Seed oil content QTL mapping and gene cloning in rapeseed

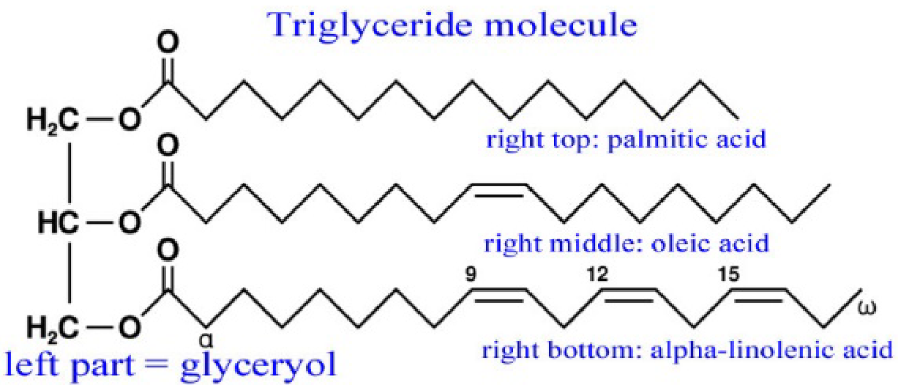
Huazhong Agricultural University
Yazhouwan National Laboratory

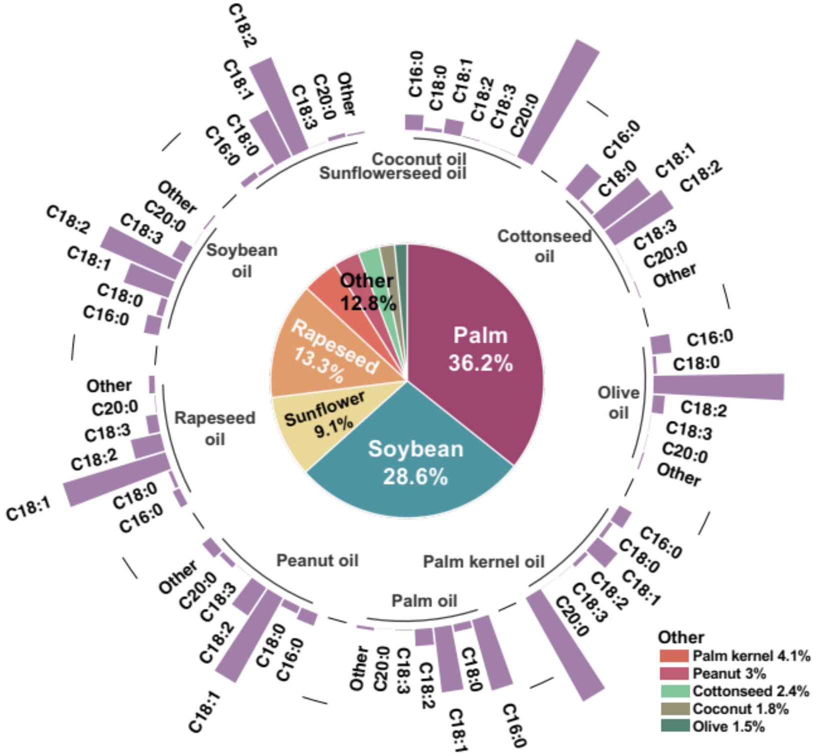
Liang Guo

September 25, 2023 Sydney

Rapeseed is the third largest source of vegetable oil

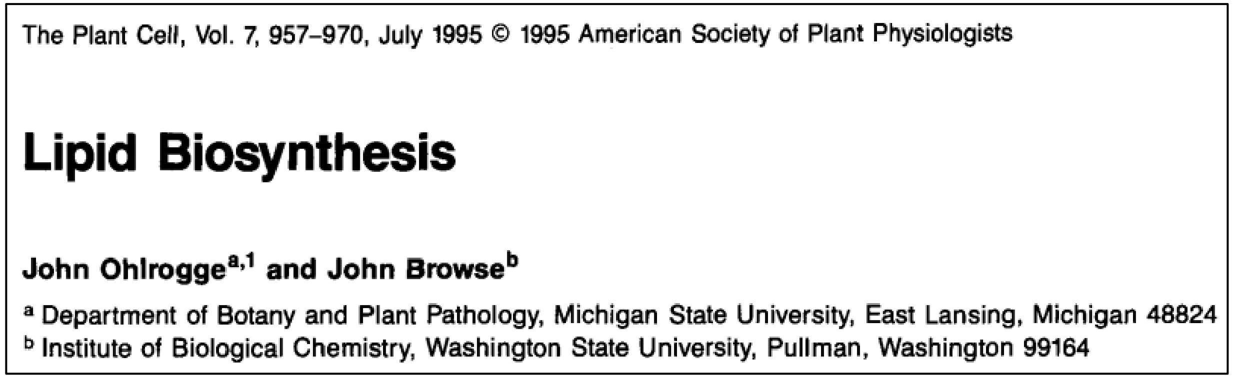
Vegetable oil sources: palm, soybean, rapeseed, peanut, maize, sunflower, cotton, sesame, olive, tea





Song et al., J Exp Bot (2021)

Plant oil biosynthesis

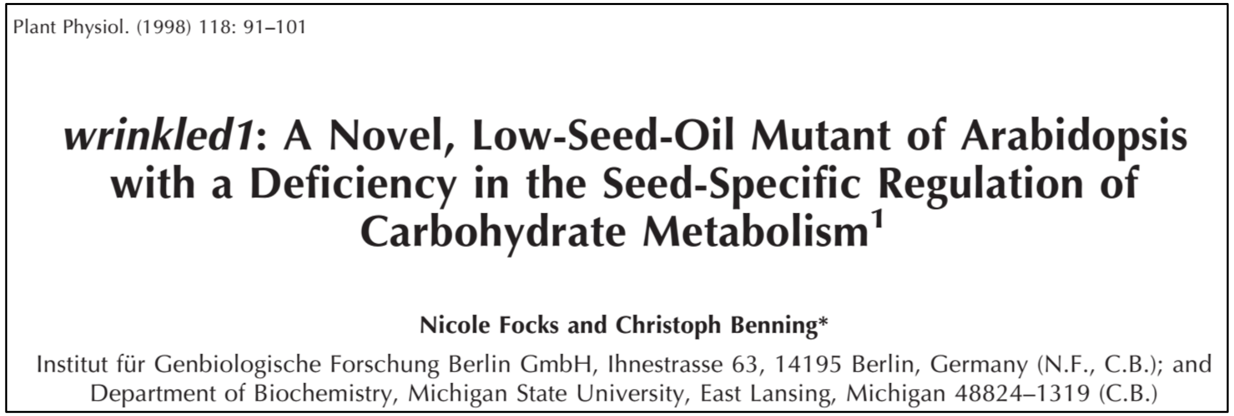


The Plant Cell, Vol. 7, 957–970, July 1995 © 1995 American Society of Plant Physiologists

Lipid Biosynthesis

John Ohlrogge[a,1] **and John Browse**[b]

[a] Department of Botany and Plant Pathology, Michigan State University, East Lansing, Michigan 48824
[b] Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164



Plant Physiol. (1998) 118: 91–101

wrinkled1: A Novel, Low-Seed-Oil Mutant of Arabidopsis with a Deficiency in the Seed-Specific Regulation of Carbohydrate Metabolism[1]

Nicole Focks and Christoph Benning*

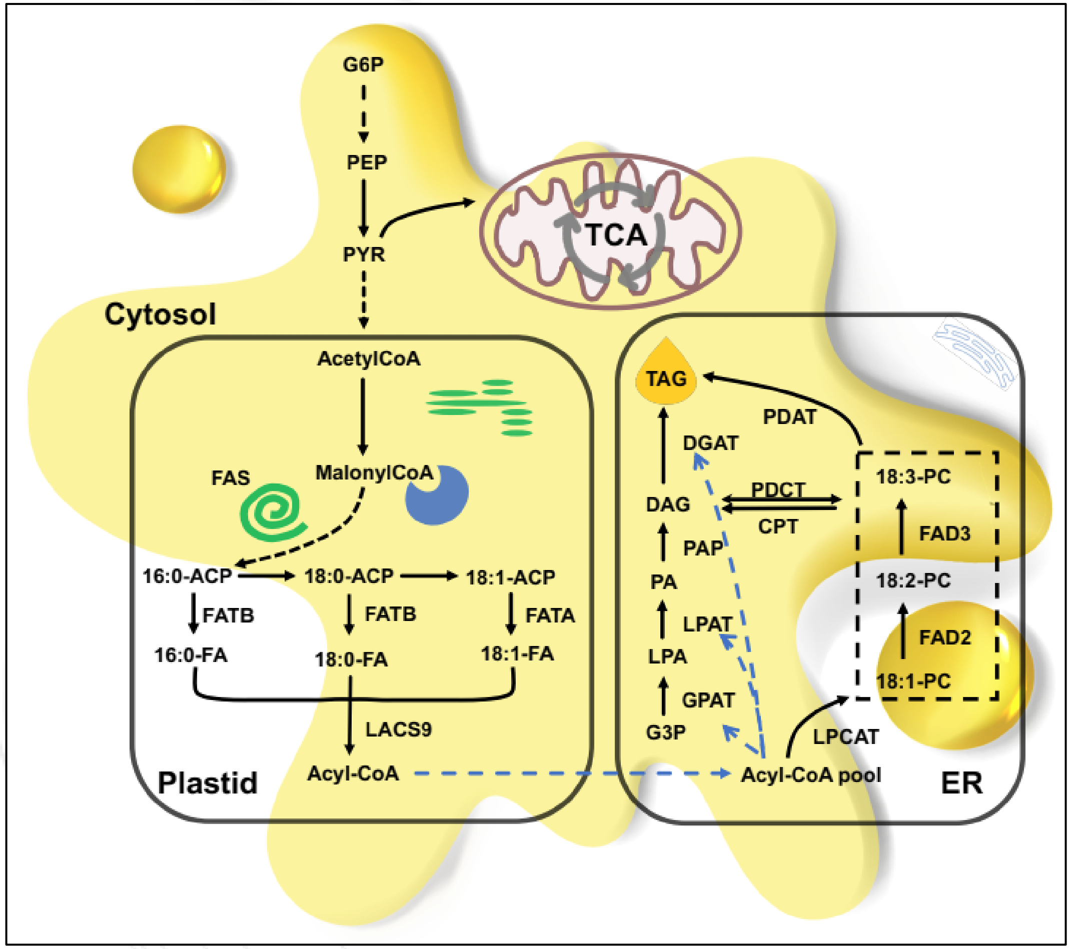
Institut für Genbiologische Forschung Berlin GmbH, Ihnestrasse 63, 14195 Berlin, Germany (N.F., C.B.); and Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824–1319 (C.B.)

An enzyme regulating triacylglycerol composition is encoded by the *ROD1* gene of *Arabidopsis*

Chaofu Lu[a,b], Zhanguo Xin[a,1], Zhonghai Ren[b], Martine Miquel[a,2], and John Browse[a,3]

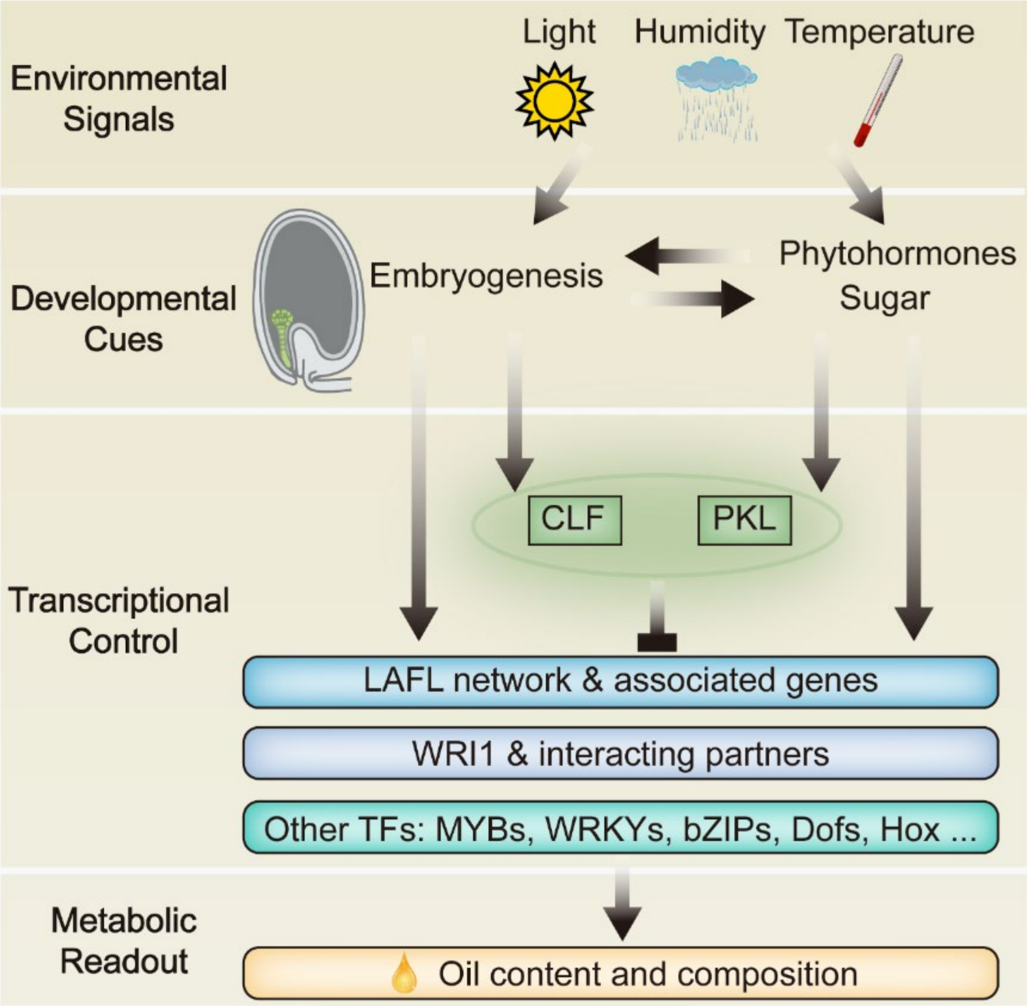
[a]Institute of Biological Chemistry, Washington State University, Pullman, WA 99164; and [b]Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, MT 59717

