

SCIENTIFIC REPORTS



OPEN

Joint genome-wide association and transcriptome sequencing reveals a complex polygenic network underlying hypocotyl elongation in rapeseed (*Brassica napus* L.)

Received: 27 September 2016

Accepted: 20 December 2016

Published: 31 January 2017

Xiang Luo, Zhifei Xue, Chaozhi Ma, Kaining Hu, Ziru Zeng, Shengwei Dou, Jinxing Tu, Jinxiong Shen, Bin Yi & Tingdong Fu

Hypocotyl elongation is considered an important typical seedling trait contributing directly to an increase in and stabilization of the yield in *Brassica napus*, but its molecular genetic mechanism is poorly understood. In the present study, hypocotyl lengths of 210 lines were measured in an illuminated culture room. A genome-wide association study (GWAS) was performed with 23,435 single nucleotide polymorphisms (SNPs) for hypocotyl length. Three lines with long hypocotyl length and three lines with short hypocotyl length from one doubled haploid line (DH) population were used for transcriptome sequencing. A GWAS followed by transcriptome analysis identified 29 differentially expressed genes associated with significant SNPs in *B. napus*. These genes regulate hypocotyl elongation by mediating flowering morphogenesis, circadian clock, hormone biosynthesis, or important metabolic signaling pathways. Among these genes, *BnaC07g46770D* negatively regulates hypocotyl elongation directly, as well as flowering time. Our results indicate that a joint GWAS and transcriptome analysis has significant potential for identifying the genes responsible for hypocotyl elongation; The extension of hypocotyl is a complex biological process regulated by a polygenic network.

Rapeseed (*Brassica napus* L., $2n = 4x = 38$; genome AACC) is one of the most important oilseed crops in the world and the genetic control of yield and yield-related traits has been studied extensively. However, less attention has been focused on elucidating the genetic mechanism of seedling traits. Well-developed seedling traits can contribute directly to an increase in and stabilization of yield and its heterosis, mineral nutrient absorption, drought tolerance, and salinity tolerance in crops^{1–10}. Thus, understanding the seedling traits will be conducive to breeding programs in *B. napus*.

In rapeseed, early seedling development traits have been shown to significantly correlate with agronomic traits². Many heterosis-related quantitative trait loci (QTLs) for seedling traits in *B. napus* are common with yield-related QTLs¹. Several promising positional and functional candidate genes have been associated with germination speed, absolute germination rate, and radicle growth in *B. napus*⁶. The number of loci detected for 14 seedling development traits, and candidate genes *GER1*, *AILP1*, *PECT*, and *FBP* were reported to strongly relate to the seedling development traits in *B. napus*⁷. Hypocotyl elongation is thought to be an importantly typical seedling trait in plants. However, almost all studies on hypocotyl elongation have focused on *Arabidopsis*. Global transcriptome analysis has revealed circadian regulation of key pathways in rhythmic hypocotyl elongation¹¹. The transcriptional regulators *CIRCADIAN CLOCK ASSOCIATED1* (*CCA1*), *EARLY FLOWERING3* (*ELF3*), *ELF4*, and *LUX ARRHYTHMO* (*LUX*) appear to link the circadian clock to diurnal control of hypocotyl growth^{12,13}. Therefore, hypocotyl elongation has a fiercely complex genetic mechanism, but little knowledge is available about hypocotyl elongation in *B. napus*.

National Key Laboratory of Crop Genetic Improvement, National Center of Rapeseed Improvement in Wuhan, Huazhong Agricultural University, Wuhan 430070, P.R. China. Correspondence and requests for materials should be addressed to C.M. (email: yuanbeauty@mail.hzau.edu.cn)

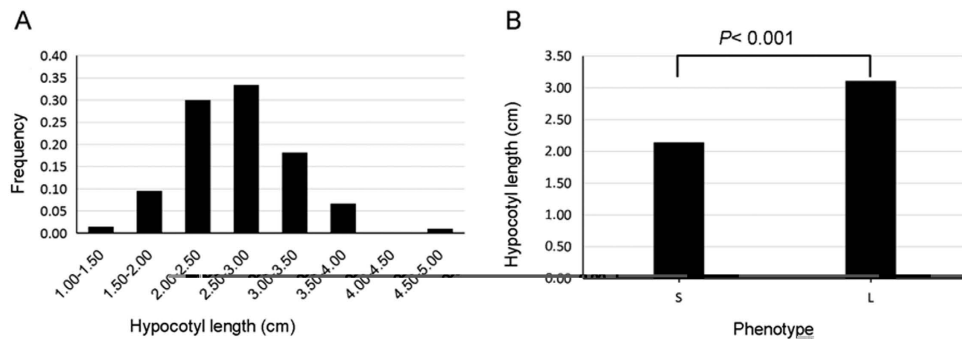


Fig. 1. Phenotypic variation in 210 accessions. (A) Frequency of phenotypic variation in 210 accessions. (B) Comparison of two phenotypes by *t*-test.

