and QTLs. Genes are shown as reverse arrows with a distinct color per co-expression module. All modules significantly correlated with SOC were shown in this network, except for module 68, module 134, and module 135. The QTLs were detected simultaneously by the significant associated genes (eQTL) and metabolites (mQTL) in this triple relationship network

indicates that there are potential regulation networks of *BnaA04.SHP1* and *BnaA05.CPC* individually, and these two genes could be responsible for the biosynthesis of flavonoids and influencing SOC in *B. napus*.

***BnaTT4* negatively impacts SOC as a key enzyme of flavonoid biosynthesis in *B. napus* seed**  
Chalcone synthase (TRANSPARENT TESTA 4, TT4) is the first step reaction of flavonoid biosynthesis [53]. mGWAS for mr1600 (luteolin C-sinapoyl hexoside) identifies a significant SNP on chromosome A03 (*mQTL421*, position = 2,231,384,  $p = 7.39 \times 10^{-14}$ , Additional file 1: Table S7). We find *BnaA03.TT4* (BnaA03g04590D) is in the mQTL region (Additional file 2: Fig. S8a-S8b). Similarly, mGWAS for mr1976 suggests a significant SNP on chromosome C02 (*mQTL2396*, position = 2,740,593,  $p = 1.16 \times 10^{-8}$ , Additional file 1: Table S7) and *BnaC02.TT4* (BnaC02g05070D) is likely the candidate gene (Additional file 2: Fig. S8c-S8d). mTWAS analysis suggests that three *BnaTT4* homologs are significantly associated with over 40 metabolites, and the majority of the associated metabolites are flavonoids (Additional file 2: Fig. S8e). *BnaTT4* homologous genes in *B. napus* show significant correlations with SOC and flavonoid metabolites in our population, especially BnaC09g43250D, BnaA10g19670D, and BnaC02g05070D are also included in module 15 (Figs. 4 and 5a; Additional file 2: Figs. S9a and S9b). Thus, we validated the function of *BnaTT4* using mutants generated by CRISPR/Cas9, which is homozygous mutants ( $T_2$ ) of BnaA02g30340D, BnaC02g05070D, BnaA03g04590D, BnaC03g06120D, BnaA10g19670D, and BnaC09g43250D (Additional file 1: Table S14). Naringenin chalcone is the direct production of TT4, which is dramatically decreased in the *BnaTT4* mutant, as well as naringenin and Chrysoeriol (Fig. 5b). *BnaTT4* mutant lines

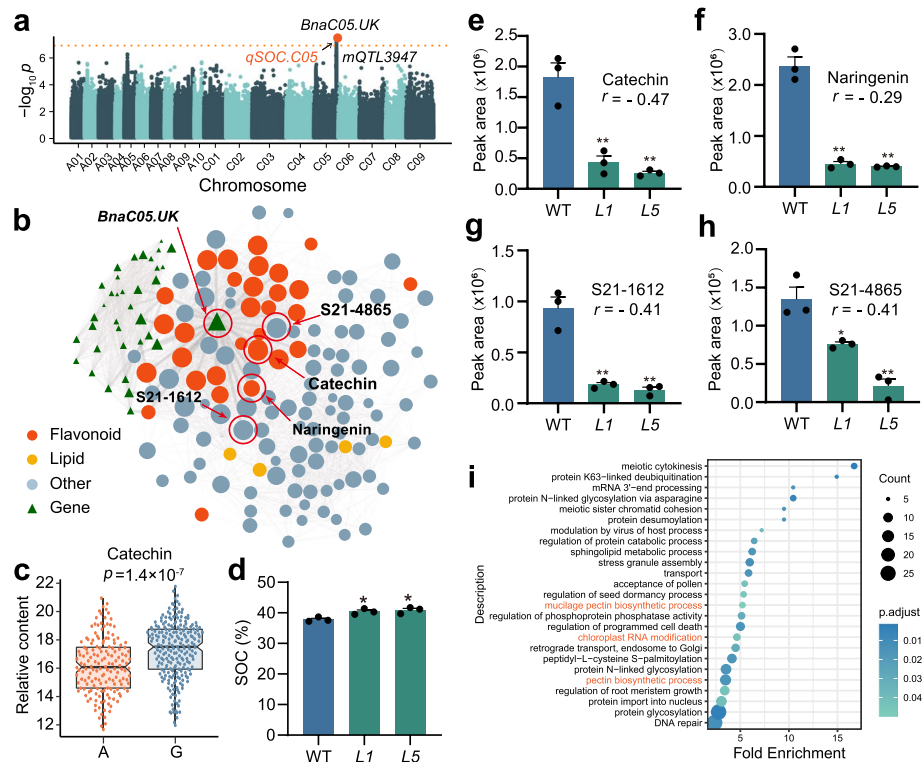




BnaC09g43250D,  $p = 8.72 \times 10^{-12}$ ; BnaA10g19670D,  $p = 5.46 \times 10^{-8}$ ; BnaA03g04590D,  $p = 5.79 \times 10^{-6}$ ; BnaC03g06120D,  $p = 9.29 \times 10^{-6}$ , Additional file 1: Table S13). Compared to WT, catechin is barely existent in *L53* ( $Bnatt4/WT = 0.05$ , Fig. 5b). Similarly, Chrysoeriol is barely existent in the mutants ( $Bnatt4/WT = 0.09$ , Fig. 5b), almost all the detected flavonoids exhibit a similar pattern. We have surveyed *BnaTT4s* mutant growth phenotypes, except for the thousand seed weight is significantly decreased, we have not observed any yield penalties (Additional file 1: Table S15). The results suggest that the synthesis of flavonoids is blocked in the *BnaTT4s* mutants and the carbon source may shift to seed oil synthesis when *BnaTT4s* are interrupted.

### *BnaC05.UK* is a novel gene negatively impacting SOC in *B. napus*

We find an mQTL hotspot with linked variations on chromosome C05, including *mQTL3946* and *mQTL3947*, which are associated with 16 SOC correlated metabolites (Fig. 6a; Additional file 2: Fig. S10a-S10t; Additional file 1: Table S7). Interestingly, this hotspot is co-localized with the SOC-related QTL identified in our previous study [25]. mGWAS for catechin suggests a significant SNP signal on chromosome C05 (Fig. 6a, position = 40,022,989,  $p = 5.66 \times 10^{-8}$ ). It belongs to the *mQTL3947* which is detected by three flavonoids and is co-localized with an SCC-related QTL



