LIÈGE université
Gembloux
Agro-Bio Tech

# Investigating the genetic basis and regulatory mechanism of folate metabolism in maize (*Zea mays*)

**Tong Lian**

**Promoter: Prof. Hervé Vanderschuren**

**Co-Promoter: Prof. Chunyi Zhang**

**2022**

COMMUNAUTÉ FRANÇAISE DE BELGIQUE
UNIVERSITÉ DE LIÈGE – GEMBLOUX AGRO-BIO TECH

# Investigating the genetic basis and regulatory mechanism of folate metabolism in maize (*Zea mays*)

**Tong Lian**

Dissertation originale présentée en vue de l'obtention du grade de docteur en sciences agronomiques et ingénierie biologique

Promoteurs: Prof. Hervé Vanderschuren
Prof. Chunyi Zhang

Année civile: 2022

# Résumé

**Tong Lian. (2022). Étude des bases génétiques et du mécanisme de régulation du métabolisme des folates chez le maïs (Zea mays) (thèse de doctorat en anglais)**. Gembloux, Belgique, Gembloux Agro-Bio Tech, Université de Liège, 132 p., 12 tableaux, 29 figures.

**Résumé** —Les folates, un groupe de vitamines B essentielles, jouent un rôle crucial dans la biosynthèse de l'ADN, le métabolisme des acides aminés et la méthylation de l'ADN au cours du développement et de la croissance des organismes. L'alimentation quotidienne est la principale source de folates pour les êtres humains. La faible consommation de folates peut entraîner une augmentation du risque de plusieurs maladies graves telles que les cancers et les anomalies du tube neural. Les teneurs en folate des plantes cultivées, en particulier des céréales, sont assez faibles. Par conséquent, il est nécessaire d'améliorer l'accumulation de folate dans les cultures, une approche appelée biofortification, pour atténuer le problème de carence en folate dans le monde. À cette fin, il est impératif de comprendre la base génétique et le mécanisme de régulation du métabolisme des folates dans les cultures. Ici, nous avons conçu deux projets pour identifier les voies métaboliques clés ou les gènes qui peuvent contribuer à l'accumulation de folate dans le grain de maïs.

**(1) L'analyse comparative du transcriptome révèle les mécanismes d'accumulation de folate dans les grains de maïs**

Auparavant, la complexité de l'accumulation de folate dans les premiers stades du développement du grain de maïs a été signalée, mais les mécanismes restent flous. Dans cette étude, deux lignées consanguines de maïs, DAN3130 et JI63, avec différents modèles et niveaux d'accumulation de folate dans les grains matures ont été utilisées pour étudier la régulation transcriptionnelle du métabolisme du folate par analyse comparative du transcriptome. Il a été démontré que l'accumulation de folate pendant le DAP 24 jusqu'à la maturité du grain était contrôlée par les voies circonjacentes de la biosynthèse du folate, telles que le métabolisme du pyruvate, le métabolisme du glutamate et le métabolisme de la sérine/glycine. De plus, les différences d'accumulation de folate entre ces deux lignées consanguines se sont révélées être liées aux gènes impliqués dans le métabolisme du folate, y compris ceux de la branche ptéridine, de la branche para-aminobenzoate, du cycle sérine/tétrahydrofolate (THF)/5-méthyltétrahydrofolate, et conversion du monoglutamate de THF en polyglutamate de THF. Ces observations ont permis de mieux comprendre les

mécanismes sous-jacents au métabolisme des folates lors de la formation du grain de maïs, ce qui a été utile pour la recherche sur la biofortification des folates chez le maïs.

**(2) Cartographie génétique des QTL de folate à l'aide d'une population ségrégée de maïs (*Zea mays L.*)**

Augmenter l'accumulation de folate dans les parties comestibles des cultures est d'une grande importance pour la santé humaine. La sélection moléculaire est une stratégie réalisable et efficace pour la biofortification des folates, mais quelque peu limitée par le manque de connaissances sur le métabolisme des folates au niveau moléculaire. Dans cette étude, nous avons rapporté la cartographie génétique des locus de caractères quantitatifs (QTL) liés aux niveaux d'accumulation de folates en utilisant une population ségrégée traversée par deux lignées de maïs, l'une riche en folates (GEMS31) et l'autre faible en folates (DAN3130). En conséquence, deux QTL sur le chromosome 5 ont été obtenus par association du séquençage de l'exome entier avec le profilage du folate du noyau. Ces QTL ont été confirmés par une analyse de ségrégation en masse à l'aide d'ADN regroupé dans F6 et d'un profilage de folate de noyau dans F7, avec un chevauchement avec les QTL identifiés dans une autre population ségréguée. Un gène candidat, nommé ZmCTM, a été identifié comme un gène codant pour une protéine de liaison aux folates qui jouait un rôle important dans le métabolisme des folates chez le maïs. La perte de la fonction ZmCTM a multiplié par trois l'accumulation de 5-méthyl-tétrahydrofolate. Nous avons conclu que ZmCTM peut participer au métabolisme du folate en convertissant le 5-méthyl-tétrahydrofolate en d'autres dérivés du folate dans le maïs.

Mots clés : métabolisme des folates, transcriptome, WGCNA, analyse QTL, BSA, maïs

# Abstract

**Tong Lian. (2022). Investigating the genetic basis and regulatory mechanism of folate metabolism in maize (*Zea mays*) (PhD Dissertation in English).** Gembloux, Belgium, Gembloux Agro-Bio Tech, University of Liège, 132 p., 12 tables, 29 figures.

**Abstract** — Folates, one group of the essential B vitamins, play a crucial role in DNA biosynthesis, amino acid metabolism, and DNA methylation during the development and growth of organisms. Daily diet is the major source of folates for human beings. The low-level intake of folates may cause increase of the risk of number of serious diseases such as cancers and neural tube defects. The folate contents in crop plants, especially cereals, are quite low. Therefore, it's necessary to improve the folate accumulation in crops, an approach called biofortification, to alleviate the folate deficiency problem worldwide. To this end, it's imperative to understand the genetic basis and regulatory mechanism of folate metabolism in crops. Here we have designed two projects to identify the key metabolic pathways or genes that may contribute to folate accumulation in maize kernel.

**(1) Comparative transcriptome analysis reveals mechanisms of folate accumulation in maize grains**

Previously, the complexity of folate accumulation in the early stages of maize kernel development was reported, but the mechanisms remain unclear. In this study, two maize inbred lines, DAN3130 and JI63, with different folate accumulation patterns and levels in mature kernels were used to investigate the transcriptional regulation of folate metabolism by comparative transcriptome analysis. It was demonstrated that the folate accumulation during DAP 24 to kernel maturity was controlled by the circumjacent pathways of folate biosynthesis, such as pyruvate metabolism, glutamate metabolism, and serine/glycine metabolism. In addition, the differences in folate accumulation between these two inbred lines were found to be related to those genes involved in folate metabolism, including those in pteridine branch, para-aminobenzoate branch, serine/tetrahydrofolate (THF)/5-methyltetrahydrofolate cycle, and conversion of THF monoglutamate to THF polyglutamate. Those observations provided insight into the mechanisms underlying folate metabolism during maize kernel formation, thus being helpful for folate biofortification research in maize.

**(2) Genetic mapping of folate QTLs using a segregated population in maize (*Zea mays L.*)**

To increase folate accumulation in edible parts of crops is of great importance for human health. Molecular breeding is a feasible and efficient strategy for folate biofortification, but somewhat constrained by shortage of the knowledge on folate metabolism at molecular level. In this study, we reported the genetic mapping of the quantitative trait loci (QTLs) linking with folate accumulation levels using a segregated population crossed by two maize lines, one high in folates (GEMS31) and the other low in folates (DAN3130). As a result, two QTLs on chromosome 5 were obtained by association of whole-exome sequencing with kernel folate profiling. These QTLs were confirmed by bulk segregant analysis using pooled DNA in F6 and kernel folate profiling in F7, with an overlap with the QTLs identified in another segregated population. One candidate gene, named Zm*CTM*, was identified as a gene encoding a folate-binding protein that played an important role in folate metabolism in maize. Loss of *ZmCTM* function enhanced 5-methyl-tetrahydrofolate accumulation by three folds. We concluded that ZmCTM may participate in folate metabolism by converting 5-methyl-tetrahydrofolate to other folate derivatives in maize.