Expressed sequence tag (EST) sequence data, array analysis, amplicon resequencing, sequences, and next-generation sequencing technologies efficiently allow genome-wide association studies (GWASs) and transcriptome analysis to be novel strategies for dissecting complex traits in plants^{14–21}. In the present study, a GWAS for hypocotyl elongation was carried out with a panel of 210 *B. napus* accessions genotyped for 23,435 SNPs. We also performed transcriptome sequencing of long and short hypocotyl phenotypes. The main objectives of this study were to obtain a better understanding of hypocotyl elongation and its relationship with yield-related traits or heterosis, dissect the genetic basis of hypocotyl elongation by combining GWAS and transcriptome analysis, and perform Gene Ontology (GO) and KEGG pathway analysis for associated genes in *B. napus*.

Results

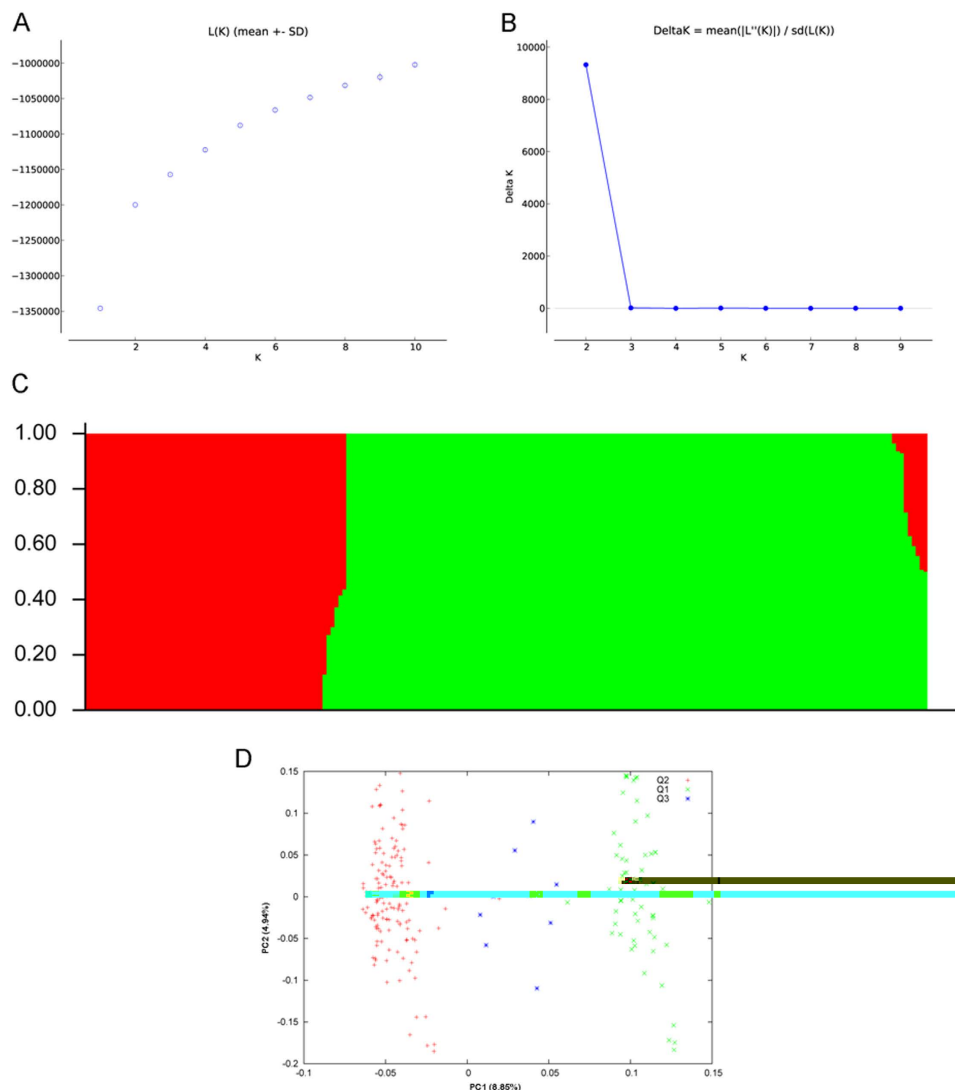
Phenotypic variations and correlation analysis. Extensive phenotypic variations in hypocotyl elongation were observed in the 210 rapeseed lines (Fig. 1A and Supplementary Tables S1 and S2). The hypocotyl elongation of the lines was normally distributed (average = 2.66, range = 1.38 to 4.81), and 63.33% of hypocotyl elongation values were between 2.00 and 3.00.