**Fig. 6** Functional study of *BnaC05.UK*. **a** Manhattan plot of mGWAS results of mr002 (2017). **b** Association analysis of genes in module 17 with SOC-correlated metabolites ( $|r| > 0.30$ ). The absolute values of Pearson's correlation coefficient values are represented by the thickness of lines. **c** Haplotype analysis of *mQTL3947* for mr002 content. **d** SOC of WT and mutant seeds. **e** Peak area of mr002, **f** naringenin, **g** S21-1612, and **h** S21-4865 in WT and mutant seeds. **i** Enrichment analysis for genes that are significantly different between WT and L5. Values are means  $\pm$  s.e.m.,  $n = 3$ . Statistical analysis is using Student's *t*-test ( $*p < 0.05$ ;  $**p < 0.01$ )

detected in our previous study [46] (Additional file 1: Table S7 and Additional file 2: Fig. S11a-S11d). Gene module analysis indicates that module 17 is not only significantly correlated with SOC but also shows significant correlations with lipid and flavonoid metabolites (Fig. 6b; Additional file 1: Table S10). We predicted a gene *BnaC05.UK* (BnaC05g43050D) with unknown biological functions as the candidate gene in *mQTL3947*. Because *BnaC05.UK*'s expression is significantly associated with 9 metabolites including catechin in mTWAS analysis (Additional file 1: Table S8). Importantly, *BnaC05.UK* is one of the core genes in module 17 (Fig. 6b). eGWAS reveals that there is a promoter region variation maybe responsible for the metabolite and SOC variations (Additional file 1: Table S8 and Additional file 2: Fig. S12a-S12c). Moreover *BnaC05.UK* is significantly negatively correlated with SOC ( $r = -0.36$ ,  $p = 9.5 \times 10^{-9}$ , Additional file 2: Fig. S12d). Haplotype analysis suggests that *BnaC05.UK* with haplotype G represent higher catechin (Fig. 6c, SOC  $r = -0.47$ ,  $p = 1.4 \times 10^{-7}$ ). We built a phylogenetic tree of *BnaC05.UK* using its cDNA sequences. There are seven sequences out of 12 belonging to *Brassica* (Additional file 2: Fig. S12e). In addition, we analyzed the basic information about *BnaC05.UK*, and its open reading frame length is 1348 bp including one exon (276 bp encoding 91 amino acids) in the Darmor genome [54]. We also analyzed the sequence of *BnaC05.UK* in the pangome of *B. napus* in BnTIR [55] (<http://yanglab.hzau.edu.cn/BnTIR>, Additional file 2: Fig. S12f). There is also no typical domain for *BnaC05.UK* analyzed by InterPro [56] (<http://www.ebi.ac.uk/interpro/>). Furthermore, there are no studies reporting the function of *BnaC05.UK* or its homologous gene.

Our multi-omics analysis suggests that *BnaC05.UK* plays important functions in affecting the level of many metabolites and may affect SOC in *B. napus*. In order to investigate the function of *BnaC05.UK*, we generated *BnaC05.UK* homozygous mutants ( $T_2$ ) using *Westar* and the mutation lines show significantly higher SOC of about 2.6% ( $L1$ ) and 3.1% ( $L5$ ) than WT (Fig. 6d; Additional file 1: Table S14). As we mentioned before, *mQTL3947* is detected by mGWAS for 16 metabolites. Metabolome analysis suggests that the level of catechin, naringenin (SOC  $r = -0.29$ ), S21-1612 (SOC  $r = -0.41$ ), and S21-4865 (SOC  $r = -0.41$ ) shows dramatic reduction (Fig. 6e-h). In addition, we performed RNA-seq using WT and  $L5$  developing seeds (37 DAF), 706 genes were upregulated, and 62 genes are downregulated. Interestingly, differentially expressed genes (DEGs) are enriched in the pectin biosynthetic process, mucilage pectin biosynthetic process, chloroplast RNA modification, and others (Fig. 6i). Among the DEGs, chloroplast RNA modification-related genes are also significantly changed (Additional file 2: Fig. S13a-S13e). In addition, the *Rhamnogalacturonan 1 rhamnosyltransferase 1* [57], *Probable galacturonosyltransferase 5* [58], and *Probable pectin methyltransferase QUA3* [59] genes have been reported to be involved in the formation of pectin and seed coat mucilage. The expressions of these genes in developing seed of mutation lines are significantly increased (Additional file 2: Fig. S13f-S13k). Three oil accumulation-related genes have increased expression levels (Additional file 2: Fig. S13l-S13n). We have also found that the mutants have more leakage of the seed coat mucilage in mature seeds compared with WT (Additional file 2: Fig. S13o). *BnaC05.UK* is localized in the chloroplast (Additional file 2: Fig. S13p), we speculate that the mutation of this gene may lead to changes in chloroplast

RNA levels and affect the accumulation of fatty acids, flavonoids, and pectin [60]. We have also surveyed *BnaC05.UK* mutant growth phenotypes, we have not observed any yield penalties (Additional file 1: Table S16).

## Discussion

Metabolites are considered as important markers for predicting traits, for example, sugar, and amino acid content could predict the accumulation of photosynthate at the final stage [7, 8, 61]. GWAS as well as TWAS have been proven to be efficient tools for genetic dissection [25–27, 62–64]. In our study, we intentionally screened 131 marker metabolites correlated to SOC, and multi-omics analysis efficiently located multiple loci potentially controlling SOC. Many SOC-associated loci (Table 1) were found to be co-localized with the ones located in our previous study [25], showing good reliability and repeatability of these two studies. The association studies and network analysis also provide a reference for future genetic improvement of *B. napus*. *BnaA05.PMT6* and *BnaC05.PMT6* were successfully cloned from two SOC-related loci *qSOC.A05* and *qSOC.C05* [25]. However, in addition to *BnaC05.PMT6*, we also found *BnaC05.UK* is a candidate gene regulating SOC in *qSOC.C05*, suggesting that there may be two or more genes regulating SOC in *qSOC.C05*. Furthermore, our study confirmed the effect of flavonoid metabolism on SOC, especially catechin and other flavonoid metabolites through functional verification of candidate genes (Figs. 5 and 6). Therefore, our study provided an effective approach for mining marker metabolites and candidate genes for agronomic traits in crops.

The huge metabolite diversity has a powerful role to explore the genetic diversity combined with multi-omics studies, and we discovered novel loci related to SOC by mGWAS. We have shown that SOC-correlated metabolites detected more signals by mGWAS (Fig. 2a–c), which enriches the information on our association population's genetic diversity for future reference. Using TWAS to dissect traits have been reported in humans [65, 66], animals [67], and plants [20, 26, 27]. In this study, TWAS has been performed using the 40 DAF population transcriptome data of *B. napus*, and united with metabolomics, we propose the concept of mTWAS for the first time. Compared with mGWAS, mTWAS tend to use the expression level of a certain gene, eliminating the uncertainty of linkage disequilibrium interval and screening candidate genes more accurately and directly [68, 69]. Our previous study comprehensively characterized the transcriptional variability in developing seeds of *B. napus* and successfully cloned two transcription factors involved in the regulation of SOC from eQTL hotspots87-88 [70]. The metabolome variation data has brought more information which helped predict novel genes regulating SOC by marker metabolite-based multi-omics analysis. In addition, we have constructed a triple relationship network. Combining co-expression modules and the relationship networks, it is worth looking forward to improving the prediction of crop agronomic traits and assisting crop breeding with marker metabolites in the future.