Key words: folate metabolism, transcriptome, WGCNA, QTL analysis, BSA, maize

# Acknowledgments

During the past six years, I was fully engaged in the doctoral training, courses, and research projects to persue my PhD degree. From the start of the projects to the completion of my thesis, I feel full of gratitude to my supervisors, friends and family. Thanks for their kind support, guidance and encouragement for my Ph. D study.

Above all, I would like to express my appreciation to the supervisors and PhD committee members: Prof. Chunyi Zhang, Prof. Hervé Vanderschuren, Dr. Ling Jiang and Prof. Ludivine Lassois. Prof. Chunyi Zhang from Biotechnology Reaserch Institute (BRI) of Chinese Academy of Agricultural Sciences (CAAS) guided the experimental design and taught me a lot in my research life. His intelligence and open-mind thinking influence me deeply not only in research works but also in my daily life. Prof. Hervé Vanderschuren from the University of Liege (Ulg) is a very nice and friendly teacher. His professional suggestions and recommendations about my research projects help me achieve my doctoral goals. During the one-year study in Belgium, his team gave me much new insight to research career. Dr. Ling Jiang is a super responsible associate professor from the BRI, CAAS. We have cooperated for over nine years on folate research projects. Thanks for her kind and careful guidance in my thesis. Also, thanks to my PhD committee member Prof. Ludivine Lassois (from Ulg), for her suggestions to my annual research progress reports.

During the four-year study period in CAAS, many teachers gave me all kinds of experimental supports. They are Prof. Huan Wang, Prof. Xiaoduo Lu, Prof. Li Pu. Meanwhile, thanks to Dr. Qiuju Liang, Dr. Weixuan Wang, Prof. Baobao Wang, Dr. Lida Han, Ms. Suhua Li, Ms. Hui Sun and Mr. Mingjun Zhang for their warmly help during my doctoral study. I also would like to express thanks to my friends and lab members from CAAS: Dr. Wenzhu Guo, Dr. Qinghe Geng, Xing Wan, JiAn Liu, Ying Ren, Xuxia Wang, Xueli Lv, Dr. Tingting Liu, Xiaowan Hou, Nan Jing, Dr. Hao Zhang, Jia Yu, Weijun Guo, F.M.S. Azam, Pirzada Khan, Dr. MD Shariful Islam and Dr. Shahid Mehmood.

In the past one-year study in Gembloux Agro-bio Tech of ULg, my classmates give me daily support and help. They are Xiaoxu Feng, Jiang Liu, Zhenxing Shi, Xiang Du, Weidong Lin, Shangwu Liu. Also, I will never forget the friendship with my colleges in Plant Genetics laboratory: Dr. Leonard Shumbe, Dr. Syed Shan, Dr. Sara Shakir, Dr. Ivan Jauregui and Dr. Kumar Vasudevan.

Meanwhile, I would like to express my sincere thanks to my other friends: Yangyang Chen, Caiyi Wang, Fan Zeng, Maoran Chen, Huaqing Wang, Shuangfeng Shi, Yang Li, Zhen Liu, Fen Xu, Jingyang Hong, Yuzhi Li, Yongfeng Li and Yuan Ji.

Finally, thanks to my parents, sister and brother-in-law. Their selfless love and careness are the power for me to insist in my PhD study.

Tong Lian

2022

Gembloux

# Tables of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| ADC | Aminodeoxychorismate |
| ADCL | Aminodeoxychorismate lyase |
| ADCS | Aminodeoxychorismate synthase |
| DHC | 5,10-Methenyl-THF cyclohydrolase/5,10-methylene-THF dehydrogenase |
| DEG | Differentially expressed genes |
| DHFS | DHF synthase |
| DHM | Dihydromonapterin |
| DHN | Dihydroneopterin |
| DHNA | Dihydroneopterin aldolase |
| DHP | Dihydropteroate |
| DHPS | Dihydropteroate synthase |
| *E. coli* | *Escherichia coli* |
| FDH | 10-Formyltetrahydrofolate dehydrogenase |
| FPGS | Folylpolyglutamate synthetase |
| FTHS | 10-Formyltetrahydrofolate synthetase |
| GCHI | GTP cyclohydrolase I |
| GCS | Glycine cleavage system |
| GDC | Glycine decarboxylase complex |
| GGH | γ-Glutamyl hydrolase |
| GO | Gene ontology |
| GS | Gene significance |
| HMDHP | 6-Hydroxymethyldihydropterin |
| HPPK | Hydroxymethyldihydropterin pyrophosphokinase |
| HPPP | 6-Hydroxymethyl-7,8-dihydropterin pyrophosphate |
| KEGG | Kyoto encyclopedia of genes and genomes |
| MS | Methionine synthase |
| MTHFD | Methylenetetrahydrofolate dehydrogenase/cyclohydrolase |
| MTHFR | Methylenetetrahydrofolate reductase |
| MTHFS | Methenyltetrahydrofolate synthetase |
| OCM | One-carbon metabolism |
| QTL | Quantitative trait loci |
| RIL | Recombination inbred lines |
| SAM | S-adenosine methionine |
| SHMT | Serine hydroxymethyltransferase |
| THF | Tetrahydrofolate |
| TS | Thymidylate synthase |

| | |
|---|---|
| WGCNA | Weighted gene co-expression network analysis |
| ρ-ABA | Para-aminobenzoate |
| ρ-ABAGlu | ρ-Aminobenzoylglutamate |
| 10-FDF | 10-Formyl-THF deformylase |
| 10-F-THF | 10-Formyl-tetrahydrofolate |
| 5,10-CH=THF | 5,10-Methenyltetrahydrofolate |
| 5,10-$CH_2$-THF | 5,10-Methylenetetrahydrofolate |
| 5-FCL | 5-Formyl-THF cycloligase |
| 5-F-THF | 5-Formyl-tetrahydrofolate |
| 5-M-THF | 5-Methyl-tetrahydrofolate |

# 1

# General introduction

Investigating the genetic basis and regulatory mechanism of folate metabolism in maize

# 1 Folate biosynthesis in plants and bacteria

Folates are a class of water-soluble B-vitamins, including tetrahydrofolate (THF) and its derivatives that play an essential role as one-carbon donors and acceptors. They are involved in DNA biosynthesis (purines and thymidylate), amino acid metabolism (methionine, glycine, and serine), pantothenate, formylmethionyl–transfer RNA and methylation in organisms (Hanson and Roje 2001, Gorelova, Ambach et al. 2017). The folate modules consist of pterin, para-aminobenzoate (ρ-ABA) and glutamate tail **(Figure 1-1)**. Mammals cannot synthesize folate *de novo* due to the lack of essential folate genes. The folate biosynthesis pathway in plants and bacteria has been well characterized where folates are synthesized *de novo* via the same pathway with minor changes (de Crecy-Lagard, El Yacoubi et al. 2007, Hanson and Gregory 2011). In general, pterin and ρ-ABA molecules are synthesized in cytosol and plastids, respectively, and then assembled into polyglutamylated THF in mitochondria. Polyglutamylated THF can be hydrolyzed to monoglutamylated THF in plant vacuoles **(Figure 1-2)**.