The correlation coefficients between hypocotyl elongation and yield-related traits showed that hypocotyl elongation positive correlated with seed yield per plant (0.29) and biomass yield per plant (0.21) at $P = 0.01$ and plant height (0.19) at $P = 0.05$ (Supplementary Table S3). Linear regression analysis of the correlated traits indicated that hypocotyls elongation can explain 3.28% of the total seed yield per plant ($P < 0.05$), 4.49% of the total biomass yield per plant ($P < 0.01$), 3.59% of the total plant height ($P < 0.05$), respectively.

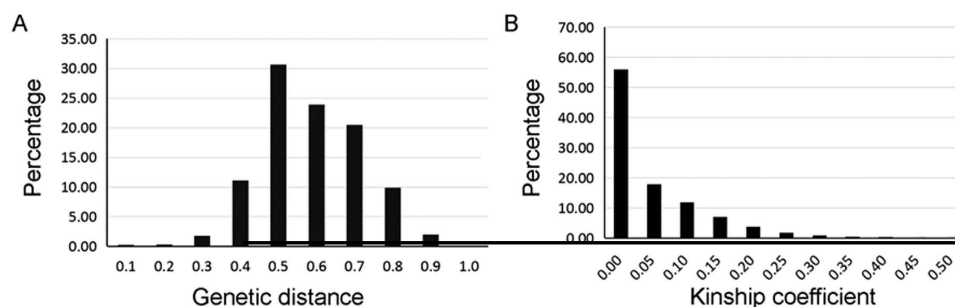
Genetic diversity, population structure, and relative kinship analysis. The genetic diversity and population structure of the 210 accessions were analyzed using 5,334 SNPs (Supplementary Table S4). Clustering inference showed that the most significant change in likelihood occurred when K increased from 2 to 3, and the highest Δk value was observed at $K = 2$ (Fig. 2A–C). Considering the probability of membership threshold of 0.70, 61 and 140 accessions were assigned to subgroups Q1 and Q2, respectively (Supplementary Table S1). The remaining nine accessions were assigned to a mixed group (Q3). The PCA also provided a pattern for the genetic structure of the GWAS population (Fig. 2D). The top two principal components clearly separated these subpopulations and explained 8.85% and 4.94% of the total SNP variations in the rapeseed panel, respectively. All of the parameters suggest that the three-group model (subgroups Q1, Q2, and Q3) sufficiently explained the genetic structure among the 210 accessions. The mean genetic distance (GD) between lines was 0.54, and 74.85% of pairs had a GD ranging from 0.5 to 0.7 (Fig. 3A). The average kinship coefficient identity by descent (IBD) within the total diversity set was 0.06 (Fig. 3B). A total of 55.93% of the pairwise kinship estimates were equal to 0, and 17.85% of pairwise kinship coefficients varied from 0 (excluding 0) to 0.05.

LD analysis. All 23,435 SNPs in the total panel were used for LD analysis. The distributions of r^2 with respect to the physical distance from each chromosome are shown in the Supplementary data (Fig. S1 and Supplementary Table S5). As expected, the mean r^2 between 0 and 500 kb decreased rapidly and continuously, followed by much slower decay at increased physical distance for both the A genome and C genome. The overall LD decay distance was 893.84 Kb when the r^2 cutoff was set to 0.1. The rate of LD decay varied over different chromosomes in both the A genome and C genome, with the shortest LD decays of 459.03 kb on chromosome A07 and 602.91 kb on chromosome C08 and the longest LD decays of 968.17 kb on chromosome A09 and 3,190.79 kb on chromosome C09. Obviously, the LD of the A genome decayed significantly faster than the LD of the C genome.