To validate the results of the meta-analysis, we have functionally analyzed two genes *BnaA03.TT4* and *BnaC02.TT4*, and *BnaTT4s* mutants have increased SOC by 4.8% and 5.6% (Fig. 5d). We found that all 26 flavonoids are associated with *TT4* expression in our mTWAS results (Additional file 1: Table S8). This may be due to

the key biological function of TT4 at the first and rate-limiting step in the flavonoid synthesis pathway [43]. These results reflect the effectiveness of our research strategy. We have also identified a novel gene *BnaC05.UK* as a negative regulator of SOC in *B. napus*, because the mutation of *BnaC05.UK* causes a 2.6% and 3.1% increase in SOC (Fig. 6d). *BnaC05.UK* is in the hotspot of metabolites, SOC, and SCC; 16 metabolites are co-localized in *mQTL3947* that are negatively correlated with SOC (Additional file 2: Figs. S10-S11). The content of most of the negative marker metabolites was significantly decreased, explaining the increased SOC in the mutants (Fig. 6e-h). Combined with the transcriptome results of developing mutant seeds, the expression levels of pectin mucilage-related genes were upregulated, and mature mutant seeds have more mucilage leakage than the wild type (Additional file 2: Fig. S13o). The increased SOC and decreased content of flavonoid metabolites in the mutants suggest that *BnaC05.UK* positively regulates flavonoids and negatively regulates oil accumulation. Since flavonoid synthesis is blocked in mutants, excess carbon sources may flow towards fatty acid and seed coat mucilage synthesis. More inspiringly, *BnaC05.UK* is expressed in the chloroplast (Additional file 2: Fig. S13p), where fatty acids are de novo synthesized [38, 71]. The expression of genes related to seed RNA editing in the chloroplast changes significantly during the development stage of the mutant seed, which may cause the disorder of RNA level in the chloroplast [60]. The expression levels of genes related to fatty acid synthesis are also significantly increased in *Palmitoyl-acyl carrier protein thioesterase* [72], *Ethylene-responsive transcription factor ABI4* [73], etc., which is also consistent with the increased SOC phenotype of the mutants (Additional file 2: Fig. S13l-S13n). Pangenome sequence analysis indicates that most of the reference genomes present one homolog of *BnaC05.UK* except for No2127. No2127 is a resynthesized *B. napus* cultivar, and its reference genome presents two homologs (Additional file 2: Fig. S12f). How *UK* gene affects the synthesis of flavonoids, fatty acids, pectin, and the biological significance of their derived evolution needs further study. TT4 catalyzes the first step of flavonoid synthesis; it starts with one molecule of coumaroyl CoA and three molecules of malonyl CoA [43]. In fatty acid synthesis, malonyl CoA is the substrate of fatty acid chain elongation [38]. Fatty acid synthesis and flavonoid synthesis compete with each other for malonyl CoA (Additional file 2: Fig. S14). Published studies have proved that blocking the genes related to flavonoid synthesis will lead to an increase in SOC [44–46]. Of the 30 flavonoids significantly correlated to SOC in this study, 27 flavonoids are significantly negatively correlated to SOC (Additional file 1: Table S5). The mGWAS and mTWAS mapping results of these metabolites also predicted *TT1*, *TT4*, *TT5*, *TT8*, *TT19*, and other known genes to be involved in flavonoid metabolism (Additional file 1: Table S13). Based on multi-omics analysis and using metabolite phenotype as a bridge, we provided direct genetic evidence to confirm the significant negative correlation between oil content and flavonoid synthesis genes (Additional file 1: Table S13). In order to explore the relationship between phenylpropanoids and fatty acid synthesis, we predicted the potential regulatory transcription factors of *BnaTT4s* and *BnaC05.UK* by multi-omics analysis and integrated them into a schematic diagram (Additional file 1: Table S17; Additional file 2: Fig. S14). Behind the simple shuffling carbon phenomenon, it usually exists more complex regulatory relationships within

flavonoids/phenolics and SOC, such as *TT2* [74] and *TT8* [75]. Whether *TT4* as a key enzyme participates in a more complexed upstream regulation network needs further investigation.

Although we primarily focused on 131 SOC-correlated metabolites, most of which were flavonoids (Additional file 1: Table S5), and the rest were also worthy of investigation. Thousand seed weight (TSW) is also an important component of *B. napus* yield [76]. It is vital to study the genetic variation of TSW in *B. napus* [77]. Combined with the metabolome map provided here, it is possible to mine more unreported TSW-related genes for breeding. *B. napus* is an important oil-producing crop, and the remaining meal can also be used as animal feed because of its rich protein [47, 48]. Therefore, the research on the protein content of the *B. napus* seeds is also of great value [78, 79]. Our metabolome data included amino acid metabolites, such as L-lysine (mr1330) and L-valine (mr1322) (Additional file 1: Table S5), which will facilitate future research on improvements for the protein content of *B. napus* seeds. In addition, we also quantified many health-related metabolites, such as melatonin (mr2076, Additional file 1: Table S2). It is proven to be effective for insomnia, mood swing, drug abuse, and even cancer treatment [80, 81]. We can unearth certain reliable loci controlling melatonin in *B. napus* [82].

It is worth noting the limitation of our extraction method, which has been previously reported to have excellent analytical confidence for water-soluble metabolites including flavonoids, phenolamines, terpenoids, amino acids, and nucleic acid derivatives [10, 15, 20]. Our study provides rich information on metabolomic diversity in *B. napus* seeds. However, most structures of metabolites remain unknown. Metabolite identification using mass spectrometry tandem chromatography (LC–MS/MS) requires high consistency of experimental conditions [83, 84]. In addition, based on the tight relationship between enzymes and metabolites, the mQTL localization can also be considered as auxiliary information for metabolite analysis [15]. Recently, utilizing software based on machine learning to resolve metabolite structures has become well-welcomed [85–87]. Comprehensive chemical standard identification combined with artificial intelligence annotation will enable large-scale annotation of the metabolome data in the future.

## Conclusions

We performed a metabolomics analysis of *B. napus* mature seeds and obtained 131 SOC-correlated marker metabolites. Based on genome, transcriptome, and metabolome data, we discovered novel genes such as *BnaC05.UK* involved in the regulation of *B. napus* SOC. This study demonstrates the advantage of marker metabolite-based multi-omics analysis to dissect the genetic basis of agronomic traits in crops.