>700 lipid metabolism genes in Arabidopsis



Ohlrogge and Browse, Plant Cell (1995); Focks and Benning (1998); Lu et al. PNAS (2009)
Chapman and Ohlrogge, JBC (2012); Song et al., J Exp Bot (2021)

Many factors affect oil accumulation in seeds



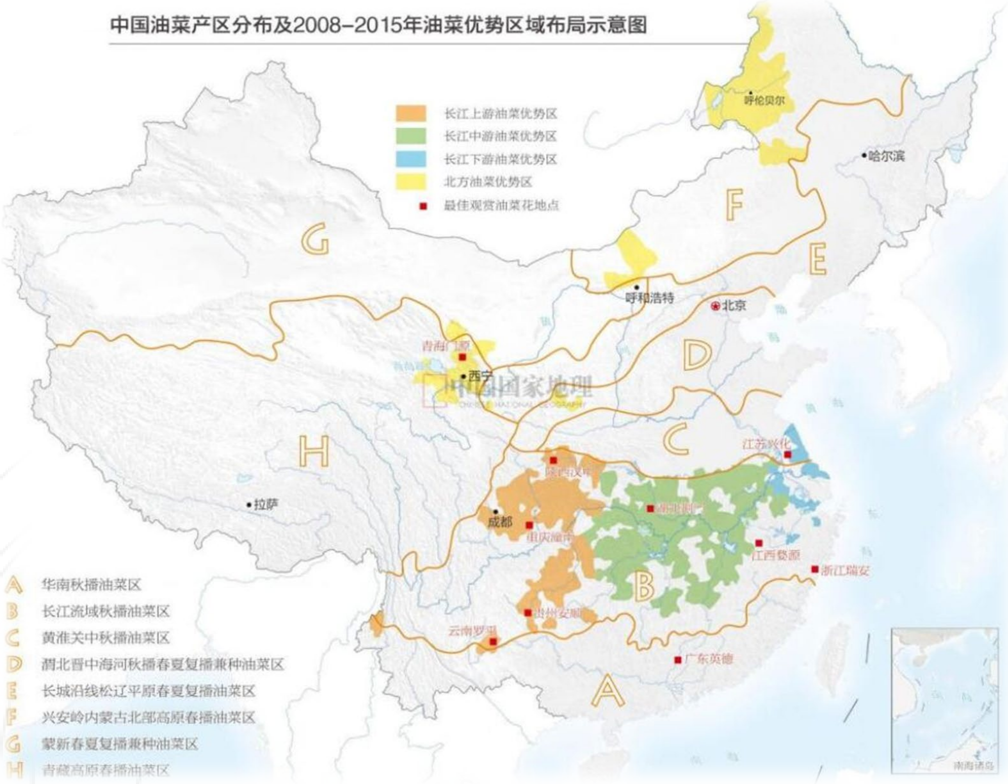
- Oil biosynthesis pathway
- Seed development
- Transcriptional regulation
- Environmental factors
- Cytoplasmic effect
- Carbon partitioning （oil-protein)
- Source-translocation-sink

……

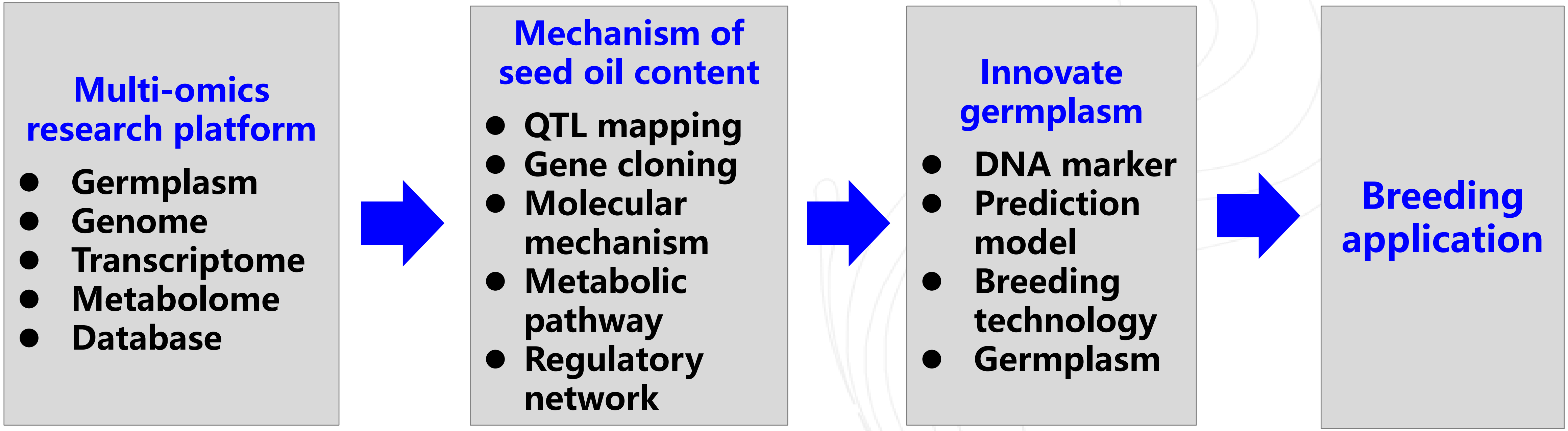
Yang et al., Plant Communications (2022)

Seed oil content is a key trait of rapeseed

- Oil content of cultivars in China is averagely 43%
- Huge potential to improve oil content (over 60%)
- Increasing oil content can increase oil production （43%→50%，oil production ↑16%）



Our focus: molecular mechanism and regulatory network of seed oil content trait in rapeseed



Multi-omics research platform

- Germplasm
- Genome
- Transcriptome
- Metabolome
- Database

Mechanism of seed oil content

- QTL mapping
- Gene cloning
- Molecular mechanism
- Metabolic pathway
- Regulatory network

Innovate germplasm

- DNA marker
- Prediction model
- Breeding technology
- Germplasm

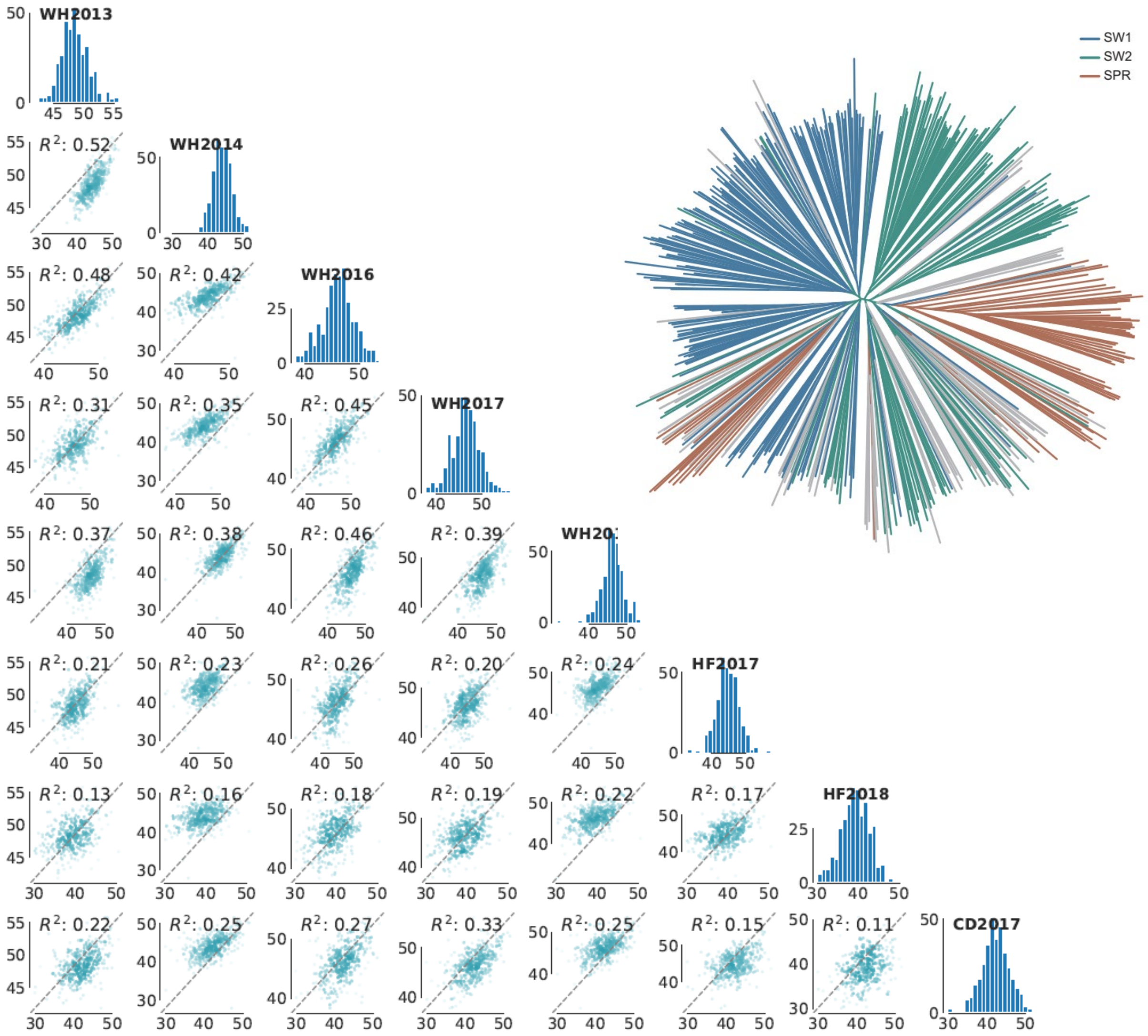
Breeding application

① **Genome- and transcriptome-wide association studies of seed oil content (SOC)**

② Multi-omics analysis dissects the genetic architecture of seed coat content （SCC)