Association mapping and candidate gene prediction. Total 23,435 polymorphisms with minor allele frequency (MAF) ≥ 0.05 were selected for association mapping of hypocotyl elongation using the BLUP value across multiple replications (Supplementary Table S1). Model comparison analyses indicated that P -values from the PCA + K model were nearer the expected P -values than those of the GLM, Q, PCA, and Q + K models (Fig. 4A). Thus, the PCA + K model was selected for association mapping of hypocotyl elongation. Five SNPs on C07 were highly significantly associated with hypocotyl elongation at $P < 2.13 \times 10^{-6}$, with a FDR of 1.0% (Fig. 4B and Table 1). All detected SNPs were located between 42.15 and 42.25 Mb on C07 and could explain 4.82% of the total phenotypic variance. Thus, the development of hypocotyl is controlled by a minor-effect polygene. In LD



F e2. A a e a c e 210 a e e d a c c e *STRUCTURE*. (A) Estimated $L(K)$ of possible clusters (k) from 1 to 10. (B) Delta K based on the rate of change of $L(K)$ between successive K values. (C) Population structure based on $k=2$. Red represents subgroup Q1; green represents subgroup Q2. (D) Principal component (PC) analysis.



F e3. G e e c d a c e a d c e c e a a b e e e a a c c e

analyses, the r^2 values were > 0.79 for all pairs of associated SNPs, suggesting that the associated SNPs were in high LD with each other (Fig. 5A and B).

According to the associated SNP variations, four haplotypes (H0, H1, H2, and H3) were identified from these *B. napus* accessions (Fig. 5C). H0, H1, H2, and H3 were observed in 3, 81, 100, and 8 lines, respectively.

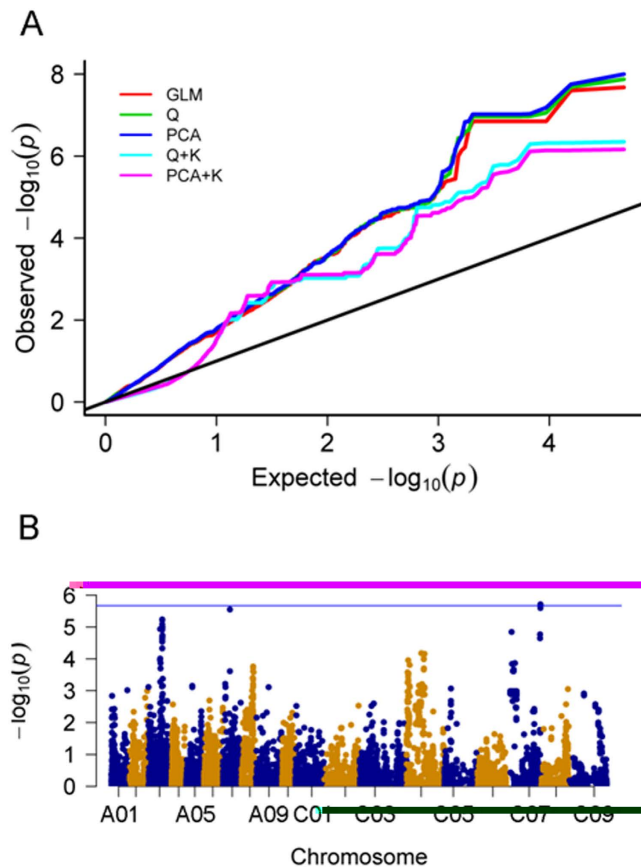


Fig. 4. Association analysis of hypocotyl elongation. (A) Quantile-quantile plots of estimated $-\log_{10}(P)$ from the association analysis of hypocotyl elongation. The black line represents expected P -values with no associations. The red line represents observed P -values using the GLM model. The green line represents observed P -values using the Q model. The blue line represents observed P -values using the PCA model. The cyan line represents observed P -values using the Q + K model. The pink line represents observed P -values using the PCA + K model (color figure online). (B) Manhattan and quantile-quantile plots generated from the genome-wide association analysis of hypocotyl elongation. The blue horizontal line depicts the Bonferroni significance threshold (2.13×10^{-6}).