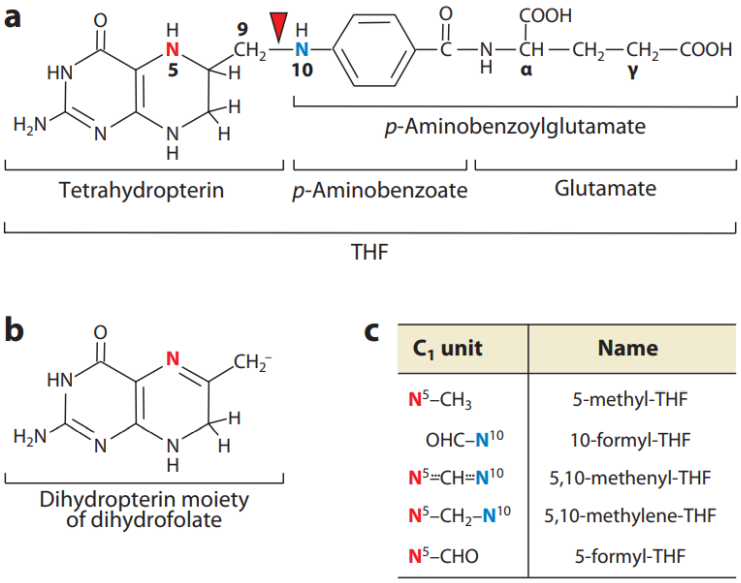



**Figure 1-1.** Structure of folate (a) Chemical structure of tetrahydrofolate (THF), monoglutamyl form. The red arrowhead marks the oxidatively labile C9–N10 bond. A polyglutamyl tail can be attached via the γ-carboxyl group of the glutamate moiety. (b) The 7,8-dihydropterin moiety of dihydrofolate. (c) The various one-carbon (C1) substituents of THF (Hanson and Gregory 2011)

## *1.1 Pterin synthesis*

The reaction of GTP to dihydroneopterin triphosphate by GTP cyclohydrolase I (GCHI) is the first step and rate-limiting step controlling the pterin biosynthesis in the cytosol (Basset, Quinlivan et al. 2002) **(Figure 1-2)**. The plant GCHI mRNAs were isolated from developing seeds, leaves, and roots demonstrating that *de novo* pterin synthesis occurs throughout the whole plant (McIntosh, Brushett et al. 2008, McIntosh and Henry 2008). GCHI mRNA and protein have strong expression in unripe tomato (*Solanum lycopersicum*) fruit, indicating that fruit purine and folate are made in situ other than imported from leaves (Basset, Quinlivan et al. 2002). In *Escherichia coli* (*E. coli*), GCHI is encoded by gene *folE*. The *E. coli* GCHI exists in the form of homodecamers, and contains one zinc-iron from each subunit. Enzyme activity assay, structure and critical binding sites of EcGCHI have been well studied. (Schoedon, Redweik et al. 1992, Rebelo, Auerbach et al. 2003).

The dihydroneopterin aldolase (DHNA) is a cytosolic protein involved in the last step of pterin synthesis **(Figure 1-2)**. It utilizes dihydroneopterin triphosphate as the substrate to produce dihydroneopterin (DHN), and then DHN is cleaved to 6-hydroxymethyldihydropterin (HMDHP). The gene encoding DHNA in *E. coli* has been cloned, characterized, and designated *FolB* (Haussmann, Rohdich et al. 1998). Comparative genomic analysis shows that the *FolB* gene is missing in diverse bacteria, heterokonts (diatoms and oomycetes), and malaria parasites (Pribat, Jeanguenin et al. 2009). In Arabidopsis, there are three genes (*AtFolB1*, *AtFolB2* and *AtFolB3*) with high similarity to the bacterial *DHNA*. *AtFolB1* and *AtFolB2* are transcriptionally expressed at a low level in the whole plant, and *AtFolB3* is only poorly expressed in siliques (Goyer, Illarionova et al. 2004).

## *1.2 ρ-ABA synthesis*

The ρ-ABA synthesis is initiated with the reaction of chorismate to aminodeoxychorismate (ADC) catalyzed by aminodeoxychorismate synthase (ADCS), with glutamine being the amino donor (Basset, Quinlivan et al. 2004) **(Figure 1-2)**. In bacteria, *PabA* encoding glutamine amidotransferase (GAT) and *PabB* encoding aminodeoxychorismate synthase (ADCS) are responsible for this step. The protein GAT works in tandem with ADCS forming a heterodimeric complex (Roux and Walsh 1992). The PabA protein from *E. coli* alone has no GAT activity, and no glutaminase activity is detected until the presence of PabB subunit (Ye, Liu et al. 1990, Roux and Walsh 1992). In plants, ADC is synthesized by a single protein, with two key domains, GAT domain in N-terminal site and ADCS domain in C-terminal site. The $NH_3$ released from the glutamine by the GAT domain is channelled toward the ADCS domain to participate in the reaction (Camara, Richefeu-Contesto et al. 2011).

Subsequently, ADC is converted to ρ-ABA by aminodeoxychorismate lyase (ADCL) in the plastids. In bacteria, ADCL is encoded by *PabC*, and the purified PabC is a monomeric protein that catalyzes the reaction with pyridoxal phosphate as a cofactor (Green and Nichols 1991, Green, Merkel et al. 1992) **(Figure 1-2)**. In contrast, ADCL protein has been characterized as a homodimeric enzyme and cloned from the Arabidopsis (Basset, Ravanel et al. 2004). It is particularly noteworthy that the

expression levels of *ADCS* and *ADCL* fall sharply during tomato fruit ripening. A similar decrease in transcript levels is observed in tomato *GCHI*, indicating that the decline of gene expression in the pterin synthesis pathway (*GCHI*) and ρ-ABA synthesis pathway (*ADCS* and *ADCL*) may be a general feature of folate synthesis during tomato fruit ripening (Basset, Ravanel et al. 2004).

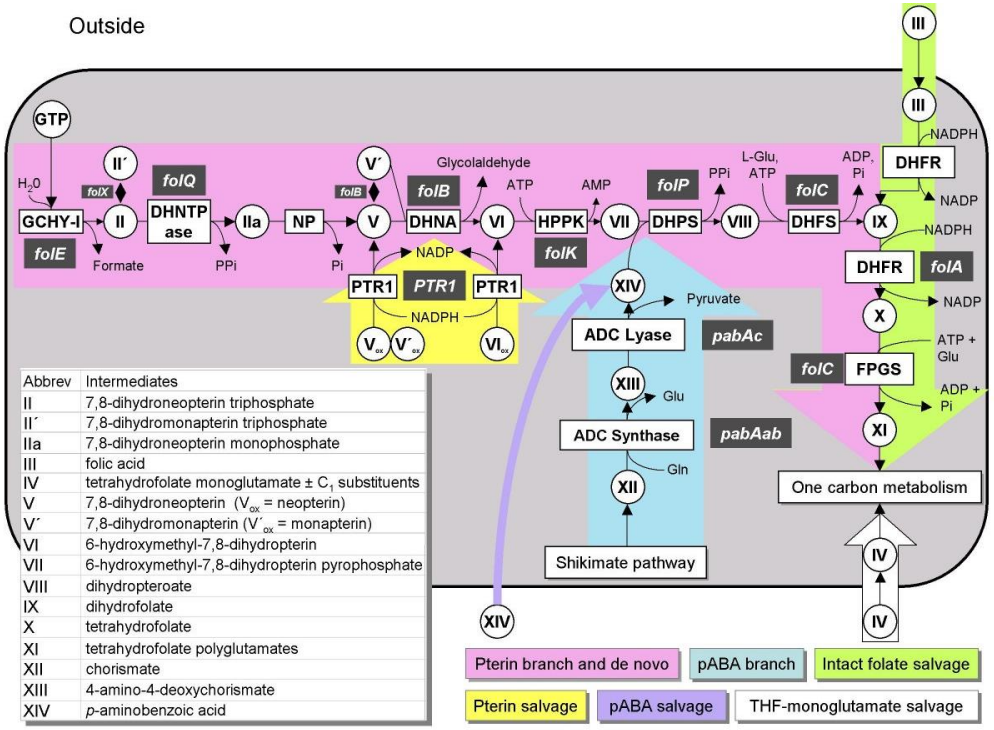
## *1.3 Folate assembly, polyglutamylation, and deglutamylation*