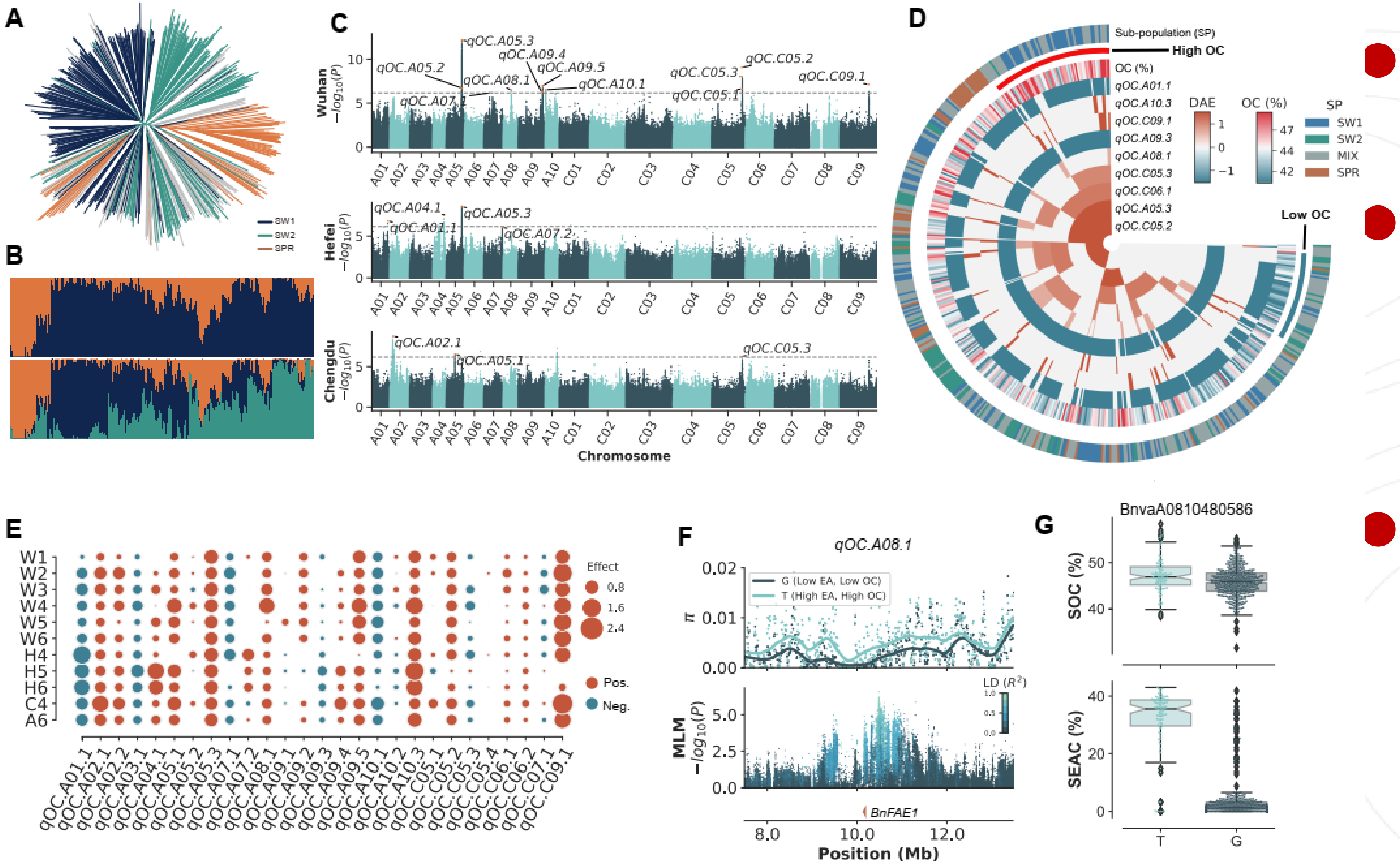
③ Mining gene impacting SOC through metabolome analysis

Re-sequenced an association population



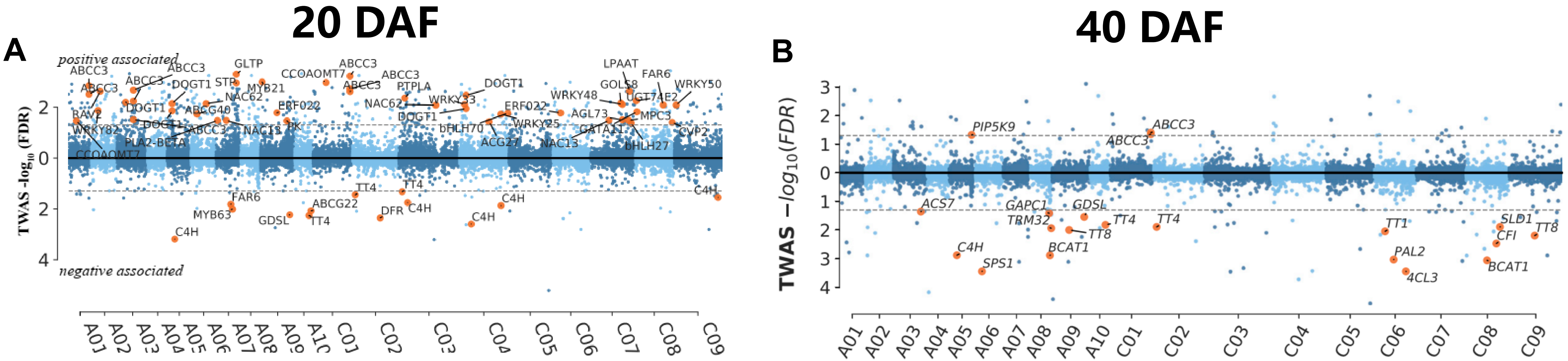
- Seed oil content (SOC) varies a lot
- Genomic variations of 505 accessions
- Seed transcriptome (20 DAF) of 309 accessions
- Seed transcriptome (40 DAF) of 274 accessions

Mapping oil-related QTLs by GWAS



- Mapped 27 reliable QTLs
- Analyzed the distribution of these QTLs in the population
- Evaluated the effect of each QTL

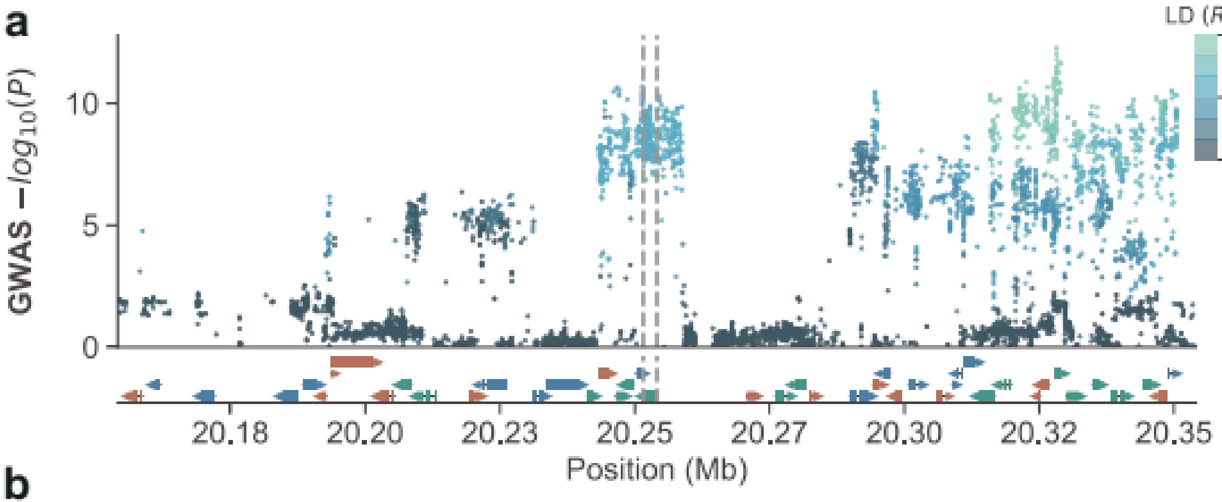
Prediction of candidate genes by TWAS



- **605 （20 DAF) and 148 （40 DAF) genes significantly associated to SOC**
- **61 genes simultaneously identified in two stages**

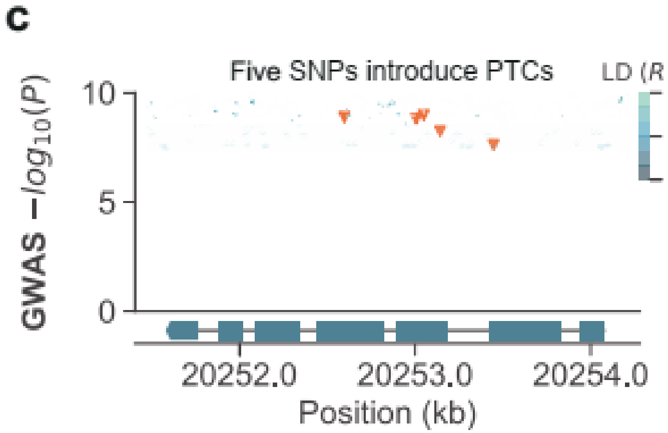
Tang et al., Molecular Plant (2021)

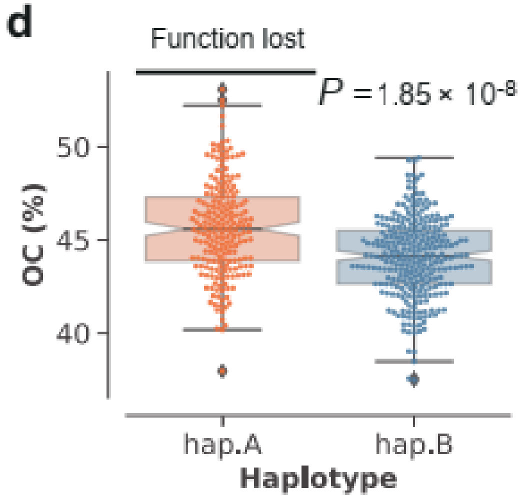
Prediction of candidate gene of *qOC.A05.3*



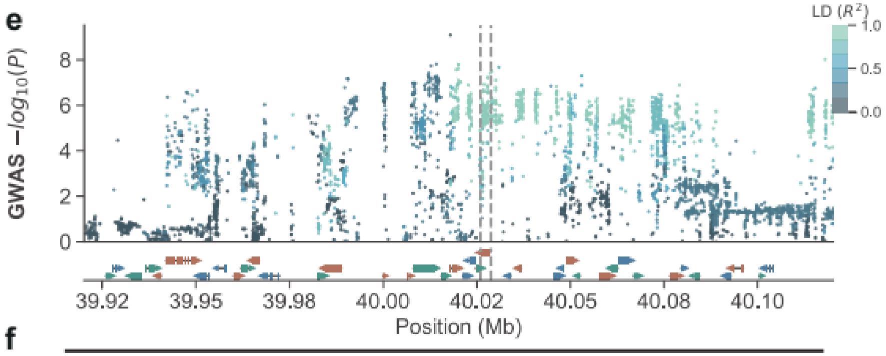
b

Gene	GF	EV	EE	EH	SS	Description
BnaA05g28570D	0.57	0.82	0.11	0.56	1.17	Probable methyltransferase PMT6
BnaA05g28620D	0.38	0.00	0.62	0.72	0.78	Phosphatidylinositol 4-phosphate 5-kinase 9
BnaA05g28720D	0.35	0.04	0.03	0.80	0.59	Protein kinase superfamily protein
BnaA05g28870D	0.39	0.31	0.34	0.39	0.56	Protein REVEILLE 8
BnaA05g28630D	0.23	0.81	0.09	0.64	0.49	Ras-related protein RABC2b
BnaA05g28740D	0.16	0.91	0.07	1.00	0.47	Lsocitrate dehydrogenase
BnaA05g28800D	0.11	1.00	1.00	0.79	0.38	Syntaxin-71
BnaA05g28750D	0.10	0.80	0.95	1.00	0.37	SNARE-like superfamily protein
BnaA05g28770D	0.16	0.79	0.03	0.56	0.32	Probable E3 ubiquitin-protein ligase LOG2
BnaA05g28830D	0.23	0.77	0.10	0.25	0.31	Minichromosome maintenance 8



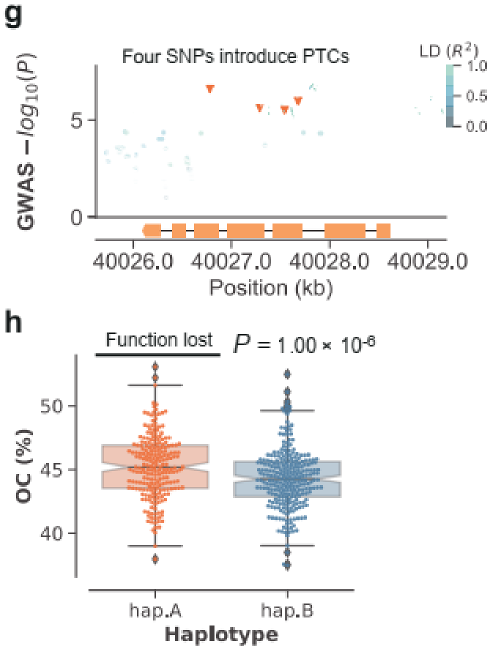


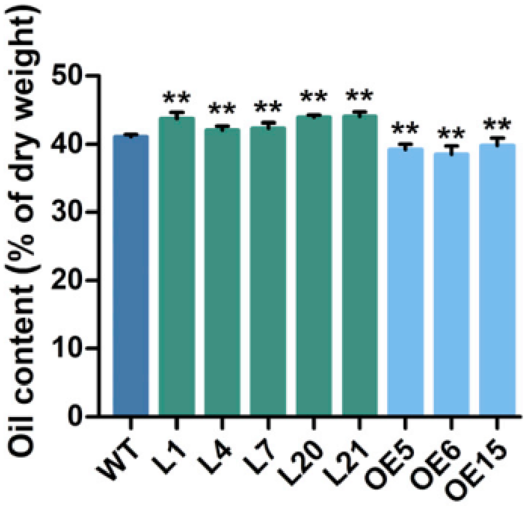
Prediction of candidate gene of *qOC.C05.3*