SNP	C	e	A e e	P	MAF	P- a e	R ² (%)
Bn-scaff 16110 1-p587456	C07		A/G	44303399	0.49	1.94E-06	4.82
Bn-scaff 16110 1-p670992	C07		G/A	44218661	0.50	7.25E-07	
Bn-scaff 16110 1-p671315	C07		A/G	44218337	0.49	7.66E-07	
Bn-scaff 16110 1-p685258	C07		A/G	44204592	0.50	6.87E-07	
Bn-scaff 16110 1-p685428	C07		C/A	44204201	0.50	7.24E-07	

Table 1. Significant SNPs associated with hypocotyl elongation. MAF, minor allele frequency; R²(%), amount of phenotypic variation for each trait explained by multiple SNPs.

Approximately 94.27% of accessions comprised H1 and H2. Therefore, H0 and H3 are rare variations, whereas H1 and H2 are conserved by artificial selection during the genetic improvement of modern *B. napus* breeding accessions. Further analysis showed that H0, H1, H2, and H3 have mean hypocotyl elongations of 2.99, 2.52, 2.79, and 2.52, respectively (Fig. 5D). H2 had a significantly ($P = 0.001$) greater hypocotyl elongation than H1. H0 and H3 were not analyzed because they are extremely rare. Thus, H2 may be a favorable haplotype and facilitate the selection of better genotypes for hypocotyl elongation in breeding *B. napus*.

Candidate genes were predicted along the ~100 Kb region between two associated SNPs (Bn-scaff 16110 1-p685428 and Bn-scaff 16110 1-p587456) according to the newly released *B. napus* genome sequence²². Only five genes (*BnaC07g46740D*, *BnaC07g46760D*, *BnaC07g46770D*, *BnaC07g46780D*, and *BnaC07g46800D*) were detected in the candidate region (Supplementary Table S6). Of these genes, *BnaC07g46770D* was previously identified to regulate the flowering time in rapeseed²³. The closest distance between *BnaC07g46770D* and a significant SNP (Bn-scaff 16110 1-p670992) was 34 Kb. Considering the LD decay of 754.95 Kb in C07, candidate genes

were also predicted in the region between 754.95 Kb upstream and downstream of the associated peak; 196 genes were obtained in the enlarged candidate region (Supplementary Table S6). All of the genes were blasted against *A. thaliana* genome data, but none of the predicted genes were homologous to the genes directly controlling hypocotyl elongation in *Arabidopsis*.

BnaC07g46660D, *BnaC07g46770D*, and *BnaC07g47470D*) were associated with the response to hormone and flower morphogenesis (Supplementary Table S8).

To explore the function of DEGs in the biosynthesis and metabolite pathways, KEGG pathway analysis was performed in two phenotypic groups (Supplementary Table S9). Six DEGs (*BnaC07g45590D*, *BnaC07g45710D*, *BnaC07g46060D*, *BnaC07g46560D*, *BnaC07g46660D*, and *BnaC07g47470D*) acted in the 30 pathways by encoding corresponding enzymes. For example, *BnaC07g46060D* and *BnaC07g46560D* regulate the lignins and phenylpropanoid biosynthesis in phenylpropanoid metabolic pathways by encoding dehydrogenase and lactoperoxidase, respectively. Furthermore, *BnaC07g46060D* and *BnaC07g47470D* participate in glycolysis/gluconeogenesis and nitrogen metabolism and carbon fixation in photosynthetic organisms by encoding dehydrogenase and aldolase, respectively. A global examination of gene expression demonstrated that genes encoding dehydrogenase regu-

was previously reported to regulate flowering time and is orthologous to *A. thaliana* *AP2* and *AT4G37750*. *AP2* belongs to the AP2/ERF gene family and is involved in plant development, in turning leaves into floral organs²⁴. *AT4G37750* belongs to the AP2/EREBP gene family and directly regulates a key clock gene (*CCA1*) that provides

In summary, this study is the first to study the hypocotyl elongation by integrating GWAS and transcriptome sequencing in *B.napus*. We demonstrated that the genes mediated by circadian clock, hormone biosynthesis, floral morphogenesis, or other metabolic signaling pathways may regulate the hypocotyl elongation in *B.napus*. These findings reveal that the phenotypic variation of the hypocotyl is a complex biological process regulated by a polygenic network in *B.napus*. Over the past decade, circadian clock and hormone effects had been linked to agronomic traits in plant^{48,49}. Hypocotyl elongation represents the best-studied model of plant circadian clock and hormone response system. Therefore, modification of these areas may have the potential for systemic effects that produce beneficial yield trait in *B.napus*.