## Methods

### Plant materials and collection of SOC phenotype data

For plant materials, 505 *B. napus* accessions were planted in Wuhan (2017, 2018) [25]. We used Foss NIRSystems 5000 near-infrared reflectance spectroscope that performed mature open-pollinated seed oil content analysis, and each accession was analyzed with 6 replications [88]. Transgenic *B. napus* plants were grown in the field for transgenic

crops at Huazhong Agricultural University. For field experiments, we followed a randomized complete block design with three replications.

### Metabolite profiling

A subpopulation was selected for profiling the metabolites in *B. napus* seeds (388 individuals harvested in 2017 and 378 individuals harvested in 2018, Additional file 1: Table S1). For each sample, 0.1 g mature *B. napus* seeds was weighted, and they were homogenized and comminuted by TissueLyser II machine (Qiagen, Germany) at 29 Hz, 60 s with 1 ml 70% methanol containing 0.1 mg/l acyclovir (internal standard). The mixture was vortexed, and the metabolites were extracted for 10 h at 4 °C. After being centrifuged at 9000 g for 10 min, all the supernatants were collected and then were pooled and filtered with a 0.22- $\mu$ m organic filter (SCAA-104; ANPEL, Shanghai, China, <http://www.anpel.com.cn/>) for LC–MS analysis.

A liquid chromatography-electrospray ionization tandem mass spectrometry system was utilized to relatively quantify the metabolites of *B. napus*. In the beginning, an MS<sup>2</sup> spectral tag (MS2T) library was established using a mixture extraction of 30 accessions, which were selected randomly from the analysis population, following the method described previously [19]. In detail, the MS2T library was established with LIT experiments which were conducted on a triple quadrupole-linear ion trap mass spectrometer (QTRAP), API 4000 Q TRAP LC/MS/MS System, in a positive ion mode and controlled by the Analyst 1.5 software (AB Sciex). In LIT experiments, MIM–EPI was carried out to screen metabolites and restricted the fragments from  $m/z$  50 to 1000 with a mass step of 0.2 Da. For each feature in the MS2T library, the accurate MS data were acquired with an Agilent 6520 accurate-mass time-of-flight (Q–TOF) mass spectrometry equipped with a dual ESI electrospray ion source in positive ion mode. The features in this MS2T library were annotated using our standards or by comparing the accurate MS data to the reference MS2 library, such as MassBank, HMDB, METLIN, and MzCloud. Quantification of metabolites was carried out using a scheduled multiple reaction monitoring method, and a total of 2173 metabolic features were detected and quantified in the mixture extraction. The 2173 metabolic features of the population were detected, quantified, and then  $\log_2$  transformed and were performed for further normalization.

### mGWAS, mTWAS, eGWAS, and co-expression analysis

For mGWAS and eGWAS, the variants were called by the meta-analysis of reads against the *B. napus* genome (*B. napus* v4.1, <http://www.genoscope.cns.fr/brassicapapus/>) as described previously [25, 54]. A total of 8,274,830, 8,288,519, 8,365,725, 8,258,336, 8,234,950 variations with minor allele frequency > 5% were selected by PLINK [89] and used to conduct mGWAS for SOC-correlated metabolites content in 2017 and 2018, SOC, SCC, and genes expression with the FaST-LMM software [90]. The significant value was set to  $1.2 \times 10^{-7}$  and finally, and the Manhattan plot was generated in the R package. For mTWAS, gene expression of 274 accessions at 40 DAF planted in Wuhan (2017) was called by the meta-analysis of RNA sequencing data, as described previously [25]. Then, the gene expression was used to conduct associative analysis with the content of SOC-correlated metabolites in *B. napus* in the same year by linear regression [91]. The significant value was set to 1/70,781. For co-expression analysis, a total of 70,781 genes

were used and analyzed with an R package named WGCNA [92]. Then, all the genes were clustered into 139 modules, with a soft-thresholding power ( $\beta=6$ ) to approximate scale-free topology within the network determined.

#### Network building for metabolome, transcriptome, and variome

The edges between SOC-correlated metabolites and genetic loci represent the significant signals of mQTLs, while the edges between genetic loci and genes represent the significant signals of eQTLs. Firstly, significant signals for 131 SOC-correlated metabolites were detected with a significant threshold value used, and the significant signals were later clustered into mQTLs according to the LD blocks of this panel (100 kb for each block). Secondly, 7316 genes related to the 131 SOC-correlated metabolites ( $p < 1.41 \times 10^{-5}$ ) were selected for conducting eGWAS; the significant signals were also clustered into eQTLs according to the LD blocks of this panel (100 kb for each block). Finally, the loci distance was calculated between mQTLs and eQTLs for linking the SOC-correlated metabolites and genes with a threshold of 100 kb utilized.

#### Model construction for SOC prediction

To achieve precise SOC classification, the SOC datasets from all five environments (2013, 2014, 2016, 2017, 2018) were collected. The inbred lines (505 individuals) were categorized into three classes based on SOC using *K*-means clustering implemented in PYTHON/SCIKIT-LEARN (*K*-means function, *n\_clusters* = 3, *random\_state* = 2). The inbred lines with high, low, and middling SOC were divided into the training set and testing set (2:1). We identified the peak SNPs according to the results of the mGWAS and co-localized QTLs of metabolites and SOC. To generate robust prediction models, random forest models to predict SOC levels were used (*n\_estimators* = 300, *max\_features* = 3) [93–95]. The training set of six peak SNPs and six random SNPs (100 times) was entered into the random forest models independently to train the prediction models. The area under the curve (AUC) value was calculated to evaluate the accuracy of the model.

#### Vector construction and plant transformation

The full-length *BnaC05.UK* CDS was cloned from accession X1072. CRISPR-Cas9 genome editing system [96] were used to generate *tt4* and *uk* mutation lines. PCR fragment and *pKSE401* vector were integrated by Golden Gate Assembly and sequenced for confirmation. Furthermore, the plant transformation of *Westar* was performed with *pKSE401* vectors to generate transgenic lines. The tissue culture method followed Dai et al. [97]. CRISPR mutant lines were comforted by sequencing, and all primers are listed in Additional file 1: Table S18.

#### SOC measurement

SOC analysis was performed on near-infrared reflectance spectroscopy with a Foss NIRSystems 5000 near-infrared reflectance spectroscope [88] and GCMS-QP2010 Ultra (SHIMADZU, Inc.) [98] using mature open-pollinated seeds. In brief, about 10 mg seed samples was weighed, adding 3 mL methanol containing 1.5% H<sub>2</sub>SO<sub>4</sub> and 0.01% butylated hydroxyl toluene (BHT) to the glass tube, then crushed the seeds. Internal standard



is heptadecanoic acid (16.2  $\mu\text{mol/mL}$ ). To extract and methylate the FA, we performed an 85 °C water bath for 2 h, using 2-mL hexane to extract fatty acid methyl esters (FAMES). Loading volume is 1  $\mu\text{L}$ , automatic injection. The instrument parameters are as follows: the injection port temperature was 230 °C; oven temperature kept at 170 °C for 1 min, temperature rises gradually (3 °C per minute) and keep at 230 °C for 3 min; and ion source temperature was 200 °C. Calculation of FA is using heptadecanoic acid.