The HMDHP and ρ-ABA are assembled to THF by hydroxymethyldihydropterin pyrophosphokinase (HPPK) and dihydropteroate synthase (DHPS) activity in the mitochondria **(Figure 1-2)**. In most bacteria, HPPK and DHPS are monofunctional enzymes. HPPK activates HMDHP, which is subsequently combined into dihydropteroate (DHP) with ρ-ABA by DHPS (Dallas, Gowen et al. 1992). The three-dimensional structures of HPPK and DHPS from *E. coli* show that HPPK monomer consists of a single globular α/β domain. Such results help us to understand how substrates and inhibitors bind with the catalytic site of the protein (Achari, Somers et al. 1997, Mouillon, Ravanel et al. 2002). In contrast to bacteria, plant HPPK, a part of bifunctional (DHNA-HPPK or HPPK-DHPS) or trifunctional (DHNA-HPPK-DHPS) enzymes (Rebeille, Macherel et al. 1997, Yan and Ji 2011), is encoded by a polycistronic gene. Aside from the mitochondrial HPPK-DHPS, cytosolic HPPK-DHPS is expressed in developing seeds and salt-stressed seedlings in Arabidopsis (Navarrete, Van Daele et al. 2012).

Next, DHP is coupled to glutamate to yield dihydrofolate (DHF) by dihydrofolate synthase (DHFS), and then dihydrofolate reductase (DHFR) converts the DHF to THF **(Figure 1-2)** (Bognar, Osborne et al. 1985, Luo, Piffanelli et al. 1993, Ravanel, Cherest et al. 2001). DHFS is a monofunctional enzyme in plants but coupled with folylpolyglutamate synthetase (FPGS) activity in bacteria. Plant DHFR is assembled with thymidylate synthase (TS) as a bifunctional enzyme DRTS (Neuburger, Rebeille et al. 1996, Cox, Robertson et al. 1999) while existing as a monofunctional enzyme in bacteria. Studies with the DHFR-deficient strains demonstrate that DHFR is not essential for the production of THF in *E. coli* (Hamm-Alvarez, Sancar et al. 1990).

Eventually, THF can be linked with polyglutamate tails via FPGS to form polyglutamated THF **(Figure 1-2)**. Arabidopsis has three FPGSs, which are targeted to the cytosol, mitochondria and plastids, respectively. This is consistent with the presence of polyglutamated THF in those subcellular compartments (Ravanel, Cherest et al. 2001). The monoglutamated folates are initially synthesized in the mitochondria and then transported to other subcellular organelles (Ravanel, Cherest et al. 2001). Vacuoles contain polyglutamates which are not from the FPGS reaction (Ravanel, Cherest et al. 2001, Orsomando, de la Garza et al. 2005), but presumably due to the polyglutamate THF transportation (Hanson and Gregory 2011). The *GLA1* encodes a protein with homology to DHFS and FPGS in Arabidopsis, and has been confirmed to affect the folate contents and plays an essential role in embryo development (Ishikawa, Machida et al. 2003). In *E. coli*, *FolC* catalyzes the addition of L-glutamate to DHP (DHFS activity) and THF (FPGS activity). The crystal structure of EcFolC consists of a different DHP binding site compared with the folate site identified in the

FPGS (Mathieu, Debousker et al. 2005). Unlike the bacteria, the polyglutamate THF are continuously imported into the vacuoles and hydrolyzed to monoglutamate THF by γ-glutamyl hydrolase (GGH) in plants (Orsomando, de la Garza et al. 2005, Akhtar, McQuinn et al. 2008, Akhtar, Orsomando et al. 2010). There are three GGHs in Arabidopsis. *AtGGH1* and *AtGGH2* are expressed throughout plants, with *AtGGH1* predominating in root and *AtGGH2* in leaves. The expression of *AtGGH3* is very low in all tissues analyzed, including roots, siliques, stems and leaves (Akhtar, Orsomando et al. 2010). To date, the factors that impact the GGH activity are still unknown. Overexpressing *GGH* in vacuoles reduces the length of folate polyglutamate tail in Arabidopsis leaves, demonstrating that this enzyme plays a role in governing polyglutamate tail length and folate stability *in vivo* (Akhtar, Orsomando et al. 2010). In general, the genes involved in folate synthesis have been characterized in *E. coli* and *Arabidopsis thaliana* (Green and Nichols 1991, Roux and Walsh 1992, Schoedon, Redweik et al. 1992, Basset, Quinlivan et al. 2002, Basset, Ravanel et al. 2004, Sahr, Ravanel et al. 2006, Gorelova, Ambach et al. 2017).




| Abbrev | Intermediates |
|---|---|
| II | 7,8-dihydroneopterin triphosphate |
| II´ | 7,8-dihydromonapterin triphosphate |
| IIa | 7,8-dihydroneopterin monophosphate |
| III | folic acid |
| IV | tetrahydrofolate monoglutamate ± $C_1$ substituents |
| V | 7,8-dihydroneopterin ($V_{ox}$ = neopterin) |
| V´ | 7,8-dihydromonapterin ($V´_{ox}$ = monapterin) |
| VI | 6-hydroxymethyl-7,8-dihydropterin |
| VII | 6-hydroxymethyl-7,8-dihydropterin pyrophosphate |
| VIII | dihydropteroate |
| IX | dihydrofolate |
| X | tetrahydrofolate |
| XI | tetrahydrofolate polyglutamates |
| XII | chorismate |
| XIII | 4-amino-4-deoxychorismate |
| XIV | *p*-aminobenzoic acid |

**Figure 1-2.** Folate synthesis and salvage pathways. Gene names are white-on-gray; all except *folQ* and *PTR1* are from *E. coli*. The *folQ* gene has been identified only in *Lactococcus lactis* and plants, and *PTR1* only in *Leishmania* and other trypanosomatids. Note that DHN aldolase also mediates epimerization of DHN to 7,8-dihydromonapterin and aldol cleavage of 7,8-dihydromonapterin. ADC, aminodeoxychorismate; DHFR, dihydrofolate reductase; DHFS, dihydrofolate synthase; DHNA, dihydroneopterin aldolase;

DHNTPase, dihydroneopterin triphosphate pyrophosphatase; DHPS, dihydropteroate synthase; FPGS, folylpolyglutamyl synthase; HPPK, hydroxymethyldihydropterin pyrophosphokinase; NP, nonspecific phosphatase; PTR1, pteridine reductase 1; the subscript ox denotes the fully oxidized forms of pterins. (de Crecy-Lagard, El Yacoubi et al. 2007)

## *1.4 Folate salvage*

Plants and some bacteria can recycle the pterin and ρ-ABA-Glu cleavage products back to folates. There are three folate-related salvage pathways **(Figure 1-2)**: (a) ρ-ABA salvage: Specifically in plants' vacuole, polyglutamate THF undergoes *in vivo* oxidative cleavage and breaks down to pterin and ρ-aminobenzoylglutamate (ρ-ABAGlu) or its polyglutamyl forms (ρ-ABAGlu$_n$) by GGH (Akhtar, McQuinn et al. 2008). Then ρ-ABAGlu can be further hydrolyzed to ρ-ABA and glutamate, which is recycled by ρ-ABAGlu hydrolase for DHF synthesis (Orsomando, Bozzo et al. 2006, Bozzo, Basset et al. 2008). (b) Pterin salvage: NADPH-dependent pterin aldehyde reductase (PTAR) can also reduce dihydropterin-6-aldehyde and pterin-6-aldehyde to form HMDHP in pea leaves and Arabidopsis seed (Orsomando, de la Garza et al. 2005). (c) Intact folate salvage: In bacteria, thymidylate synthase (TS) catalyzes folate derivatives to produce DHF. DHFR is required to continually utilize such DHF to synthesize THF (Cox, Robertson et al. 1999). In *E. coli*, DHFR and TS are encoded by separate genes *FolA* and *ThyA*. This is different from that in plants, where DHFR and TS are assembled as a bifunctional enzyme DRTS (Neuburger, Rebeille et al. 1996, Cox, Robertson et al. 1999). In Arabidopsis, the structural and evolutionary analysis of two bifunctional DRTS enzymes indicates that the DHFR and TS domains have evolved at different rates (Lazar, Zhang et al. 1993).