Materials and Methods

Plant materials and trait collection. A set of 210 elite inbred rapeseed lines with abundant phenotypic variation were collected to construct an association panel (Supplementary Table S1); 55 lines (X001-X055) were used to isolate and characterize the sucrose transporter (SUT) gene⁵⁰, and 155 lines (X056-X210) were derived from an association mapping population genotyped using the 60 K Illumina[®] Infinium SNP array⁵¹. The yield-related traits of these lines were measured in a previous study^{50,51}. The 210 lines were grown with 20 replications in 10 × 10 culture plates. When cotyledons were fully developed, all of the lines were sprayed with nutrient solution as described previously⁵². To control environmental conditions, the seedlings were grown in an illuminated culture room under 16 L:8 D conditions at 20 °C and measurements performed on day 20. Photographs of seedlings were analyzed using AutoCAD software (<http://www.autodesk.com.cn/education/free-software/featured>). Three long hypocotyl (L) and three short hypocotyl (S) lines were used for transcriptome sequencing. These lines were selected from a doubled haploid (DH) population (DH-6004) developed from 2011 5515 137 × Gui01A10 F1 (field code 9 6004), in which ‘2011 5515 137’ exhibits early flower and ‘Gui01A10’ moderate flower.

SNP genotyping. Fifty-five lines (X001-X055) and six DHs (DH1, DH2, DH3, DH4, DH5 and DH6) were genotyped using the *Brassica* 60 K Illumina[®] Infinium SNP array. Combined with genotype information obtained previously for the other 155 lines, 26,016 SNPs were mapped in silico using the probe sequences of 52,157 SNPs to perform a Blast N search against *B.napus* genome sequences⁵³. Only the top hits, using an E-value cut-off of 1E-15 against the *B.napus* genome sequences, were considered. Hits with AA or BB frequency equal to zero (i.e., monomorphic), call frequency <0.8, or minor frequency <0.05 were excluded. Thus, a total of 23,435 SNPs were filtered for association analysis (Supplementary Table S4). Genetic diversity and Nei’s genetic distance⁵⁴ were estimated using *PowerMarker* version 3.25⁵⁵.

Population structure, relative kinship, and linkage disequilibrium. The population structure was inferred using the software package *STRUCTURE* v2.3.4⁵⁶ based on 5,334 SNPs with AA or BB frequency >0.05, call frequency ≥0.9 and minor frequency >0.2. Five independent runs were performed with a K-value (the putative number of genetic groups) from 1 to 10, with both the length of the burning period and the number of Markov Chain Monte Carlo (MCMC) replications after burning set to 100,000 iterations under the ‘admixture model’. The most likely k-value was determined by the log probability of data |LnP(D) and ad hoc statistic Δk based on the rate of change of LnP(D) between successive k values as described previously⁵⁷. Accessions with a probability of membership >0.7 were assigned to corresponding clusters, and those with a probability of membership <0.7 were assigned to a mixed group. The relative kinship matrix comparing all pairs of accessions was calculated using the software package *SPAGeD*⁵⁸. Negative values between two individuals were set to 0⁵⁹. Principal component analysis (PCA) based on SNPs was carried out using the *EIGENSTRAT* tool⁶⁰. The linkage disequilibrium (LD) parameter r^2 was calculated using the software *TASSEL* 3.0 with 1,000 permutations⁶¹.

GWAS and statistical analysis. The effects of population structure (Q, PC) and kinship (K) on the traits were evaluated by a GWAS using five models (GLM, Q, PCA, PCA + K, and Q + K). Significant loci were identified by comparing *P*-values with the Bonferroni threshold (0.05/23,435 = 2.13E-06). Quantile-quantile plots of the estimated $-\log_{10}(P)$ values in the association mapping model were created using the observed *P*-values from marker-trait associations versus the expected *P*-values. In addition, false discovery rates (FDRs) were calculated as $(m \times P)/n \times 100\%$, where *m* is the total number of SNPs (23,435 in this study), *P* is the *P*-value threshold for detecting a significant association, and *n* is the total number of significant associations per trait⁶².