### Subcellular localization

The *BnaC05.UK* coding region was cloned and integrated into the vector pMDC83 to express BnaC05.UK-GFP fusion protein. The primers used are listed in Additional file 1: Table S18. After the recombinant constructs were transformed into *Agrobacterium* strain GV3101, it was infiltrated into tobacco leaves with P19 for transient expression. We observed BnaC05.UK-GFP proteins in epidermal cells using confocal laser scanning microscopy (FV1200, Olympus, Japan).

### Ruthenium red staining of seed-coat mucilage

We followed the method of Li et al. [99]. In general, dry mature seeds were infiltrated with 0.05 mM EDTA (pH 8.5) for 1 h, then stained with 0.01% ruthenium red (w/v) for another 1 h and washed by ddH<sub>2</sub>O. Seed-coat mucilage was photted using an optical microscope (BX53M, Olympus, Japan).

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-023-02984-z>.

**Additional file 1: Table S1.** Summary of accessions for mGWAS. **Table S2.** The metabolites detected in the study. **Table S3.** The metabolite content in 2017. **Table S4.** The metabolite content in 2018. **Table S5.** The metabolites correlated to SOC. **Table S6.** Hypergeometric test of the 2,173 metabolites. **Table S7.** mGWAS results of 2 years SOC-correlated metabolites and 2017 SCC. **Table S8.** Summary of significant genes in mTWAS for SOC-correlated metabolites at 40 DAF. **Table S9.** The genes' correlation information of 40 DAF transcriptome. **Table S10.** The statistics of modules calculated by WGCNA. **Table S11.** Information for eQTLs. **Table S12.** The integrated network of mGWAS and eQTL. **Table S13.** The candidate genes associated with SOC from mGWAS, eQTL and mTWAS. **Table S14.** The genotype of BnaTT4 and BnaC05.UK mutant lines in the T2 generation. **Table S15.** The mean value of yield traits for the BnaTT4 CRISPR lines. **Table S16.** The mean value of yield traits for the BnaC05.UK CRISPR lines. **Table S17.** Transcriptional factors prediction of BnaTT4s and BnaC05.UK. **Table S18.** Primers used in this study.

**Additional file 2: Fig. S1.** Key metabolites structures and mass spectrum plots. **Fig. S2.** Enrichment analysis for genes in modules significantly correlated with SOC. **Fig. S3.** mGWAS results related to the loci on chromosome A09. **Fig. S4.** The performance of metabolite-based Random Forest model for predicting the SOC level. **Fig. S5.** The genome-wide distribution of eQTLs. **Fig. S6.** Genetic analysis of BnaA04g01810D. **Fig. S7.** Genetic analysis of BnaA05g01400D. **Fig. S8.** mGWAS and mTWAS results of BnaTT4s. **Fig. S9.** Correlation analysis of BnaTT4s with SOC and SOC-correlated flavonoids. **Fig. S10.** mGWAS results related to the loci on chromosome C05. **Fig. S11.** mGWAS results related to SCC and mr002. **Fig. S12.** Functional study of *BnaC05.UK*. **Fig. S13.** The relative expression of differentially expressed genes between WT and *L5*. **Fig. S14.** Phenylpropanoids and fatty acid synthesis relationship and predicted transcriptional regulation of BnaTT4s and BnaC05.UK.

**Additional file 3.** Review History.

### Peer review information

Wenjing She was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

### Review history

The review history is available as Additional file 3.

**Authors' contributions**

L.G. and W.C. designed and supervised this study. L.L., J.C., and Xiaowei.W. performed the experiments. Zhitao.T., L.L., Zengdong.T., Y.Z., and H.Z. analyzed the data. L.L. Zhitao.T. wrote the manuscript. L.G., W.C., W.W., and X.Y. revised the manuscript.

**Funding**

This work was supported by grants from the National Science Fund for Distinguished Young Scholars (32225037), Hubei Hongshan Laboratory (2021HSZD004), HZAU-AGIS Cooperation Fund (SZYJY2022008, SZYJY2021004), Fundamental Research Funds for the Central Universities (2662022ZKPY001), and Higher Education Discipline Innovation Project (B20051).

**Availability of data and materials**

The genome resequencing data of 382 *B. napus* inbred lines and gene expression data of developing seeds (40 DAF) of 274 inbred lines are available under Genome Sequence Archive (GSA; <https://bigd.big.ac.cn/gsa/>) with Bioproject ID PRJCA002835 [100] and PRJCA002836 [101]. Mass spectrometry data of seed metabolome of 382 *B. napus* inbred lines are deposited at China National Center for Bioinformation (CNCB; <https://www.cncb.ac.cn>) with Bioproject ID of PRJCA017446 [102]. No other scripts and software were used other than those mentioned in the "Methods" section. All the materials in this study are available upon request.

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Competing interests**

The authors declare no competing interests.