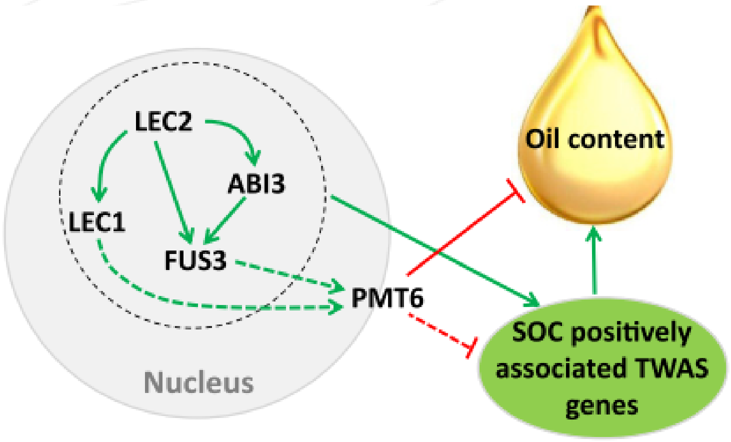
f

Gene	GF	EV	EE	EH	SS	Description
BnaC05g42890D	0.60	1.00	0.12	0.63	1.42	Probable methyltransferase PMT6
BnaC05g42850D	0.30	0.97	0.26	1.00	0.96	Plant protein 1589 of unknown function
BnaC05g42990D	0.41	0.00	0.54	0.84	0.91	Sequence-specific DNA binding transcription factors
BnaC05g42900D	0.34	0.00	0.47	0.89	0.77	Probable calcium-binding protein CML36
BnaC05g42980D	0.36	0.27	0.06	0.77	0.67	Threonine dehydratase biosynthetic
BnaC05g42780D	0.50	0.00	0.45	0.22	0.44	Phenylalanine ammonia-lyase 4
BnaC05g42910D	0.18	0.99	0.10	0.68	0.44	Gibberellin-regulated protein 13
BnaC05g42790D	0.35	0.00	0.25	0.47	0.42	Glycosyltransferase family 61 protein
BnaC05g43050D	0.26	0.00	1.00	0.29	0.41	
BnaC05g42920D	0.17	0.70	0.41	0.66	0.40	Probable inactive purple acid phosphatase 16





- *PMT6* negatively regulates SOC
- *PMT6* functions downstream of key TF-regulating SOC



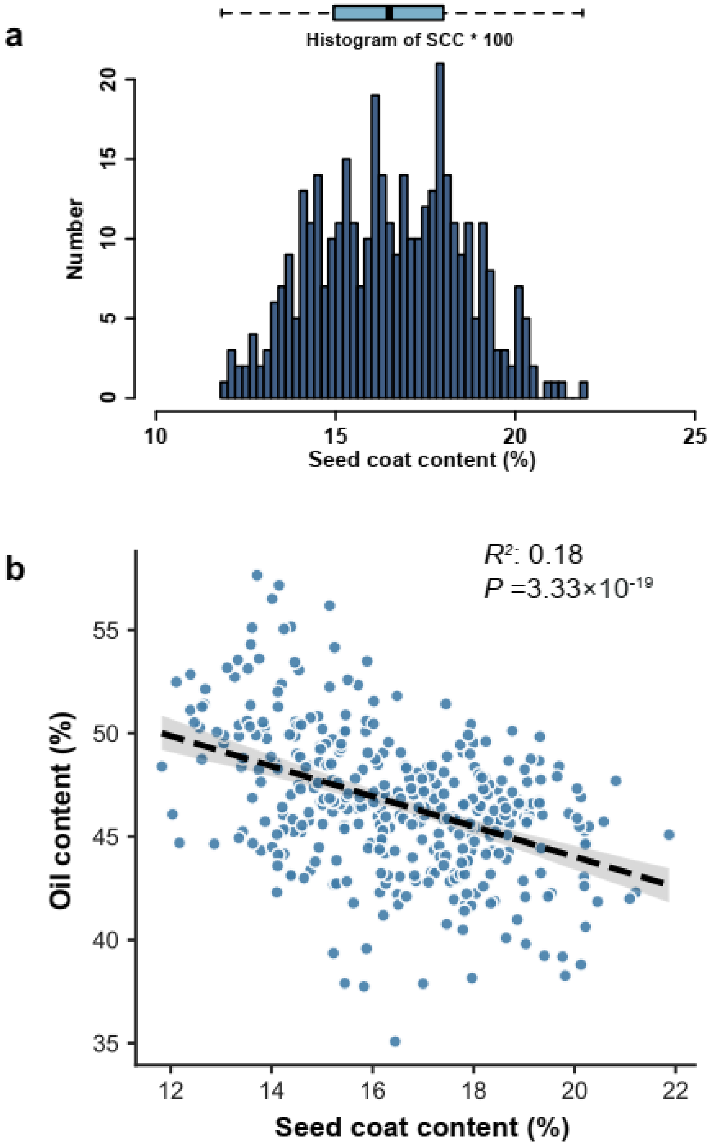
Tang et al., Molecular Plant (2021)

① Genome- and transcriptome-wide association studies of seed oil content (SOC)

② **Multi-omics analysis dissects the genetic architecture of seed coat content (SCC)**

③ Mining gene impacting SOC through metabolome analysis

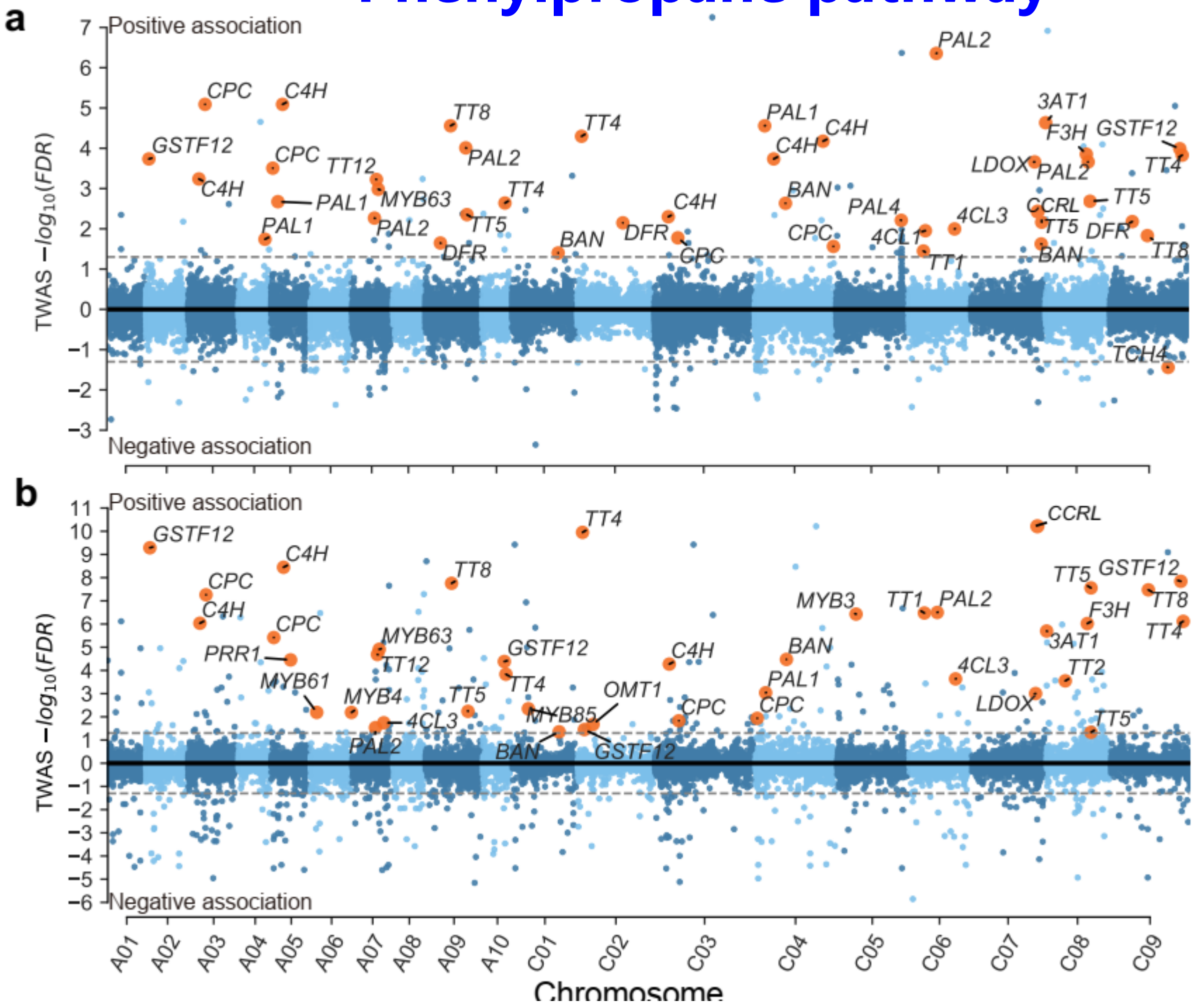
SCC is an important seed trait

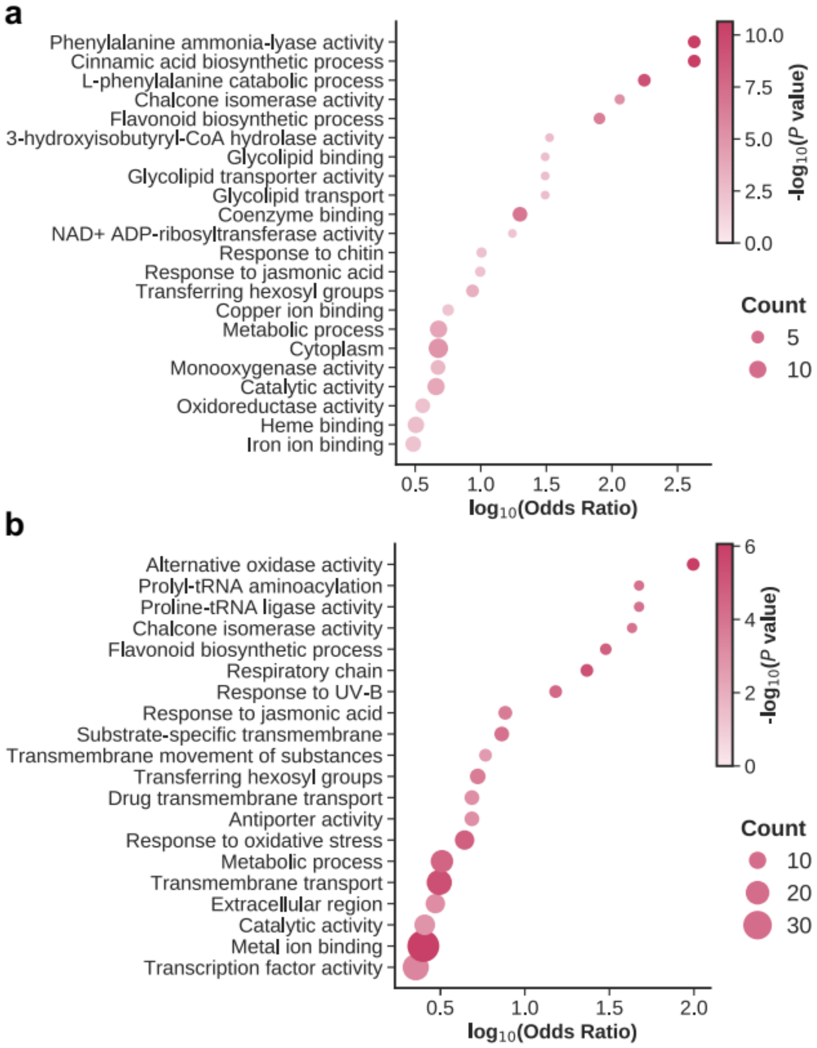


- Rape seeds are mainly composed of seed coat and embryo, and embryo is the main tissue storing oil and protein
- SCC represents percentage of seed coat weight in seeds
- SCC negatively correlates with SOC
- Reducing the seed coat content can increase the oil content of seed