## *1.5 Folate regulation in plants and bacteria*

There is no evidence on the regulation of ρ-ABA synthesis during folate biosynthesis in bacteria. The absence of feedback inhibition by ρ-ABA or end products of folate metabolism has been reported for plants and bacterial ADCS enzymes (Viswanathan, Green et al. 1995, Basset, Quinlivan et al. 2004) **(Figure 1-3)**. ADCS enzyme activity is strongly feedback inhibited by DHP and DHF *in vitro* (Mouillon, Ravanel et al. 2002). However, it is unlikely that this effect will happen *in vivo* because DHP and DHF pools are mainly produced and stored in mitochondrial, while ADCS is in the plastid (Sahr, Ravanel et al. 2006). Meanwhile, the *ADCS* is coregulated with *ADCL* and seems not to be feedback inhibited by physiological concentration of ρ-ABA, THF, 5-M-THF and 5-F-THF in higher plants (Basset, Ravanel et al. 2004). The inhibition of GCHI activity by tetrahydrobiopterin was observed in *E. coli* (Schoedon, Redweik et al. 1992). However, this type of feedback-inhibitory mechanism is absent in plants due to a lack of tetrahydrobiopterin. The folate enhancement in tomatoes provides evidence for feedforward control of folate biosynthesis at transcriptional level **(Figure 1-3)**. The expression level of *DHNA* can be induced by pterin accumulation, *ADCL1* by ADC, and *mFPGS* possibly by THF (Waller, Akhtar et al. 2010). For bifunctional HPPK-DHPS protein, HPPK activity in plants limits the overall reaction and is highly dependent on DHPS activity. The rate

of dihydropterin $PP_i$ formation is dependent on its utilization by DHPS. HPPK activity is strongly feedback-inhibited by dihydropterin$PP_i$. Meanwhile, the DHPS domain is feedback inhibited by DHP, DHF and THF, which postulate that the DHPS domain plays an essential regulatory role in the folate biosynthesis pathway (Mouillon, Ravanel et al. 2002). The regulatory mechanism of polyglutamylated THF biosynthesis is very complicated. During the folate assembly stage, inhibiting DHFR to block folate synthesis in Arabidopsis results in folate depletion and increased transcript level of cytosolic *FPGS*. Such results imply that folate polyglutamates operate a feedback control on *FPGS* gene expression (Loizeau, De Brouwer et al. 2008). Notably, the uridylyltransferase/uridylyl-removing enzyme and tRNA modification enzyme allosterically control the *FolC* to regulate the polyglutamylated THF biosynthesis in *E. coli* (Rodionova, Goodacre et al. 2018). Overexpression of *folKE* (biofunctional gene encoding HPPK and GCHI) in *Lactococcus Lactis* reduces the average polyglutamyl tail length, resulting in enhanced excretion of folate (Sybesma, Van Den Born et al. 2003). As polyglutamyl tail length increases, the ability of folate transporters to shuttle folates across membranes tend to be reduced. The plant FPGS enzyme also contains an adverse effect of a high level of extracellular folate, whereas increased FPGS activity is associated with higher intracellular folate levels (Akhtar, Orsomando et al. 2010). . In addition, the *GGH* also is an indicator of folate contents in potato (*Solanum tuberosum*) tubers (Robinson, Garcia Salinas et al. 2019). The extent of GGH-mediated hydrolysis of vacuolar polyglutamates affects the total cellular folate levels in tomatoes (Akhtar, Orsomando et al. 2010).

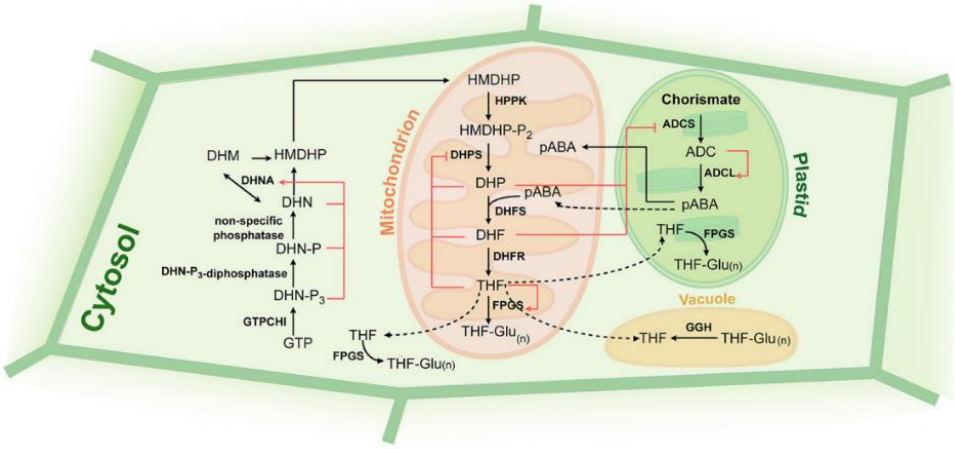



**Figure 1-3.** Folate biosynthesis and regulation in plants. Black arrows indicate biosynthetic steps, red arrows and blunt-end arrows show activation and inhibition, respectively, of enzymatic steps by folate precursors (Gorelova, Ambach et al. 2017)

## *1.6 Other metabolism associated with folate synthesis*

The carbohydrate metabolism regulates the pools of hexoses, pentoses and trioses in the plastid and cytosol. The upstream precursors of folate biosynthesis mostly come from carbohydrate metabolism, which provides erythrose-P and phosphoenolpyruvate for chorismate biosynthesis and provides ribose 5-P for GTP synthesis. (Roje 2007, Asensi-Fabado and Munne-Bosch 2010). GTP is a common precursor of riboflavin and folates, demonstrating that folate synthesis might be associated with riboflavin metabolism (Roje 2007). For the pterin branch, formate and glycolaldehyde can be released by GCHI and DHNA, respectively, which demonstrate that formate metabolism and glycolaldehyde metabolism are involved in folate synthesis (Rebelo, Auerbach et al. 2003, Wang, Li et al. 2007). In the ρ-ABA branch, ADCS converts chorismate and glutamine to ADC and glutamate (Viswanathan, Green et al. 1995). Whereas PabA alone has no glutaminase activity, PabB alone can convert chorismate into ADC in the presence of $NH_3$ (Camara, Richefeu-Contesto et al. 2011). The scientists construct a stable recombinant GAT-ADCS from Arabidopsis to investigate that $NH_3$ perform a role as channeling between these two domains (Camara, Richefeu-Contesto et al. 2011). Meanwhile, pyruvate can be produced via ADCL in *E. coli* indicating that folate synthesis is associated with pyruvate metabolism (Green, Merkel et al. 1992). Additionally, folate metabolism is critical for nitrogen metabolism in Arabidopsis under nitrate-limited conditions (Meng, Xu et al. 2017). Furthermore, Ayala-Rodriguez's research results illustrate that folate connects to plant hormones in Arabidopsis. (Ayala-Rodriguez, Barrera-Ortiz et al. 2017).