Received: 10 December 2022 Accepted: 8 June 2023

Published online: 19 June 2023

**References**

- Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L. Innovation - metabolite profiling: from diagnostics to systems biology. *Nat Rev Mol Cell Bio.* 2004;5:763–9.
- Obata T, Fernie AR. The use of metabolomics to dissect plant responses to abiotic stresses. *Cell Mol Life Sci.* 2012;69:3225–43.
- Sulpice R, McKeown PC. Moving toward a comprehensive map of central plant metabolism. *Annu Rev Plant Biol.* 2015;66:187–210.
- Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Törjék O, Fiehn O, Eckardt A, Willmitzer L, Selbig J, Altmann T. The metabolic signature related to high plant growth rate in *Arabidopsis thaliana*. *P Natl Acad Sci.* 2007;104:4759.
- Sulpice R, Pyl E-T, Ishihara H, Trenkamp S, Steinfath M, Witucka-Wall H, Gibon Y, Usadel B, Poree F, Piques MC, et al. Starch as a major integrator in the regulation of plant growth. *P Natl Acad Sci.* 2009;106:10348.
- Bijlsma S, Bobeldijk I, Verheij ER, Ramaker R, Kochhar S, Macdonald IA, van Ommen B, Smilde AK. Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. *Anal Chem.* 2006;78:567–74.
- Carreno-Quintero N, Acharjee A, Maliepaard C, Bachem CW, Mumm R, Bouwmeester H, Visser RG, Keurentjes JJ. Untargeted metabolic quantitative trait loci analyses reveal a relationship between primary metabolism and potato tuber quality. *Plant Physiol.* 2012;158:1306–18.
- Do PT, Prudent M, Sulpice R, Causse M, Fernie AR. The influence of fruit load on the tomato pericarp metabolome in a *solanum chmielewskii* introgression line population. *Plant Physiol.* 2010;154:1128–42.
- Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE. Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *P Natl Acad Sci.* 2012;109:8872–7.
- Wen WW, Li D, Li X, Gao YQ, Li WQ, Li HH, Liu J, Liu HJ, Chen W, Luo J, Yan JB. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat Commun.* 2014;5:3438.
- Sun F, Liu J, Hua W, Sun X, Wang X, Wang H. Identification of stable QTLs for seed oil content by combined linkage and association mapping in *Brassica napus*. *Plant Sci.* 2016;252:388–99.
- Luo Z, Wang M, Long Y, Huang Y, Shi L, Zhang C, Liu X, Fitt BDL, Xiang J, Mason AS, et al. Incorporating pleiotropic quantitative trait loci in dissection of complex traits: seed yield in rapeseed as an example. *Theor and Appl Genet.* 2017;130:1569–85.
- Delourme R, Falentin C, Huteau V, Clouet V, Horvais R, Gandon B, Specel S, Hanneton L, Dheu JE, Deschamps M, et al. Genetic control of oil content in oilseed rape (*Brassica napus* L.). *Theor and Appl Genet.* 2006;113:1331–45.
- Liu S, Fan C, Li J, Cai G, Yang Q, Wu J, Yi X, Zhang C, Zhou Y. A genome-wide association study reveals novel elite allelic variations in seed oil content of *Brassica napus*. *Theor and Appl Genet.* 2016;129:1203–15.
- Chen W, Gao YQ, Xie WB, Gong L, Lu K, Wang WS, Li Y, Liu XQ, Zhang HY, Dong HX, et al. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat Genet.* 2014;46:714–21.
- Chen W, Wang WS, Peng M, Gong L, Gao YQ, Wan J, Wang SC, Shi L, Zhou B, Li ZM, et al. Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat Commun.* 2016;7:12767.

17. Matsuda F, Nakabayashi R, Yang ZG, Okazaki Y, Yonemaru J, Ebana K, Yano M, Saito K. Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *Plant J.* 2015;81:13–23.
18. Wen W, Li K, Alseekh S, Omranian N. Genetic determinants of the network of primary metabolism and their relationships to plant performance in a maize recombinant inbred line population. *Plant Cell.* 2015;27:1839–56.
19. Chen W, Gong L, Guo Z, Wang W, Zhang H, Liu X, Yu S, Xiong L, Luo J. A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol Plant.* 2013;6:1769–80.
20. Zhu GT, Wang SC, Huang ZJ, Zhang SB, Liao QG, Zhang CZ, Lin T, Qin M, Peng M, Yang CK, et al. Rewiring of the fruit metabolome in tomato breeding. *Cell.* 2018;172:249–+.
21. Zhan C, Lei L, Liu Z, Zhou S, Yang C, Zhu X, Guo H, Zhang F, Peng M. Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. *Nat Plants.* 2020;6:1447–54.
22. Chen J, Hu X, Shi T, Yin H, Sun D, Hao Y, Xia X, Luo J, Fernie AR, He Z, Chen W. Metabolite-based genome-wide association study enables dissection of the flavonoid decoration pathway of wheat kernels. *Plant Biotechnol J.* 2020;18:1722–35.
23. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, Jansen R, de Geus EJC, Boomsma DI, Wright FA, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet.* 2016;48:245–52.
24. Gusev A, Mancuso N, Won H, Kousi M, Finucane HK, Reshef Y, Song L, Safi A, McCarroll S, Neale BM, et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat Genet.* 2018;50:538–48.
25. Tang S, Zhao H, Lu S, Yu L, Zhang G, Zhang Y, Yang QY, Zhou Y, Wang X, Ma W, et al. Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. *Mol Plant.* 2021;14:470–87.
26. Li Z, Wang P, You C, Yu J, Zhang X, Yan F, Ye Z, Shen C, Li B, Guo K, et al. Combined GWAS and eQTL analysis uncovers a genetic regulatory network orchestrating the initiation of secondary cell wall development in cotton. *New Phytol.* 2020;226:1738–52.
27. Ma Y, Min L, Wang J, Li Y, Wu Y, Hu Q, Ding Y, Wang M, Liang Y, Gong Z, et al. A combination of genome-wide and transcriptome-wide association studies reveals genetic elements leading to male sterility during high temperature stress in cotton. *The New phytol.* 2021;231:165–81.
28. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, et al. Genetics of gene expression surveyed in maize, mouse and man. *Nature.* 2003;422:297–302.
29. Jaffe AE, Straub RE, Shin JH, Tao R, Gao Y, Collado-Torres L, Kam-Thong T, Xi HS, Quan J, Chen Q, et al. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat Neurosci.* 2018;21:1117–25.
30. Criado-Mesas L, Ballester M, Crespo-Piazuelo D, Castelló A, Fernández AI, Folch JM. Identification of eQTLs associated with lipid metabolism in *Longissimus dorsi* muscle of pigs with different genetic backgrounds. *Sci Rep.* 2020;10:9845.
31. Gu J, Hou D, Li Y, Chao H, Zhang K, Wang H, Xiang J, Raboanatahiry N, Wang B, Li M. Integration of proteomic and genomic approaches to dissect seed germination vigor in *Brassica napus* seeds differing in oil content. *BMC Plant Biol.* 2019;19:21.
32. Xiao Z, Zhang C, Tang F, Yang B, Zhang L, Liu J, Huo Q, Wang S, Li S, Wei L, et al. Identification of candidate genes controlling oil content by combination of genome-wide association and transcriptome analysis in the oilseed crop *Brassica napus*. *Biotechnol Biofuels.* 2019;12:216.
33. Elahi N, Duncan RW, Stasolla C. Modification of oil and glucosinolate content in canola seeds with altered expression of *Brassica napus* LEAFY COTYLEDON1. *Plant Physiol Biochem.* 2016;100:52–63.
34. Li-Beisson Y, Shorosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, Debono A, Durrett TP, et al. Acyl-lipid metabolism. *Arabidopsis Book.* 2013;11:e0161.
35. Cronan JE Jr, Waldrop GL. Multi-subunit acetyl-CoA carboxylases. *Prog Lipid Res.* 2002;41:407–35.
36. Sasaki Y, Nagano Y. Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Biosci Biotechnol Biochem.* 2004;68:1175–84.
37. Zu X, Zhong J, Luo D, Tan J, Zhang Q, Wu Y, Liu J, Cao R, Wen G, Cao D. Chemical genetics of acetyl-CoA carboxylases. *Molecules.* 2013;18:1704–19.
38. Chapman KD, Ohlrogge JB. Compartmentation of triacylglycerol accumulation in plants. *J Biol Chem.* 2012;287:2288–94.
39. Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M. Genetics and biochemistry of seed flavonoids. *Annu Rev Plant Biol.* 2006;57:405–30.
40. Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* 2015;20:176–85.
41. Xu W, Grain D, Bobet S, Le Gourrierc J, Thévenin J, Kelemen Z, Lepiniec L, Dubos C. Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in *Arabidopsis* seed. *New Phytol.* 2014;202:132–44.
42. Pollastri S, Tattini M. Flavonols: old compounds for old roles. *Ann Bot.* 2011;108:1225–33.
43. Xuan L, Zhang C, Yan T, Wu D, Hussain N, Li Z, Chen M, Pan J, Jiang L. TRANSPARENT TESTA 4-mediated flavonoids negatively affect embryonic fatty acid biosynthesis in *Arabidopsis*. *Plant Cell Environ.* 2018;41:2773–90.
44. Zhai Y, Yu K, Cai S, Hu L. Targeted mutagenesis of BnTT8 homologs controls yellow seed coat development for effective oil production in *Brassica napus* L. *Plant Biotechnol J.* 2020;18:1153–68.
45. Xie T, Chen X, Guo T, Rong H, Chen Z, Sun Q, Batley J, Jiang J, Wang Y. Targeted knockout of BnTT2 homologues for yellow-seeded *Brassica napus* with reduced flavonoids and improved fatty acid composition. *J Agric Food Chem.* 2020;68:5676–90.
46. Zhang Y, Zhang H, Zhao H, Xia Y, Zheng X, Fan R, Tan Z, Duan C, Fu Y, Li L, et al. Multi-omics analysis dissects the genetic architecture of seed coat content in *Brassica napus*. *Genome Biol.* 2022;23:86.