Genes significantly associated with SCC

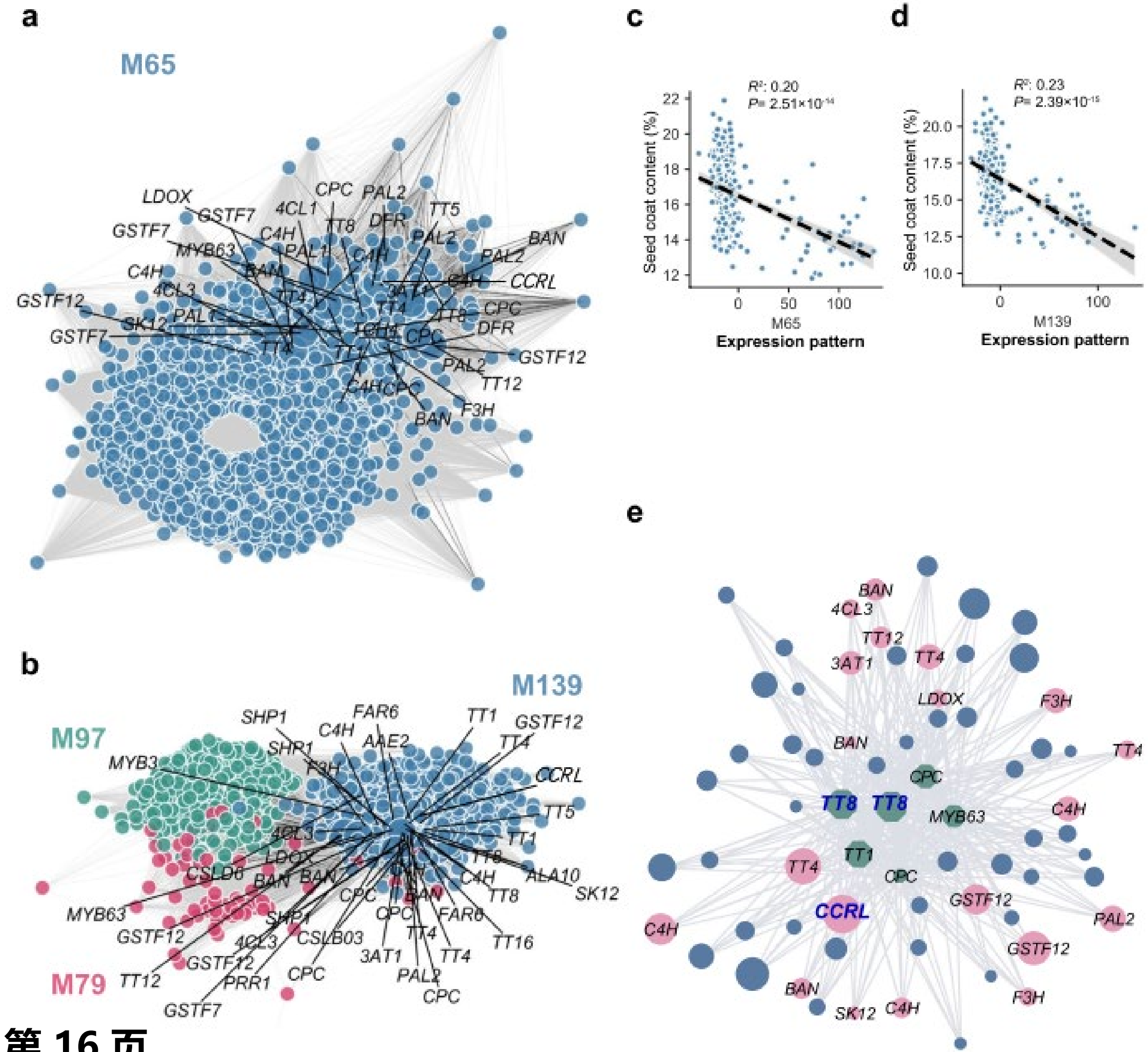
Phenylpropane pathway



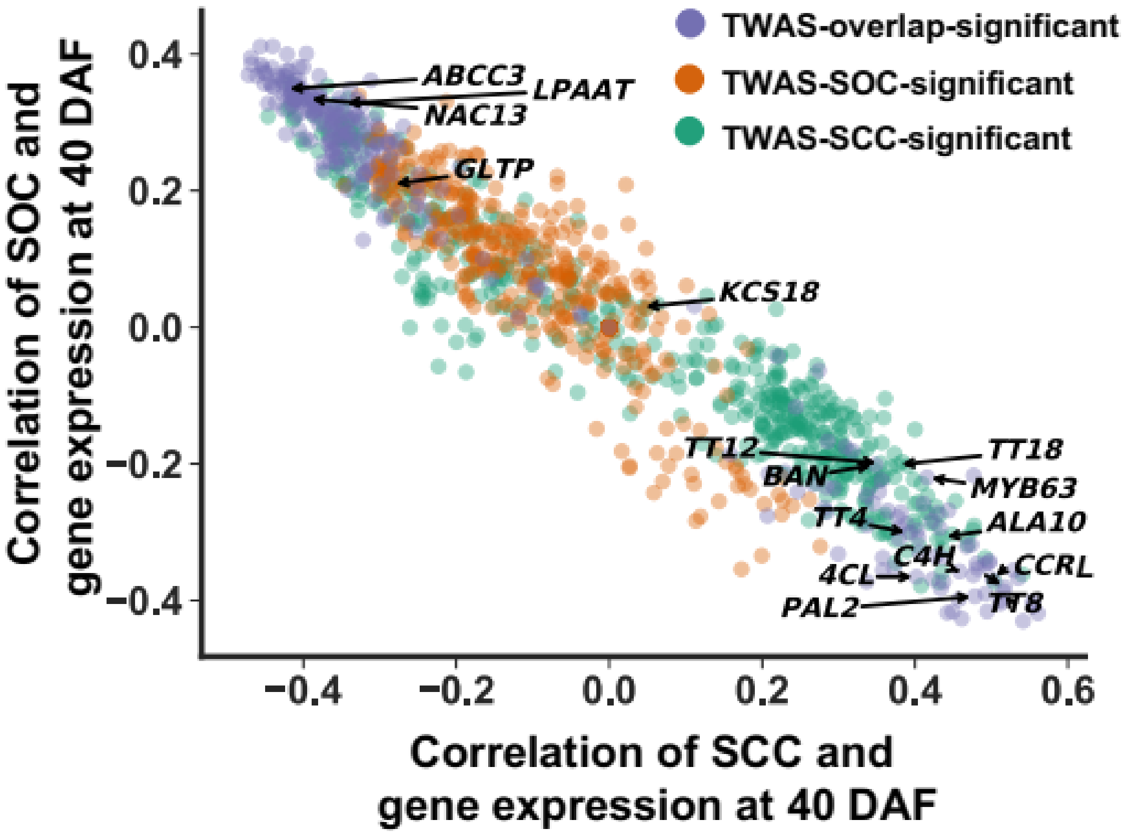


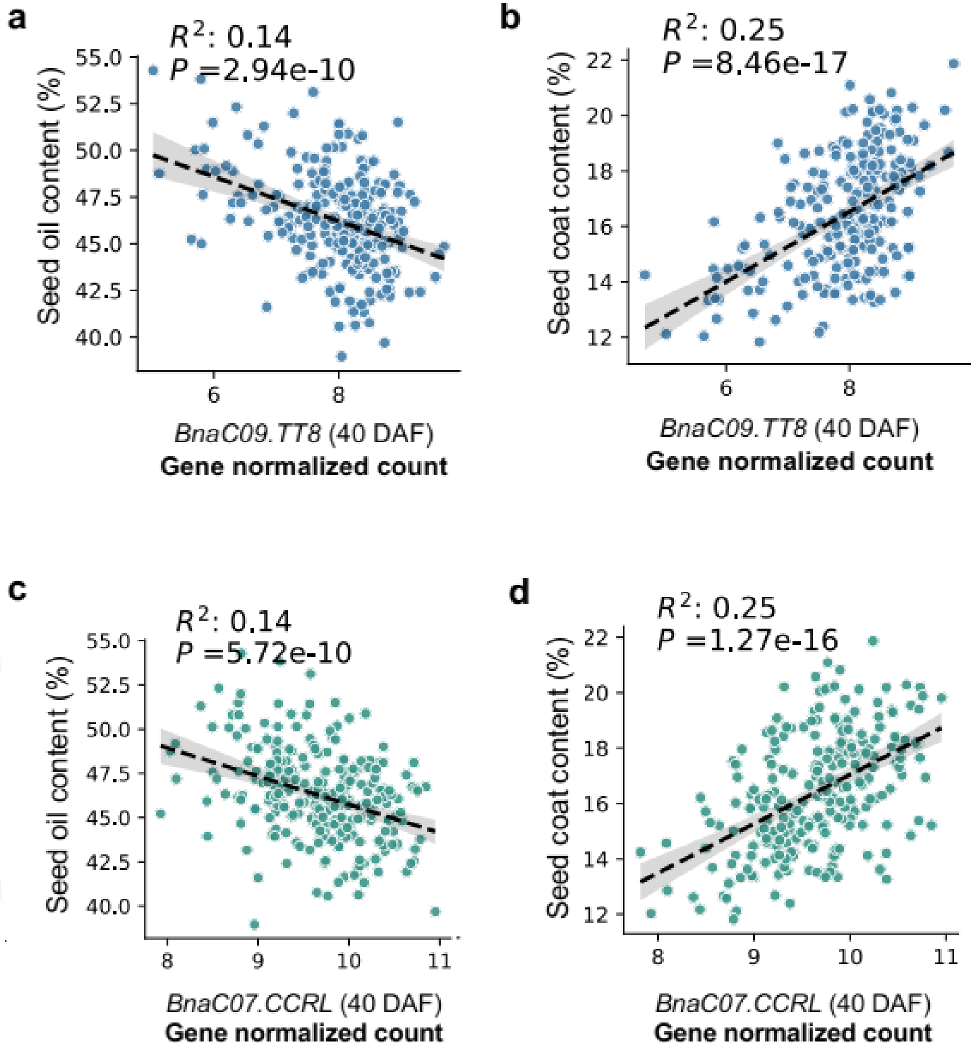
Zhang et al., Genome Biology (2022)