# 2 One-carbon metabolism

C1 units such as formyl, methylene and methyl, at various oxidation levels such as can be attached to the $N_5$ and/or $N_{10}$ positions of THF to generate different folate derivatives, such as 5,10-methylenetetrahydrofolate (5,10-$CH_2$-THF), 5,10-methenyltetrahydrofolate (5,10-CH=THF), 10-formyl tetrahydrofolate (10-F-THF), 5-formyl tetrahydrofolate (5-F-THF) and 5-methyl tetrahydrofolate (5-M-THF) (Figure 1-1C). The interconversions occur on different folate derivatives with the synthesis of methionine, S-adenosine methionine (SAM), thymidylate and purines, called one-carbon metabolism (OCM) **(Figure 1-4)**.

## *2.1 OCM in bacteria*

**A. The interconversion between 5,10-$CH_2$-THF and THF:** In bacteria, three reactions are involved in the conversion between THF and 5,10-$CH_2$-THF. First, the enzyme serine hydroxymethyltransferase (SHMT) utilizes the 5,10-$CH_2$-THF and glycine as the substrate to generate the serine and THF. SHMT enzyme from *E.coli*, encoded by gene *GlyA*, shows a similar reaction mechanism and structure with the rabbit liver isoenzymes (Schirch, Hopkins et al. 1985). Functional studies of SHMT reveal that *GlyA* gene plays a key role in lysostaphin resistance (Batool, Ko et al. 2020). In *E. coli*, methionine limitation can cause the depression of SHMT activity and show a high correlation between the ratio of homocysteine to SAM and the rate of SHMT synthesis (Greene and Radovich 1975, Dev and Harvey 1984). Second,

mitochondrial glycine decarboxylase complex (GDC), which is known as the glycine cleavage system (GCS), catalyzes the reaction of THF to 5,10-$CH_2$-THF. GCS/GDC consist of four component proteins: P-protein (encoded by gene *GCV2*), H-protein (encoded by gene *GCV3*), T-protein (encoded by gene *GCV1*) and L-protein (encoded by gene *LPD1*). P-protein catalyzes the initial partial reaction of glycine degradation, and H-protein serves as a co-substrate. T-protein catalyzes the release of ammonia and transfer of the one-carbon unit to THF (Fujiwara and Motokawa 1983) and interacts with H-protein by the N-terminal region. (Fujiwara, Okamura-Ikeda et al. 1984, Okamura-Ikeda, Kameoka et al. 2003). In the glycine cleavage reaction, L-protein catalyzes the reoxidation of the reduced lipoate attached to the H-protein (Faure, Bourguignon et al. 2000, Neuburger, Polidori et al. 2000). The lipoic acid prosthetic group of H-protein interacts with the active site of the above three enzymes (P-, T-, L-proteins) (Okamura-Ikeda, Fujiwara et al. 1999). The reaction mechanism of the GCS has been well-reviewed (Kikuchi, Motokawa et al. 2008). In *Saccharomyces cerevisiae* (yeast), all the GCV genes show increased transcription levels after adding exogenous glycine into the medium, but the level of 5,10-$CH_2$-THF is decreased with excess glycine (Piper, Hong et al. 2000). Third, ThyA protein, which belongs to the TS family, catalyzes the 5,10-$CH_2$-THF to produce DHF, and then DHFR utilizes such DHF to synthesize THF. ThyX protein also shows TS activity, both ThyA and ThyX catalyze deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP). However, their reductive mechanism is different. ThyX performs the reaction to form THF, while ThyA catalysis results in the formation of DHF (Leduc, Escartin et al. 2007). In most bacteria, SHMT, DHFR and TS constitute the DHFR-Thy/folate cycle (Carreras and Santi 1995).

**B. The interconversion between 5,10-CH=THF and 5-F-THF:** SHMT not only converts 5,10-$CH_2$-THF to THF and generates glycine, but also catalyzes the 5,10-CH=THF to 5-F-THF in the presence of glycine (Stover and Schirch 1990). The *E. coli* mutants lacking SHMT enzyme activity do not produce 5-F-THF (Stover and Schirch 1990). Next, 5-formyl-THF cycloligase (5-FCL) enzyme removes the 5-F-THF. However, some prokaryotes have no gene encoding the 5-FCL protein, and the glutamate formiminotransferase functional replaces 5-FCL in some organisms (Jeanguenin, Lara-Nunez et al. 2010).

**C. The conversion of 5,10-$CH_2$-THF to methionine:** The 5,10-methylenetetrahydrofolate reductase (MTHFR), encoded by *MetF* in *E. coli*, catalyzes the reduction of 5,10-$CH_2$-THF to 5-M-THF using the FAD as the cofactor. The essential folate binding site of MTHFR (Aspartate 120) reveals the relationship between complex formation and catalytic effectiveness (Trimmer, Ballou et al. 2005). Following that, the *MetE* encodes the cobalamin-independent methionine synthase (MS), which utilizes the 5-M-THF to produce methionine under aerobic conditions. While the MetH protein drives this reaction under anaerobic conditions (Mordukhova and Pan 2013).

**D. The conversion of 5,10-$CH_2$-THF to 10-F-THF:** The 5,10-methenyl-THF cyclohydrolase/5,10-methylene-THF dehydrogenase (MTD/MTCH), encoded by gene *FolD*, is a bifunctional protein in most bacteria. The dehydrogenase activity of

FolD catalyzes $NADP^+$-dependent oxidation of 5,10-$CH_2$-THF to 5,10-CH=THF. Then, 5,10-CH=THF is converted to 10-F-THF by the cyclohydrolase activity of FolD (D'Ari and Rabinowitz 1991). In contrast to the *Peptostreptococcus* and *Clostridium formicoaceticum*, FolD has monofunctional dehydrogenase activity (Clark and Ljungdahl 1982, Wohlfarth, Geerligs et al. 1991). The *FolD* deletion strain in *E. coli* is auxotrophic for glycine and purine/formate, and the mutant also shows hypersensitivity to trimethoprim (Sah, Aluri et al. 2015).

**E. The conversion between 10-F-THF and THF:** In addition to *FolD*, 10-F-THF also can be generated from THF, ATP and formate by 10-formyltetrahydrofolate synthetase (FTHS, encoded by *Fhs*) (Paukert and Rabinowitz 1980). Many bacteria have both FolD activity and FTHS-mediated mechanisms to meet the requirement of 10-F-THF (Sah, Aluri et al. 2015). Finally, 10-F-THF can serve as the formyl group donor for the formylation of $Met\text{-}tRNA^{fMet}$ to produce $fMet\text{-}tRNA^{fMet}$ by fMet formyltransferase (encoded by gene *Fmt*). The disruption of *Fmt* results in a significant reduction of growth in *E. coli* (Guillon, Mechulam et al. 1992). The mechanism of OCM regulating the translation initiation in bacteria has been well deciphered (Shetty and Varshney 2021). In general, the metabolites involved in OCM, such as methionine, 5,10-$CH_2$-THF, 5-M-THF and 10-F-THF, have been shown to regulate the rate of protein synthesis and DNA translation (Shetty and Varshney 2021).