47. Marles MS, Gruber MY. Histochemical characterisation of unextractable seed coat pigments and quantification of extractable lignin in the *Brassicaceae*. *J Sci Food Agr*. 2004;84:251–62.
48. Meng J, Shi S, Gan L, Li Z, Qu X. The production of yellow-seeded *Brassica napus* (AACC) through crossing interspecific hybrids of *B. campestris* (AA) and *B. carinata* (BBCC) with *B. napus*. *Euphytica*. 1998;103:329–33.
49. Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF. SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. *Nature*. 2000;404:766–70.
50. Brambilla V, Battaglia R, Colombo M, Masiero S, Bencivenga S, Kater MM, Colombo L. Genetic and molecular interactions between BELL1 and MADS box factors support ovule development in *Arabidopsis*. *Plant Cell*. 2007;19:2544–56.
51. Wada T, Tachibana T, Shimura Y, Okada K. Epidermal cell differentiation in *Arabidopsis* determined by a Myb homolog CPC. *Science*. 1997;277:1113–6.
52. Zhu H-F, Fitzsimmons K, Khandelwal A, Kranz RG. CPC, a single-repeat R3 MYB, is a negative regulator of anthocyanin biosynthesis in *Arabidopsis*. *Mol Plant*. 2009;2:790–802.
53. Bharti AK, Khurana JP. Molecular characterization of transparent testa (tt) mutants of *Arabidopsis thaliana* (ecotype Estland) impaired in flavonoid biosynthetic pathway. *Plant Sci*. 2003;165:1321–32.
54. Chalhoub B, Denoëud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*. 2014;345:950–3.
55. Liu D, Yu L, Wei L, Yu P, Wang J, Zhao H, Zhang Y, Zhang S, Yang Z, Chen G, et al. BnTIR: an online transcriptome platform for exploring RNA-seq libraries for oil crop *Brassica napus*. *Plant Biotechnol J*. 2021;19:1895–7.
56. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res*. 2020;49:D344–54.
57. Takenaka Y, Kato K, Ogawa-Ohnishi M, Tsuruhama K, Kajiura H, Yagyu K, Takeda A, Takeda Y, Kunieda T, Hara-Nishimura I, et al. Pectin RG-I rhamnosyltransferases represent a novel plant-specific glycosyltransferase family. *Nat Plants*. 2018;4:669–76.
58. Lund CH, Stenbæk A, Atmodjo MA, Rasmussen RE, Møller IE, Erstad SM, Biswal AK, Mohnen D, Mravec J, Sakuragi Y. Pectin synthesis and pollen tube growth in *Arabidopsis* involves three GAUT1 Golgi-anchoring proteins: GAUT5, GAUT6, and GAUT7. *Front Plant Sci*. 2020;11:585774.
59. Miao Y, Li HY, Shen J, Wang J, Jiang L. QUASIMODO 3 (QUA3) is a putative homogalacturonan methyltransferase regulating cell wall biosynthesis in *Arabidopsis* suspension-cultured cells. *J Exp Bot*. 2011;62:5063–78.
60. Sung TY, Tseng CC, Hsieh MH. The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in *Arabidopsis* mitochondria. *Plant J*. 2010;63:499–511.
61. Lisec J, Meyer RC, Steinfath M, Redestig H, Becher M, Witucka-Wall H, Fiehn O, Törjék O, Selbig J, Altmann T, Willmitzer L. Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J*. 2008;53:960–72.
62. Yano K, Yamamoto E, Aya K, Takeuchi H, Lo PC, Hu L, Yamasaki M, Yoshida S, Kitano H, Hirano K, Matsuoka M. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat Genet*. 2016;48:927–34.
63. Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet*. 2012;44:32–9.
64. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*. 2010;42:961–7.
65. Gamazon ER, Wheeler HE. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet*. 2015;47:1091–8.
66. Gusev A, Lawrenson K, Lin X, Lyra PC Jr, Kar S, Vavra KC, Segato F, Fonseca MAS, Lee JM, Pejovic T, et al. A transcriptome-wide association study of high-grade serous epithelial ovarian cancer identifies new susceptibility genes and splice variants. *Nat Genet*. 2019;51:815–23.
67. Doke T, Huang S, Qiu C, Liu H, Guan Y, Hu H, Ma Z, Wu J, Miao Z, Sheng X, et al. Transcriptome-wide association analysis identifies DACH1 as a kidney disease risk gene that contributes to fibrosis. *J Clin Invest*. 2021;131:e141801.
68. Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function. *Am J Hum Genet*. 2018;102:717–30.
69. Wainberg M, Sinnott-Armstrong N. Opportunities and challenges for transcriptome-wide association studies. *Nat Genet*. 2019;51:592–9.
70. Tan Z, Peng Y, Xiong Y, Xiong F, Zhang Y, Guo N, Tu Z, Zong Z, Wu X, Ye J, et al. Comprehensive transcriptional variability analysis reveals gene networks regulating seed oil content of *Brassica napus*. *Genome Biol*. 2022;23:233.
71. Hong Y, Xia H, Li X, Fan R, Li Q, Ouyang Z, Tang S, Guo L. *Brassica napus* BnaNTT1 modulates ATP homeostasis in plastids to sustain metabolism and growth. *Cell Rep*. 2022;40:111060.
72. Bonaventure G, Salas JJ, Pollard MR, Ohlrogge JB. Disruption of the FATB gene in *Arabidopsis* demonstrates an essential role of saturated fatty acids in plant growth. *Plant Cell*. 2003;15:1020–33.
73. Penfield S, Li Y, Gilday AD, Graham S, Graham IA. *Arabidopsis* ABA INSENSITIVE4 regulates lipid mobilization in the embryo and reveals repression of seed germination by the endosperm. *Plant Cell*. 2006;18:1887–99.
74. Chen M, Wang Z, Zhu Y, Li Z, Hussain N, Xuan L, Guo W, Zhang G, Jiang L. The effect of transparent TESTA2 on seed fatty acid biosynthesis and tolerance to environmental stresses during young seedling establishment in *Arabidopsis*. *Plant Physiol*. 2012;160:1023–36.
75. Chen M, Xuan L, Wang Z, Zhou L, Li Z, Du X, Ali E, Zhang G, Jiang L. TRANSPARENT TESTA8 inhibits seed fatty acid accumulation by targeting several seed development regulators in *Arabidopsis*. *Plant Physiol*. 2014;165:905–16.
76. Moles AT, Ackerly DD, Webb CO, Tweddle JC, Dickie JB, Pitman AJ, Westoby M. Factors that shape seed mass evolution. *P Natl Acad Sci*. 2005;102:10540–4.
77. Song X-J, Huang W, Shi M, Zhu M-Z, Lin H-X. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet*. 2007;39:623–30.
78. Pal L, Sandhu SK, Bhatia D, Sethi S. Genome-wide association study for candidate genes controlling seed yield and its components in rapeseed (*Brassica napus*). *Physiol Mol Biol Plants*. 2021;27:1933–51.