SCC-related gene modules and co-expression networks



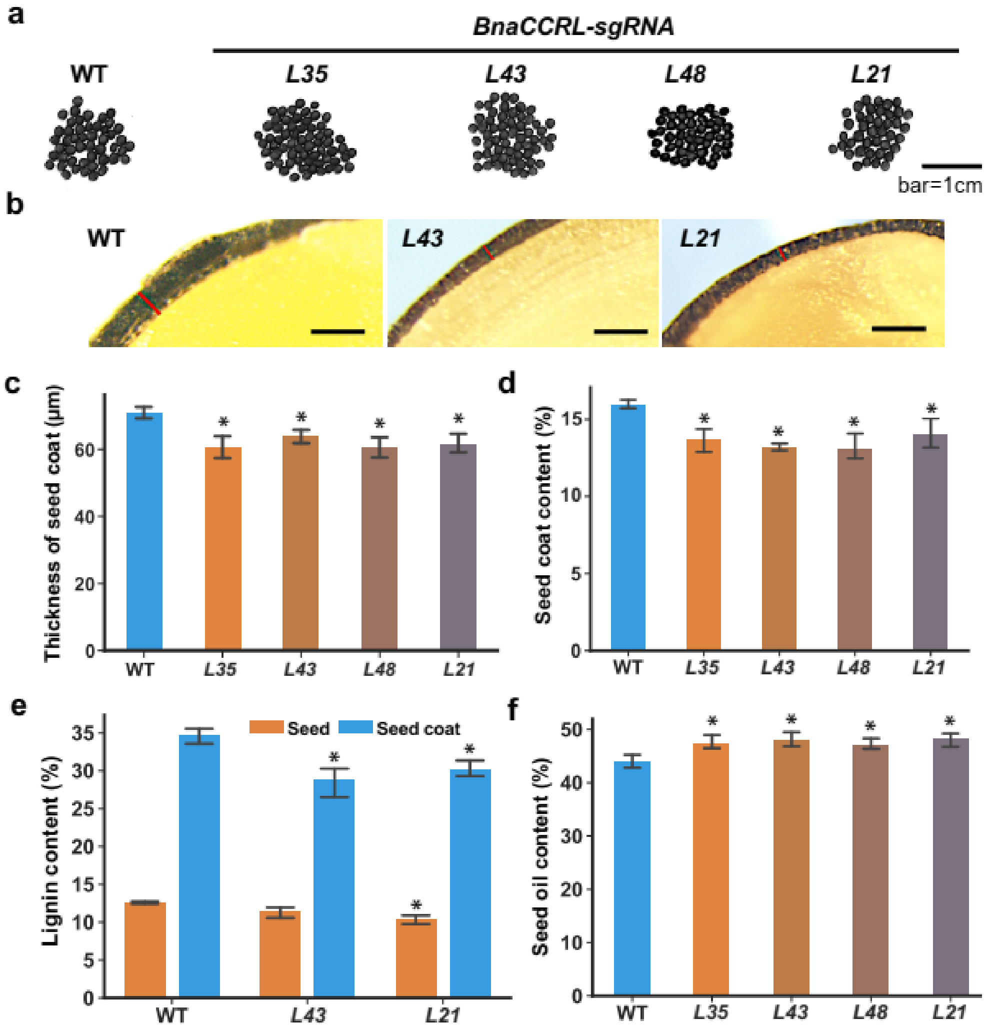
Further mining key genes

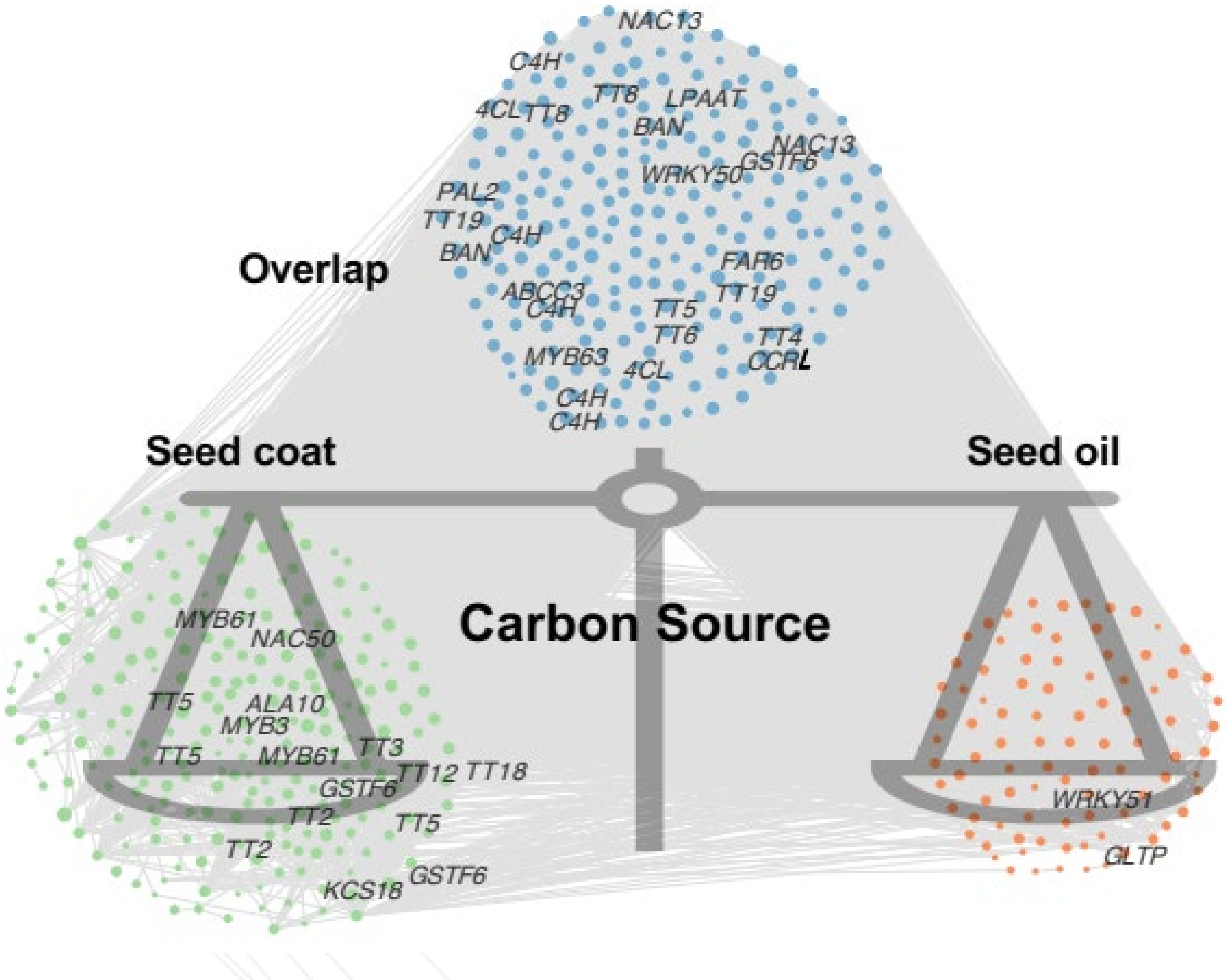




Zhang et al., Genome Biology (2022)

BnaCCRL affects both SCC and SOC





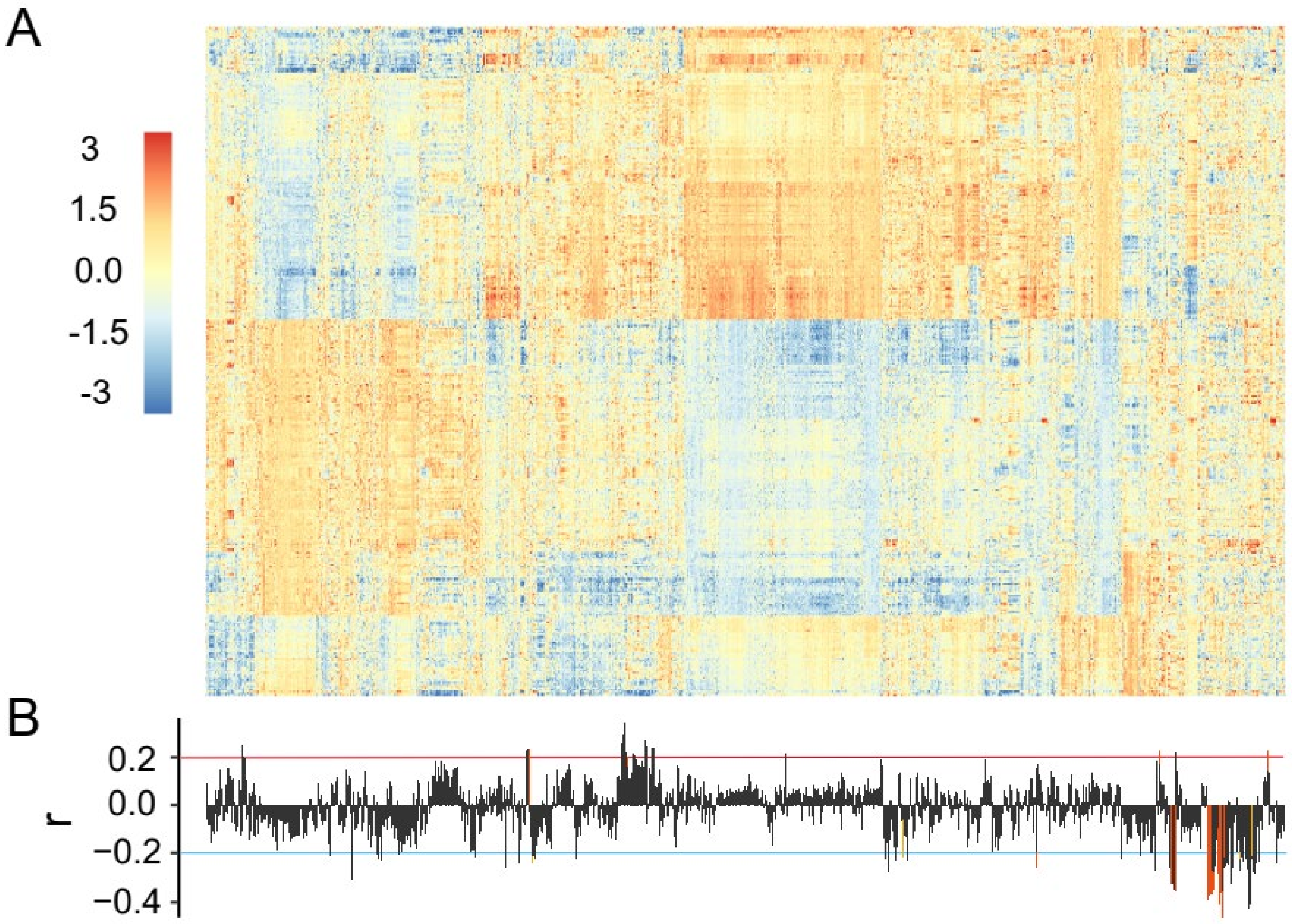
Zhang et al., Genome Biology (2022)

① Genome- and transcriptome-wide association studies of seed oil content (SOC)

② Multi-omics analysis dissects the genetic architecture of seed coat content (SCC)

③ **Mining gene impacting SOC through metabolome analysis**

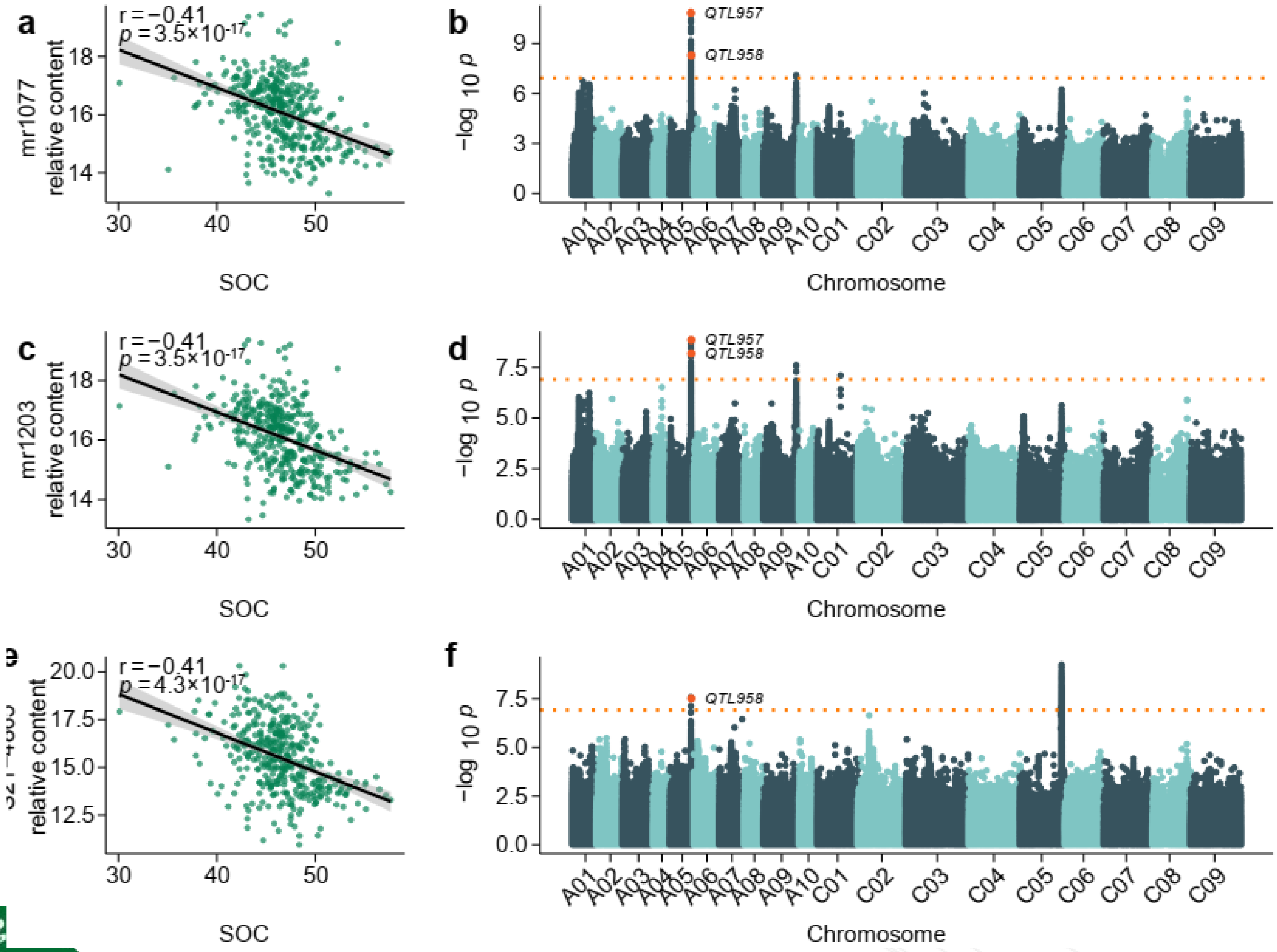
Mining loci/gene through metabolome analysis



- **2,172 metabolites of mature seeds were quantified by LC-MS/MS**
- **131 marker metabolites were identified to be correlated with SOC (*r*>0.2)**

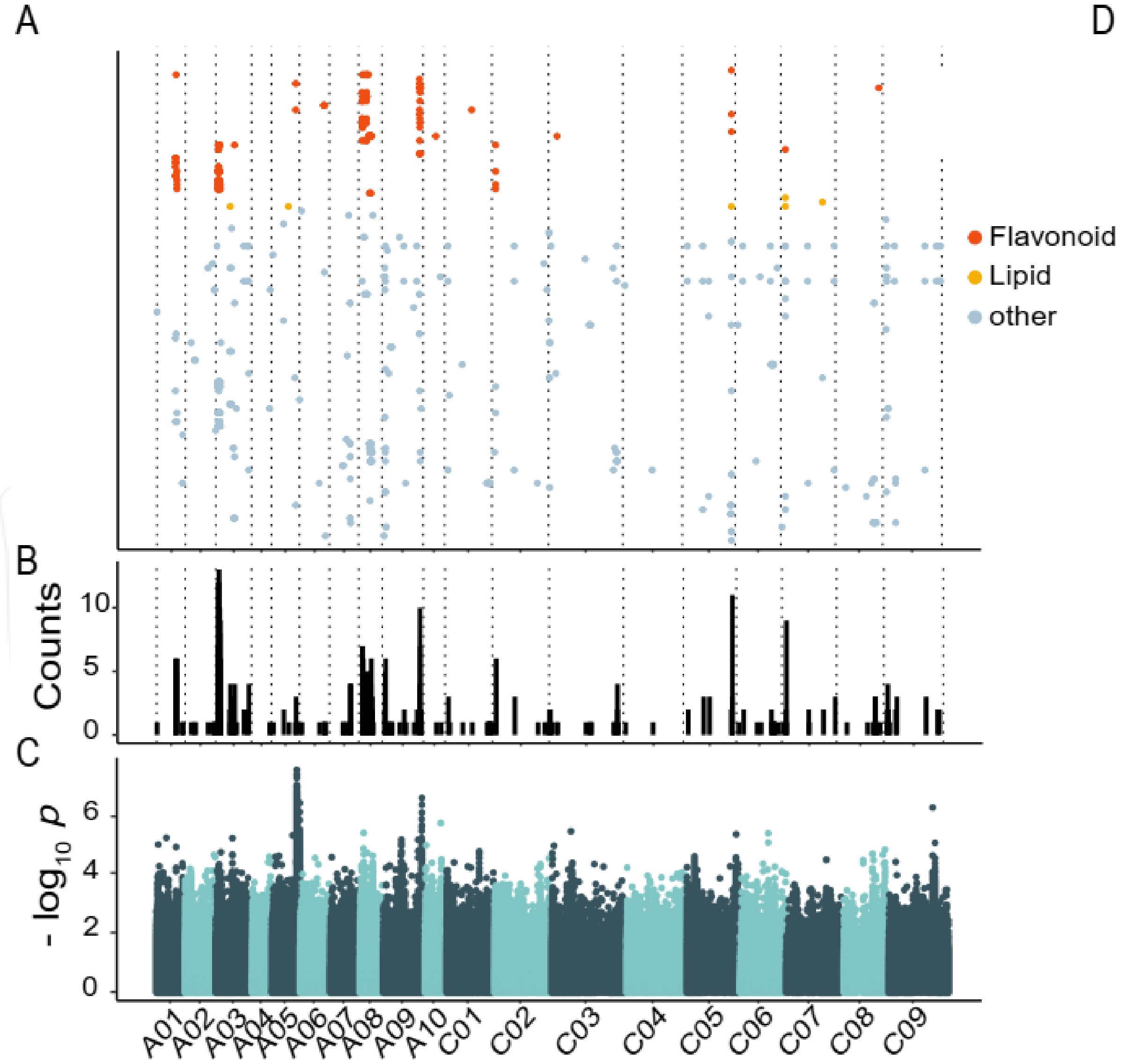
Li et al., Genome Biology (2023)