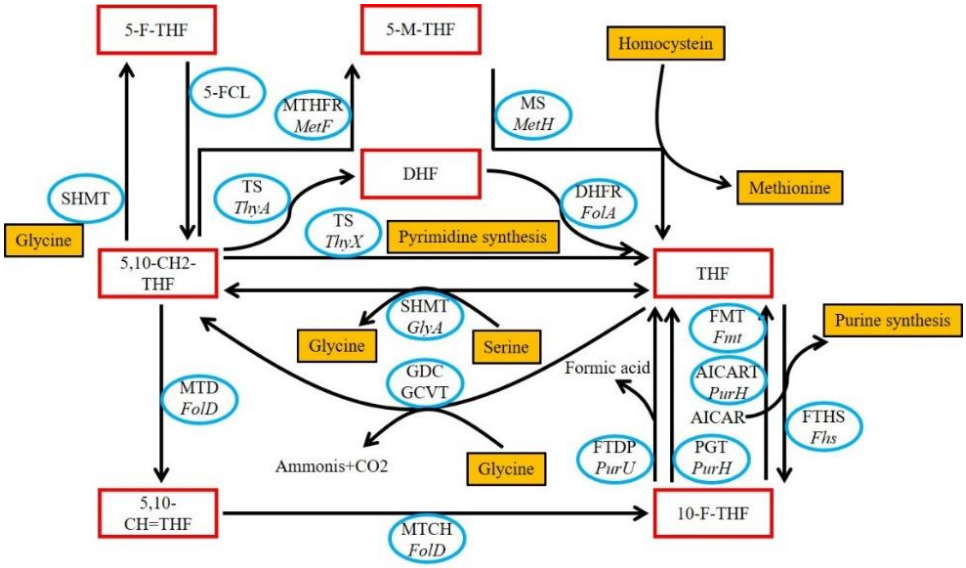



**Figure 1-4.** One-carbon metabolism in bacteria. Red frame: folate derivatives; Blue circle: folate-related genes; Orange frame: major metabolites. 5-M-THF, 5-methyltetrahydrofolate; 5,10-$CH_2$-THF, 5,10-methylenetetrahydrofolate; 5,10-CH=THF, 5,10-methenyltetrahydrofolate; 10-F-THF, 10-formyltetrahydrofolate; AICAR, aminoamidazolecarboxamide ribotide; AICART, aminoimidazolecarboxamide ribotide transformylase; DHFS, dihydrofolate synthase; FTDP, formyltetrahydrofolate deformylase; GCVT, glycine cleavage system aminomethyltransferase; GDC, glycine decarboxylase; MS,

methionine synthase; MTCH, 5,10-methylenetetrahydrofolate cyclohydrolase; MTD, 5,10-methylenetetrahydrofolate dehydrogenase; MTHFR, 5,10-methylenetetrahydrofolate reductase; PGT, phosphoribosyl glycinamidetransformylase; SHMT, serine hydroxymethyltransferase; TS, thymidylate synthase; modified from (Leduc, Escartin et al. 2007).

## *2.2 Other metabolisms associated with OCM in bacteria*

During the process of OCM, 5,10-CH=THF and THF are the major sources for the reversible interconversion of serine and glycine. 5-M-THF is the primary substrate to synthesize methionine. 10-F-THF, which provides the formyl group, contributes to the biosynthesis of purine and formylation of amino acids. In *E. coli*, the *FolD* knock-out strain is auxotrophic for glycine and purine (Sah, Aluri et al. 2015). In *Streptococcus* mutants, the deletion of gene *Fhs* also results in purine auxotrophy (Crowley et al., 1997). In particular, the dysfunction in *Fhs* confers a growth advantage to *E. coli* under hypoxia (Sah et al., 2015). Taken together, the C1 units from OCM used in the biosynthesis of methyl groups, purine and thymidylate demonstrate that amino acid metabolism and biosynthesis of DNA are associated with OCM (Schirch, Hopkins et al. 1985).

In addition, the activity of OCM in bacteria is related to the lengths of polyglutamate tail attached in THF. The OCM depends on glutamate levels and is directly linked to carbon and nitrogen metabolism. Moreover, the inner conversion among different folate derivatives relies on redox regulators like NAD+/NADH, demonstrating that the cellular redox status regulates the activity of OCM in the bacteria (Maynard and Kanarek 2020). Different metabolites from OCM regulate protein synthesis and impact translation (Shetty and Varshney 2021): First, methionine is a crucial amino acid required for aminoacylation of the initiator $tRNA^{fMet}$ and the elongator $tRNA^{Met}$. Both tRNAs participate in the process of translation directly. Second, SAM is vital in ribosome biogenesis and function. Third, 10-F-THF also provides the formyl group donor for the formylation of Met-$tRNA^{fMet}$ in bacteria. Forth, 5,10-$CH_2$-THF is the methyl donor for tRNA modifications. In summary, carbon and nitrogen metabolism, the ratio of NAD+/NADH and the protein translation process are associated with OCM in bacteria (Shetty and Varshney 2021).

## *2.3 OCM in plants*

The OCM reactions in plants are similar to that in bacteria and supply C1 units to synthesize nucleic acid, pantothenate and amino acid. Different folate derivatives are interconverted between different oxidation statuses. The plant OCM has been described in detail (Hanson and Roje 2001). First, THF and 5,10-$CH_2$-THF are interconvertible *via* SHMT and GDC (GCS). Next, the interconversion among 10-F-THF, 5,10-CH=THF, and 5,10-$CH_2$-THF are responsible for the biosynthesis of purine, dTMP, pantothenate and formylMet-Trna. The bifunctional 5,10-methenyl-THF cyclohydrolase/5,10-methylene-THF dehydrogenase (DHC), MTHFR, 10-formyl THF deformylase (10-FDF), FTHS and 5-FCL are involved in these processes **(Figure 1-5)**.

In detail, SHMT is present in cytosol, mitochondria and plastids, while GDC performs the reaction only in mitochondria (Bauwe and Kolukisaoglu 2003). Sustaining research studies have drawn four conclusions as below: First, the GDC and SHMT reactions in mitochondria are tightly connected via a shared pool of 5,10-$CH_2$-THF. Second, almost all the5,10-$CH_2$-THF generated by GDC is the substrate for serine synthesis (Hanson and Roje 2001). Third, the availability of cytosolic THF limits the rate of SHMT reaction (Mouillon, Aubert et al. 1999). Forth, the glycine formed by the cytosolic SHMT reaction is rapidly metabolized through the mitochondrial GDC/SHMT system (Li, Moore et al. 2003). DHC activity is NADP-dependent and present in cytosol, mitochondria, and plastids, whereas the MTHFR activity is NADH-dependent. This difference in coenzyme requirement probably renders the reaction in plants reversible (Roje, Wang et al. 1999). In Arabidopsis, two types of MTHFR catalyze the reversible reaction in the cytosol (Roje, Wang et al. 1999). As one of the critical proteins involved in the methionine biosynthesis cycle, the MS in the higher plant is cobalamin-independent. Three MS genes are identified from Arabidopsis that display high similarity with other land plants (Ravanel, Block et al. 2004). The formate activates the synthesis pathway of 10-F-THF by FTHS in the cytosol, mitochondria, and chloroplasts, respectively (Chen, Chan et al. 1997). 5-FCL, which is highly abundant in leaf mitochondria, catalyzes the conversion of 5-F-THF to 5,10-CH=THF (Roje, Janave et al. 2002).

Adding or removing one-carbon units to perform the primary function in metabolite biosynthesis and regulation is very important to the growth and development of plants. Folate derivatives such as 5,10-$CH_2$-THF, which is the substrate to participate in the reaction of dUMP to dTMP, play an essential role in maintaining genome stability (Neuburger, Rebeille et al. 1996). 10-F-THF is utilized for purine and formyl-methionyl-tRNA synthesis to affect the nucleotide pool maintenance and translation, and 5-M-THF provides its methyl group for methylation involved in the regulation of the genome (Gorelova, Ambach et al. 2017). In Arabidopsis, the folate biosynthesis pathway is inhibited by inhibitor sulfamethazine causing reduced DNA methylation and histone methylation that further shows the release of epigenetic silencing (Zhang, Deng et al. 2012). Due to such a significant role in methylation, several conclusions have also been proved that folate level is associated with epigenetic silencing and auxin signalling, then further affects plants development of embryo, root and hypocotyl (Gorelova, Ambach et al. 2017). The plant embryos are autonomous for folate synthesis, and the *fpgs1/fpgs2* and *dhfs* (*gla1*) knock-out mutants of Arabidopsis have a significant embryo lethal phenotype (Mehrshahi, Gonzalez-Jorge et al. 2010). In Arabidopsis, the loss function of *mFPGS* (*atdfb* mutant) seedlings has a severely shortened primary root, demonstrating the crucial role of folate in root development (Srivastava, Ramos-Parra et al. 2011, Reyes-Hernandez, Srivastava et al. 2014). In addition, *atdfc* causes decreased folate that shows a significantly abnormal phenotype of seedling growth under nitrogen-limited conditions. The shorter length of root, lower nitrate content and decreased abundance of nitrogen storage amino acids compared with wild-type, illustrate that folate is required for nitrogen utilization during seedling development (Jiang, Liu et al. 2013). Indeed, *fpgs1* and *ms1* mutation