Phenotypic variation, correlation and linear regression analyses were performed using *SPSS* version 19.0 (IBM Corp., Armonk, NY, USA).

Nuclear RNA extraction and RNA sequencing. When the second cotyledons were fully expanded in the illuminated culture room, the seedlings of three S lines (DH1, DH2 and DH3) and three L lines (DH4, DH5 and DH6) were pooled to long hypocotyl bulk and short hypocotyl bulk, respectively, then immediately frozen in liquid nitrogen and stored at -80 °C. Total nuclear RNA was extracted from ~100 mg of frozen plants using the RNeasy Pure Plant Kit (TIANGEB BIOTECH, Beijing, China) according to the manufacturer’s instructions in two biological replicates. NanoDrop ND 1000 (NanoDrop technologies) was used to evaluate the quality of the extracted RNA. RNA with an RNA Integrity Number (RIN) > 8 as assessed by Agilent Technologies 2100 Bioanalyzer (Agilent) was used to prepare the c-DNA library. The sequencing library was generated using the Illumina RNA Library Prep Kit (NASDAQ: ILMN, America) following the manufacturer’s recommendations. The library preparations were sequenced on an Illumina HiSeq 200 platform, and 100-bp paired-end reads were generated.

DEG identification and gene annotations. The sequenced data were trimmed by removing adapters, low-quality sequences or bases, and contaminations or overrepresented sequences using *Trimmomatic* software version 0.33. The clean reads were aligned to the *B.napus* reference genome²² using *Hisat* software version 0.1.6

and then assembled using TopHat 2.0.0 and Cufflinks⁶³. Fragments per kilobase million (FPKM) was determined to estimate gene expression levels. Differentially expressed genes (DEGs) between two genotypes were identified by Cuffdiff based on the criteria $P < 0.05$ and $|\log^2(L/S)| > 1$. To identify possible homologous genes, DEGs were blasted against the *A. thaliana* genome database (<http://www.arabidopsis.org/>). The GO enrichment analysis for DEGs was implemented by *Blast2GO* and significantly enriched GO terms ($P < 0.05$) displayed using the online tool WEGO (<http://wego.genomics.org.cn>). The enrichment of DEGs was determined by KEGG pathway analysis using the KOBAS2.0 website (<http://kobas.cbi.pku.edu.cn/home.do>). To analyze the metabolic pathway and functional classification of DEGs, expression data were mapped to metabolic pathways using *MapMan* software⁶⁴.

References

1. Basunanda, P. *et al.* Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **120**, 271–281, doi: 10.1007/s00122-009-1133-z (2010).
2. Korber, N. *et al.* Seedling development in a *Brassica napus* diversity set and its relationship to agronomic performance. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **125**, 1275–1287, doi: 10.1007/s00122-012-1912-9 (2012).
3. Liu, S. *et al.* Genome-wide analysis of ZmDREB genes and their association with natural variation in drought tolerance at seedling stage of *Zea mays* L. *PLoS Genet* **9**, e1003790, doi: 10.1371/journal.pgen.1003790 (2013).
4. Adolf, V. I., Jacobsen, S.-E. & Shabala, S. Salt tolerance mechanisms in quinoa (*Chenopodium quinoa* Willd.). *Environmental and Experimental Botany* **92**, 43–54 (2013).
5. Ahmad, M., Shahzad, A., Iqbal, M., Asif, M. & Hirani, A. H. Morphological and molecular genetic variation in wheat for salinity