Marker metabolites correlated with SOC



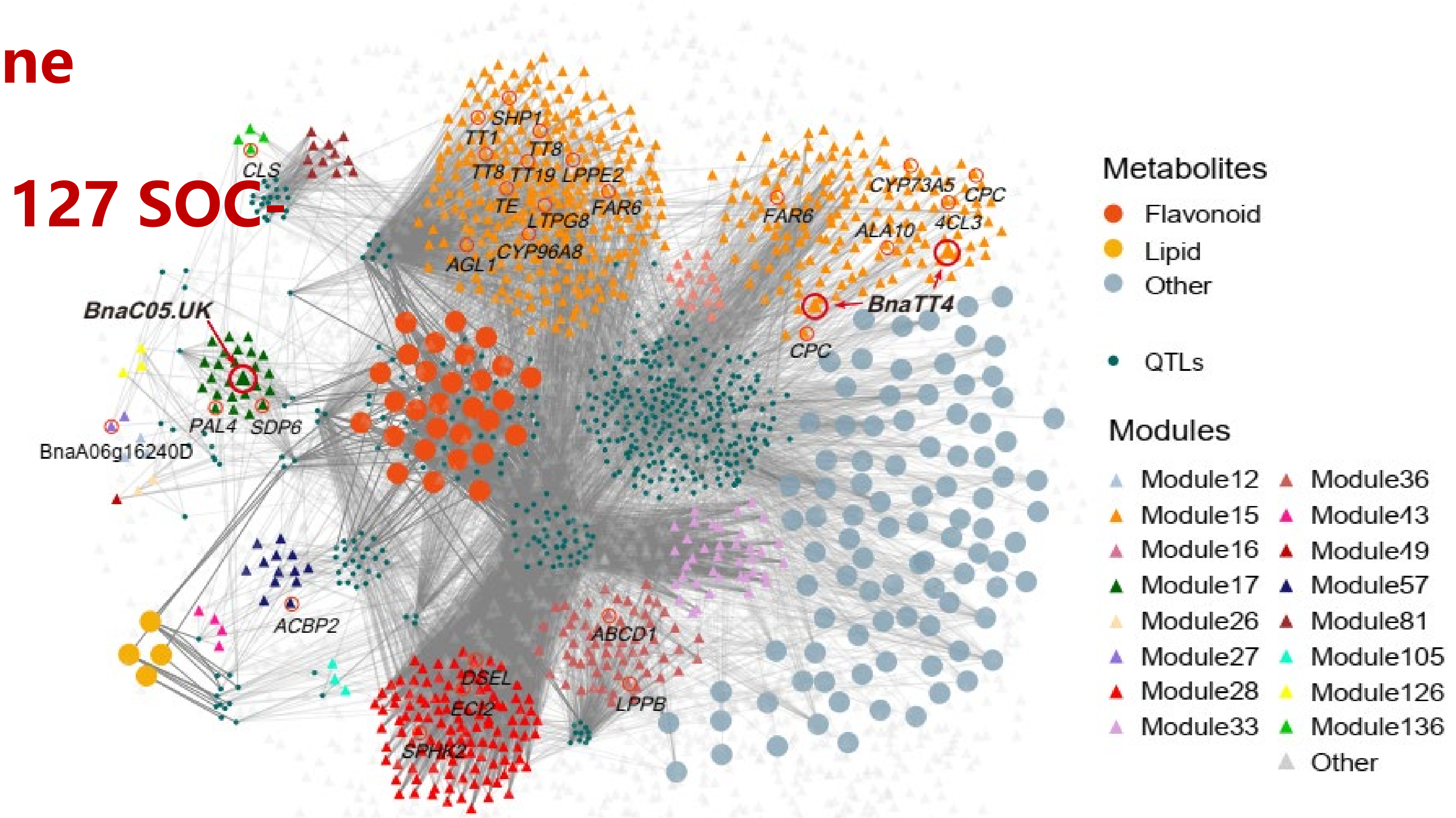
Mining genes regulating SOC through mGWAS

- **230 mQTLs have been identified**
- **23 mQTLs on A01, A03, A08, A09, C02, C05 and C07 were detected by more than 5 SOC-correlated metabolites**
- **6 mQTLs co-localized with our previous mapped SOC QTLs**

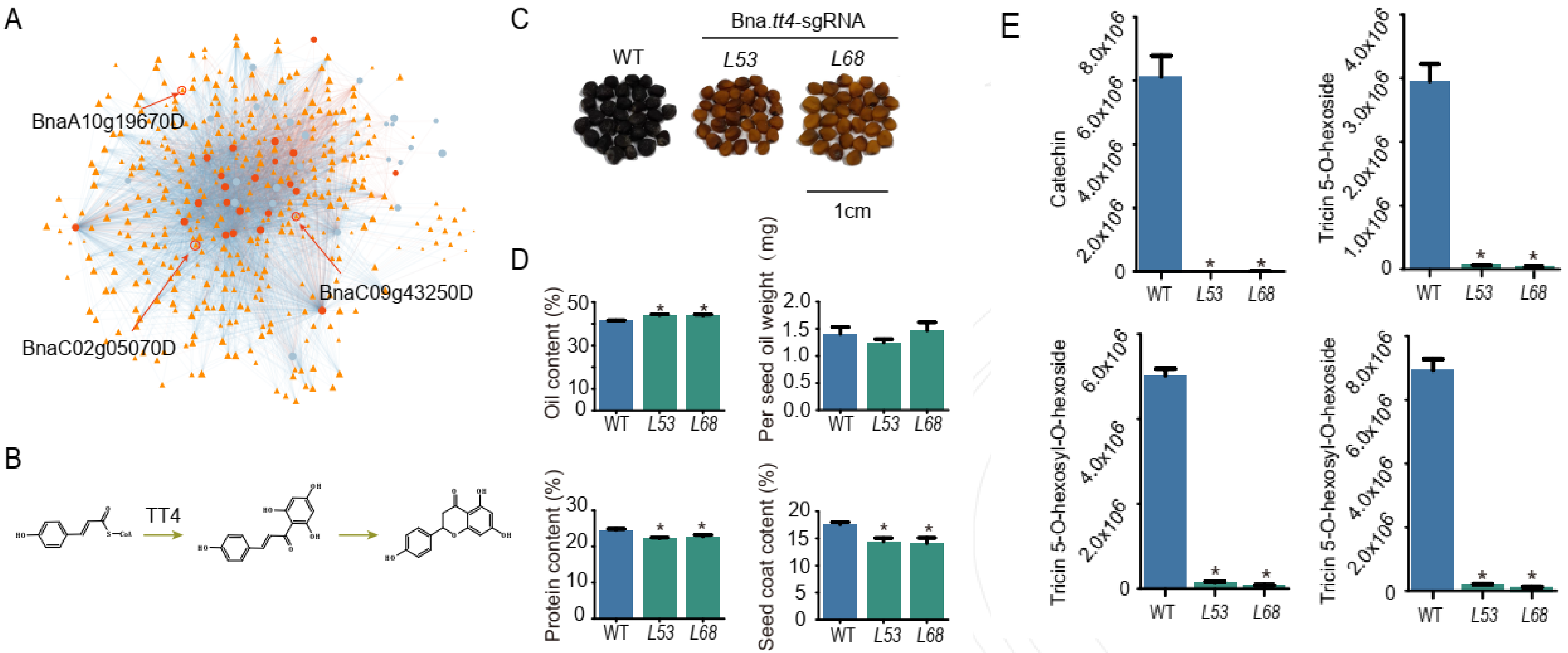


Conduction of triple relationship network discovers promising genes for improving SOC

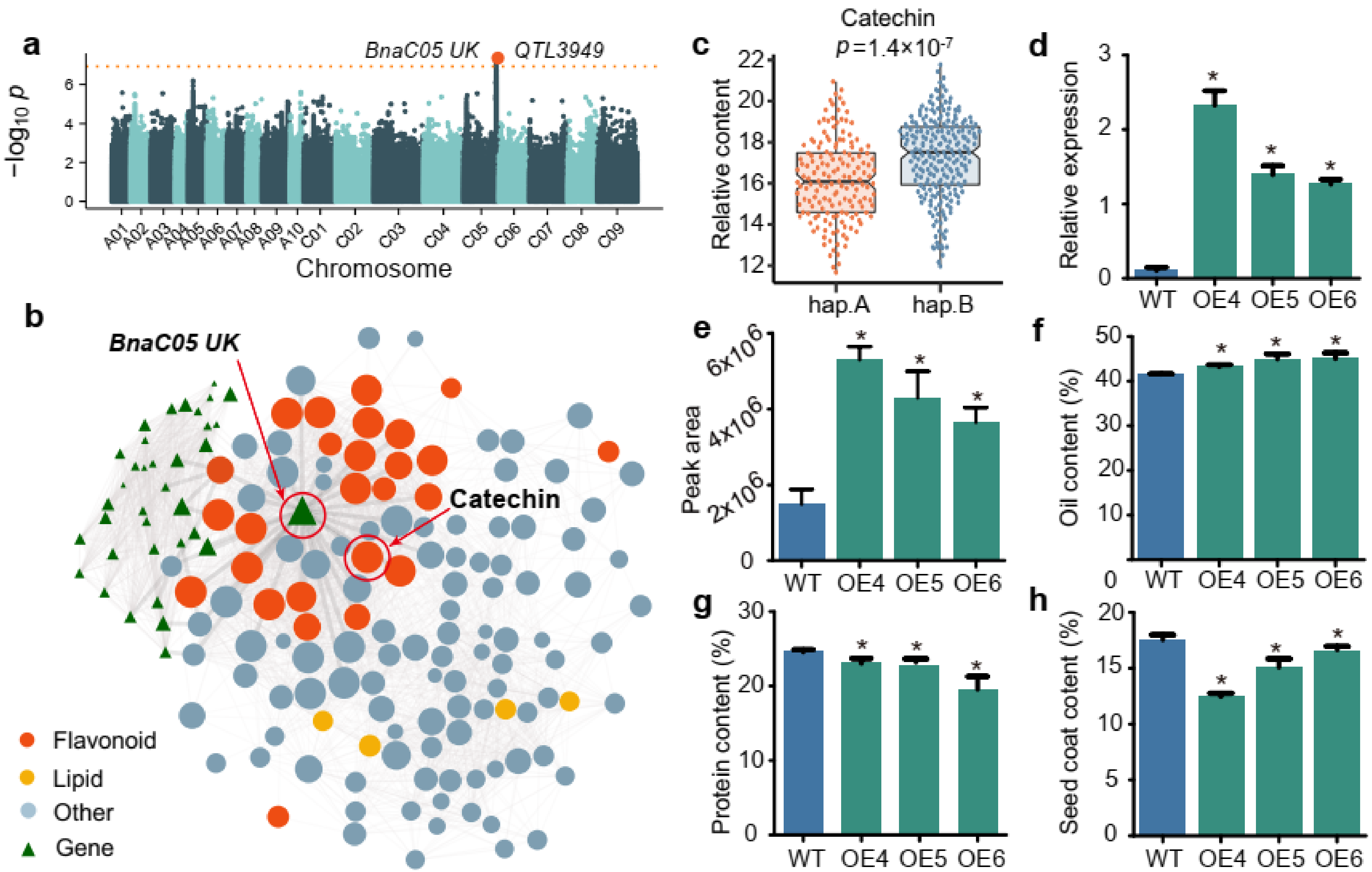
- **Metabolite-mQTL-eQTL-gene**
- **444 QTLs, 3,813 genes and 127 SOC-related metabolites**