79. So KKY, Duncan RW. Breeding canola (*Brassica napus* L.) for protein in feed and food. *Plants* (Basel). 2021;10:2220.
80. Pohanka M. New uses of melatonin as a drug, a review. *Curr Med Chem*. 2022;29:3622–37.
81. Liu J, Clough SJ, Hutchinson AJ, Adamah-Biassi EB, Popovska-Gorevski M, Dubocovich ML. MT1 and MT2 melatonin receptors: a therapeutic perspective. *Annu Rev Pharmacol Toxicol*. 2016;56:361–83.
82. Huangfu L, Chen R, Lu Y, Zhang E, Miao J, Zuo Z, Zhao Y, Zhu M, Zhang Z, Li P. OsCOMT, encoding a caffeic acid O-methyltransferase in melatonin biosynthesis, increases rice grain yield through dual regulation of leaf senescence and vascular development. *Plant Biotechnol J*. 2022;20:1122–39.
83. Naz S, Gallart-Ayala H, Reinke SN, Mathon C, Blankley R, Chaleckis R, Wheelock CE. Development of a liquid chromatography–high resolution mass spectrometry metabolomics method with high specificity for metabolite identification using all ion fragmentation acquisition. *Anal Chem*. 2017;89:7933–42.
84. Chaleckis R, Naz S, Meister I, Wheelock CE. LC-MS-based metabolomics of biofluids using all-ion fragmentation (AIF) acquisition. *Methods Mol Biol*. 2018;1730:45–58.
85. Blaženović I, Kind T, Ji J, Fiehn O. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites*. 2018;8:31.
86. Kind T, Tsugawa H, Cajka T, Ma Y, Lai Z, Mehta SS, Wohlgemuth G, Barupal DK, Showalter MR, Arita M, Fiehn O. Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom Rev*. 2018;37:513–32.
87. Chaleckis R, Meister I, Zhang P, Wheelock CE. Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics. *Curr Opin Biotechnol*. 2019;55:44–50.
88. Gan L, Sun X, Jin L, Wang G, Xu J, Wei Z, Fu T. Establishment of math models of NIRS analysis for oil and protein contents in seed of *Brassica napus*. *Zhongguo nong ye ke xue*. 2003;36:1609–13.
89. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
90. Widmer C, Lippert C, Weissbrod O, Fusi N, Kadie C, Davidson R, Listgarten J, Heckerman D. Further improvements to linear mixed models for genome-wide association studies. *Sci Rep*. 2014;4:6874.
91. Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, Hattori C, Werner P, Bancroft I. Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nat Biotechnol*. 2012;30:798–802.
92. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
93. Nam H, Chung BC, Kim Y, Lee K, Lee D. Combining tissue transcriptomics and urine metabolomics for breast cancer biomarker identification. *Bioinformatics*. 2009;25:3151–7.
94. Wu BL, Abbott T, Fishman D, McMurray W, Mor G, Stone K, Ward D, Williams K, Zhao HY. Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics*. 2003;19:1636–43.
95. Chen TL, Cao Y, Zhang YN, Liu JJ, Bao YQ, Wang CR, Jia WP, Zhao AH. Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection. *Evid Based Complement Alternat Med*. 2013;2013:298183.
96. Gao X, Yan P, Shen W, Li X, Zhou P, Li Y. Modular construction of plasmids by parallel assembly of linear vector components. *Anal Biochem*. 2013;437:172–7.
97. Dai C, Li Y, Li L, Du Z, Lin S, Tian X, Li S, Yang B, Yao W, Wang J, et al. An efficient *Agrobacterium*-mediated transformation method using hypocotyl as explants for *Brassica napus*. *Mol Breeding*. 2020;40:96.
98. Yang Q, Fan C, Guo Z, Qin J, Wu J, Li Q, Fu T, Zhou Y. Identification of FAD2 and FAD3 genes in *Brassica napus* genome and development of allele-specific markers for high oleic and low linolenic acid contents. *Theor Appl Genet*. 2012;125:715–29.
99. Li D, Guo Y. Melatonin represses oil and anthocyanin accumulation in seeds. *Plant Physiol*. 2020;183:898–914.
100. Tang S, Zhao H, Lu S, Yu L, Zhang G, Zhang Y, Yang Q, Zhou Y, Wang X, Ma W, et al. Genome-wide re-sequencing data of *Brassica napus*. Datasets. Genome Sequence Archive. <https://ngdc.cnbc.ac.cn/bioproject/browse/PRJCA002835>. 2021.
101. Tang S, Zhao H, Lu S, Yu L, Zhang G, Zhang Y, Yang Q, Zhou Y, Wang X, Ma W, et al. Transcriptome-wide data of seed of *Brassica napus*. Datasets. Genome Sequence Archive. <https://ngdc.cnbc.ac.cn/bioproject/browse/PRJCA002836>. 2021.
102. Li L, Tian Z, Chen J, Tan Z, Zhang Y, Zhao H, Wu X, Yao X, Wen W, Chen W, Guo L. Metabolome-wide data of seed in *Brassica napus*. Datasets. China National Center for Bioinformation. <https://ngdc.cnbc.ac.cn/omix/release/OMIX004266>. 2023.

## Publisher's Note

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