35. Kumar, V. *et al.* Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Research* **22**, 133–145, doi: 10.1093/dnares/dsu046 (2015).
36. Xu, L. *et al.* Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.). *DNA Research*, dsv035, doi: 10.1093/dnares/dsv035 (2015).
37. Famoso, A. N. *et al.* Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genet* **7**, e1002221, doi: 10.1371/journal.pgen.1002221 (2011).
38. Lu, Y. *et al.* Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 19585–19590, doi: 10.1073/pnas.1006105107 (2010).
39. Cadic, E. *et al.* Combined linkage and association mapping of flowering time in Sunflower (*Helianthus annuus* L.). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **126**, 1337–1356, doi: 10.1007/s00122-013-2056-2 (2013).
40. Mir, R. R. *et al.* Genetic dissection of grain weight in bread wheat through quantitative trait locus interval and association mapping. *Molecular Breeding* **29**, 963–972, doi: 10.1007/s11032-011-9693-4 (2012).
41. Tian, F. *et al.* Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature genetics* **43**, 159–162, doi: 10.1038/ng.746 (2011).
42. Kover, P. X. *et al.* A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS genetics* **5**, e1000551, doi: 10.1371/journal.pgen.1000551 (2009).
43. Wei, L. *et al.* Genome-wide association analysis and differential expression analysis of resistance to Sclerotinia stem rot in *Brassica napus*. *Plant Biotechnol J* **14**, 1368–1380, doi: 10.1111/pbi.12501 (2016).
44. Kwon, Y. *et al.* A novel *Arabidopsis* MYB-like transcription factor, MYBH, regulates hypocotyl elongation by enhancing auxin accumulation. *Journal of experimental botany* **64**, 3911–3922, doi: 10.1093/jxb/ert223 (2013).
45. Zhao, Y. *et al.* Trp-dependent auxin biosynthesis in *Arabidopsis*: involvement of cytochrome P450s CYP79B2 and CYP79B3. *Genes & development* **16**, 3100–3112 (2002).
46. Mas, P. Dual Role of TOC1 in the Control of Circadian and Photomorphogenic Responses in *Arabidopsis*. *The Plant Cell Online* **15**, 223–236, doi: 10.1105/tpc.006734 (2002).
47. Vandenbussche, F., Verbelen, J.-P. & Van Der Straeten, D. Of light and length: Regulation of hypocotyl growth in *Arabidopsis*. *BioEssays* **27**, 275–284, doi: 10.1002/bies.20199 (2005).
48. Murai, N. Review: Plant Growth Hormone Cytokinins Control the Crop Seed Yield. *American Journal of Plant Sciences* **05**, 2178–2187, doi: 10.4236/ajps.2014.514231 (2014).
49. Bendix, C., Marshall, C. M. & Harmon, F. G. Circadian Clock Genes Universally Control Key Agricultural Traits. *Molecular plant* **8**, 1135–1152, doi: 10.1016/j.molp.2015.03.003 (2015).
50. Li, F. *et al.* Characterization of sucrose transporter alleles and their association with seed yield-related traits in *Brassica napus* L. *BMC plant biology* **11**, 168 (2011).
51. Luo, X. *et al.* Unravelling the complex trait of harvest index in rapeseed (*Brassica napus* L.) with association mapping. *BMC Genomics* **16**, 379, doi: 10.1186/s12864-015-1607-0 (2015).
52. Shi, T. *et al.* *Brassica napus* root mutants insensitive to exogenous cytokinin show phosphorus efficiency. *Plant and Soil* **358**, 61–74, doi: 10.1007/s11104-012-1219-2 (2012).
53. Chalhouh, B. *et al.* Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**, 950–953 (2014).
54. Nei, M. & Takezaki, N. Estimation of genetic distances and phylogenetic trees from DNA analysis. *Proc 5th World Cong Genet Appl Livestock Prod* **21**, 405–412 (1983).
55. Liu, K. & Muse, S. V. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**, 2128–2129 (2005).
56. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
57. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology* **14**, 2611–2620 (2005).
58. Hardy, O. J. & Vekemans, X. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular ecology notes* **2**, 618–620 (2002).
59. Yu, J. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics* **38**, 203–208 (2006).
60. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904–909, doi: 10.1038/ng1847 (2006).
61. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
62. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300 (1995).
63. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**, 562–578 (2012).
64. Thimm, O. *et al.* mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant Journal for Cell & Molecular Biology* **37**, 914–939 (2004).

Acknowledgements

This work was supported by grants from the National Key Research and Development Program of China (No.2016YFD0100803) and the National Natural Science Foundation of China (31571706).

Author Contributions

C.M. and X.L. conceived and designed the experiments. X.L., Z.X., Z.Z. and S.D performed the experiments. X.L., Z.X. and K.H. analyzed the data. C.M., J.T., J.S., B.Y. and T.F contributed materials. X.L. and C.M. wrote the paper. All authors reviewed the manuscript.

Additional Information

Supplementary Information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permissions: Luo, X. *et al.* Joint genome-wide association and transcriptome sequencing reveals a complex polygenic network underlying hypocotyl elongation in rapeseed (*Brassica napus* L.). *Sci. Rep.* **7**, 41561; doi: 10.1038/srep41561 (2017).

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