Validation of a key gene impacting SOC



Validation of a novel gene impacting SOC

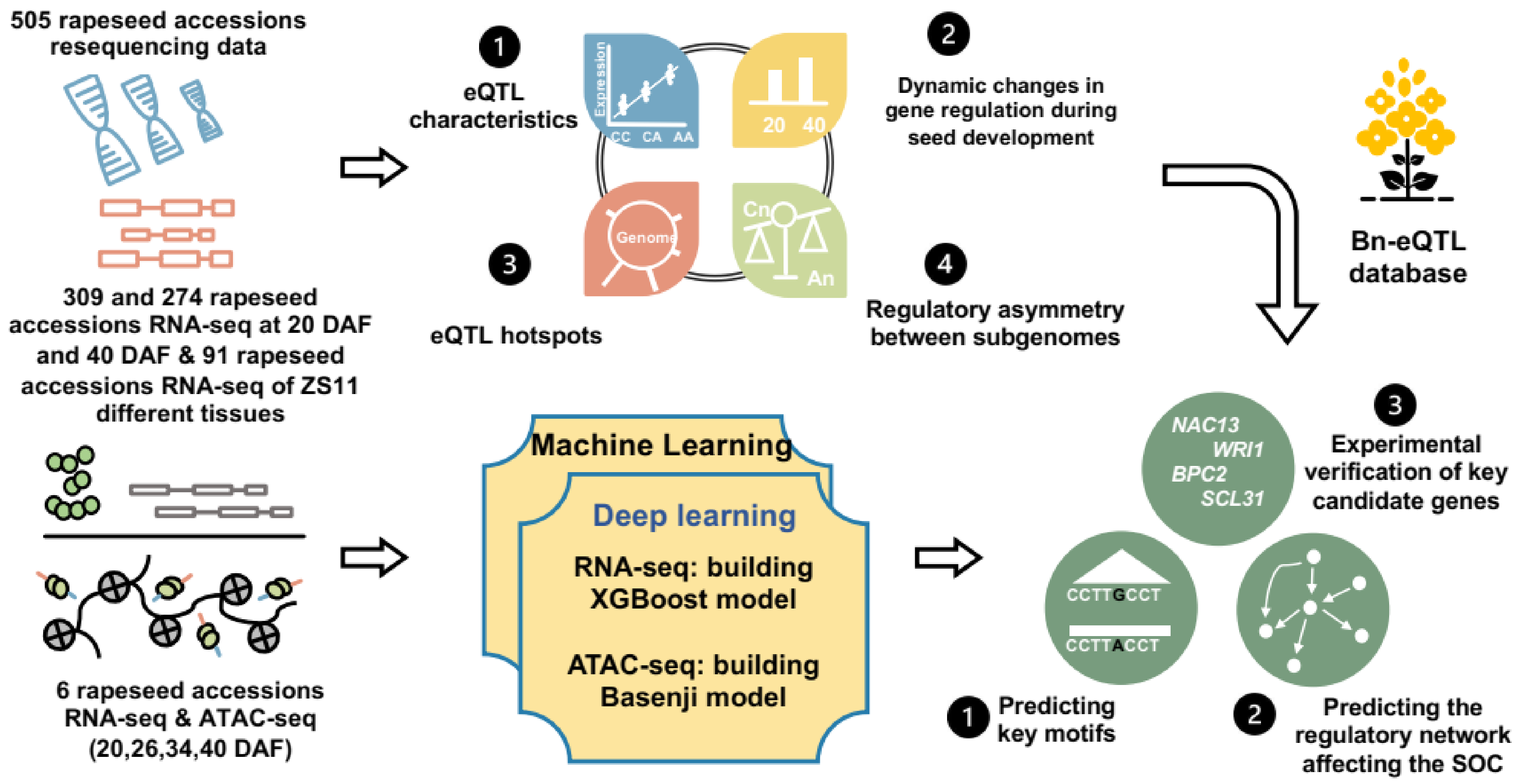


Li et al., Genome Biology (2023)

Successfully cloned more than 10 genes

Gene	Publications	Verification
BnaNPC6	Cai et al., New Phytol (2020)	Arabidopsis, Camelina, and oilseed rape
BnaPMT6	Tang et al., Mol Plant (2021)	*B. napus*
BnaCCRL	Zhang et al., Genome Biol (2022)	*B. napus*
BnaNTT1	Hong et al., Cell Rep (2022)	*B. napus*
BnaPPT1, BnaPPT2	Tang et al., J Adv Res (2022)	*B. napus*
BnaNTT2	Xia et al., Mol Breed (2022)	*B. napus*
BnaBASS2	Tang et al., Plant Biotechnol J (2022)	*B. napus*
BnaA09.SCL31	Tan et al., Genome Biol (2022)	Arabidopsis
BnaFAX1	Li, et al., Mol Breed (2023)	*B. napus*
LncRNAs	Li et al., Biotechnol Biofuels Bioprod (2023)	*B. napus*
BnaC05.UK	Li et al., Genome Biol (2023)	*B. napus*

Integrate multi-layered data to establish the gene networks involved in seed oil accumulation



505 rapeseed accessions resequencing data

309 and 274 rapeseed accessions RNA-seq at 20 DAF and 40 DAF & 91 rapeseed accessions RNA-seq of ZS11 different tissues

6 rapeseed accessions RNA-seq & ATAC-seq (20,26,34,40 DAF)

① eQTL characteristics

② Dynamic changes in gene regulation during seed development

③ eQTL hotspots

④ Regulatory asymmetry between subgenomes

Bn-eQTL database

Machine Learning

Deep learning

RNA-seq: building XGBoost model

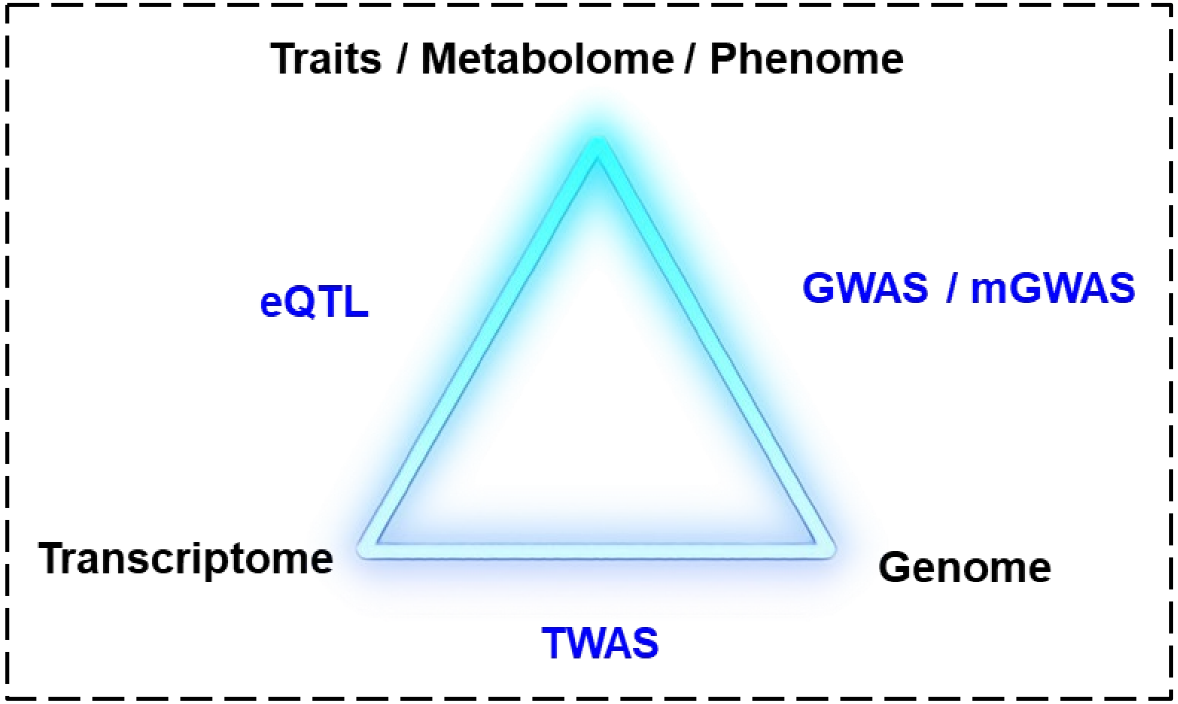
ATAC-seq: building Basenji model

① Predicting key motifs

② Predicting the regulatory network affecting the SOC

③ Experimental verification of key candidate genes

Construction of comprehensive molecular regulatory network for oil synthesis


Traits / Metabolome / Phenome

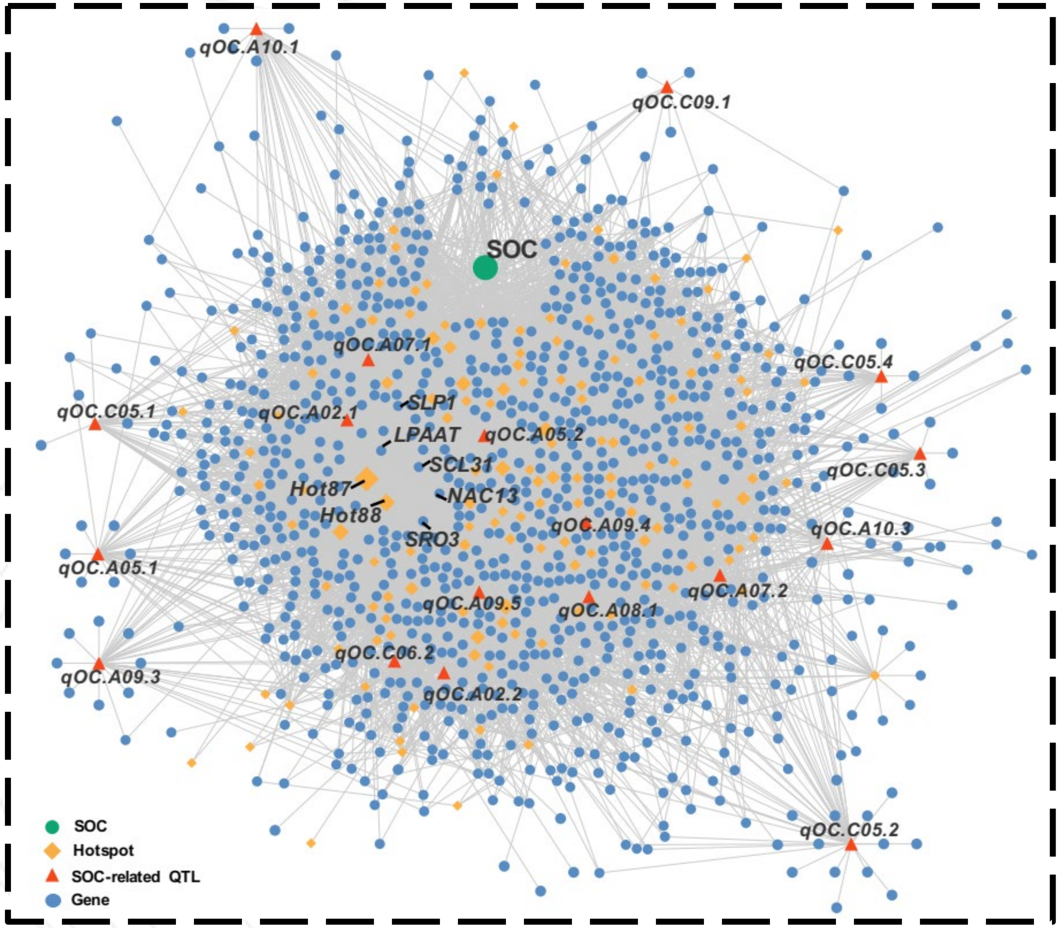
eQTL

GWAS / mGWAS

Transcriptome

Genome

TWAS



Tan et al., Genome Biology (2022)

Summary

- Generated multi-omics data resources
- Efficient gene mining based on multi-omics analysis
- Prediction of gene function by multi-omics analysis
- Construction of regulatory network
- Precisely breed rapeseed with high seed oil content driven by knowledge of multi-omics analysis

Acknowledgement

Collaboration

Huazhong Agricultural University
Prof. Kede Liu
Prof. Weibo Xie
Prof. Lingling Chen
Prof. Qingyong Yang

Donald Danforth Plant Science Center
Prof. Sam Wang

University of North Texas
Prof. Kent Chapman

Kansas State University
Prof. Ruth Welti

Rapeseed group

Guo Lab

Prof. Hu Zhao
Dr. Shan Tang
Dr. Zengdong Tan
Dr. Yuting Zhang
Dr. Long Li
Hui Zhang

国家自然科学基金委员会
NSFC