in Arabidopsis reduces DNA methylation and releases chromatin silencing, indicating that folate polyglutamylation plays a role in chromatin silencing by maintaining global DNA methylation and histone $H_3K_9$ dimethylation (Zhou, Zhang et al. 2013, Yan, Ma et al. 2019). Meanwhile, a high folate demand in leaves is necessary for photosynthesis, which connects folate metabolism to the light response (Gorelova, Ambach et al. 2017). The *Atshmt1* mutant grows under ambient $CO_2$ that shows a lethal photorespiratory phenotype (Voll, Jamai et al. 2006), suggesting that *SHMT1* plays a conserved function in photorespiration. Furthermore, several analyses provide evidence that folate metabolism can be affected by stress conditions and contribute to balancing the ratio of NADPH/NADP (Fan, Ye et al. 2014, Gorelova, De Lepeleire et al. 2017, Xiang, Hu et al. 2020). The folate metabolism connects to oxidative stress via NADPH production and links with ROS metabolism through homocysteine. Further, overexpression of *OsSHMT3* in Arabidopsis confers tolerance against salt stress suggesting that the folate metabolism might be associated with response to salinity stress (Mishra, Jain et al. 2019). The role of folate in plant development suggests that different folate derivatives have a specific function in many essential processes. A deeper understanding of the regulatory mechanism of folate accumulation in the plant not only provides the acknowledge for folate biofortification but also helps us better understand the plant's growth and development.

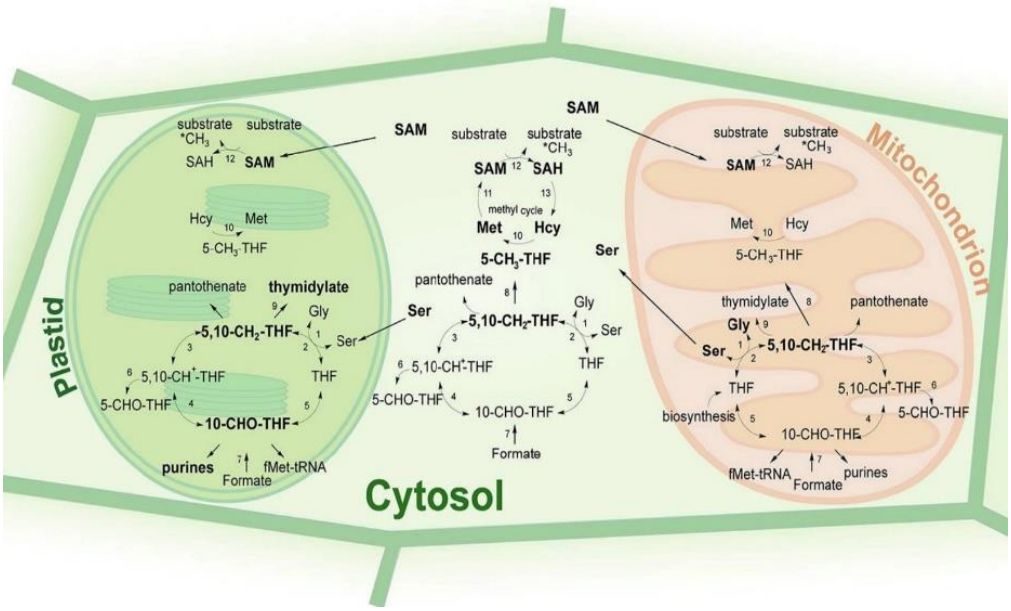



**Figure 1-5.** The reactions of plant C1 metabolism. THF, tetrahydrofolate; 5-CH3-THF, 5-methyltetrahydrofolate; 5,10-$CH_2$-THF, 5,10-methylenetetrahydrofolate; 5,10-CH=THF, 5,10-methenyltetrahydrofolate; 5-CHO-THF, 5-formyltetrahydrofolate; 10-CHO-THF, 10-formyltetrahydrofolate; SAM, S-adenosyl-methionine; SAH, S-adenosyl-homocysteine; Hcy, homocysteine; Met, methionine; Ser, serine; Gly, glycine. 1, serine

hydroxymethyl transferase; 2, glycine decarboxylase complex; 3, 5,10-methylenetetrahydrofolate dehydrogenase; 4, 5,10-methenyltetrahydrofolate cyclohydrolase; 5, 10-formyltetrahydrofolate synthetase; 6, 5-formyltetrahydrofolate cyclohydrolase; 7, 10-formyltetrahydrofolte synthetase; 8, 5,10-methylenetetrahydrofolate reductase; 9, thymidylate synthase; 10, methionine synthase; 11, S-adenosylmethionine synthetase; 12, methyltransferase; 13, SAH hydrolase. Bolded reactions are specific for/ prevalent in a given subcellular compartment (Gorelova, Ambach et al. 2017).

## *2.4 Other metabolism associated with OCM in plants*

The connection between folate metabolism and other metabolites has been investigated to understand the mechanism of folate accumulation in plants. Similar to bacteria, the biosynthesis of amino acid, purine and thymidylate appear tightly connected to OCM in plants. Most vitamins have the same upstream precursors from the pentose and triose pool, which produce chorismate for the biosynthesis of tocochromanols, phylloquinone, and folates, and ribose 5-P for the biosynthesis of thiamine, riboflavin, and folates (DellaPenna 2005, Sahr, Ravanel et al. 2006, Roje 2007, Verberne, Sansuk et al. 2007, Asensi-Fabado and Munne-Bosch 2010). Glutamine and glutamate are involved in the reactions within the specific and common biosynthetic pathways of B vitamins and purine that connect to thiamine, riboflavin and folates (Asensi-Fabado and Munne-Bosch 2010). Besides, the dysfunctions of folate metabolism have impacts on the balance of NADPH/NADP ratio, nitrogen metabolism and photorespiration in plants (Collakova, Goyer et al. 2008, Jiang, Liu et al. 2013, Fan, Ye et al. 2014, Gorelova, De Lepeleire et al. 2017, Meng, Xu et al. 2017, Chen, Zhang et al. 2019, Li, Liang et al. 2021). In addition, foliar salicylic acid treatment improves the accumulation of folates in Arabidopsis with the increased expression of folate binding protein (FBP), evince that folate metabolism also correlates with the salicylic acid via FBP (Puthusseri, Divya et al. 2018). Exogenously supplying plant growth regulators (containing abscisic acid) in callus cultures substantially also increases folate contents (Puthusseri, Divya et al. 2012). The above results reveal that folate metabolism and hormone metabolism are closely connected. Meanwhile, OCM is associated with the synthesis of several amino acids, both directly and indirectly. For example, using serine and homocysteine as the substrate to form glycine, methionine, and cysteine via different reactions (Piper, Hong et al. 2000). Indeed, the small pool of methanol is metabolized via the route: methanol → formaldehyde → formate → $CO_2$, and some of the formates participate in C1 folate pools (Janave, Ramaswamy et al. 1993). On the other side, the lignin metabolism has been suggested to be associated with OCM in Arabidopsis (*Arabidopsis thaliana*), maize (*Zea mays*) and sorghum (*Sorghum bicolor*) (Li, Hill-Skinner et al. 2015, Srivastava, Chen et al. 2015, Adeyanju, Sattler et al. 2021). For instance, in Arabidopsis, the expression of genes in the lignin biosynthetic pathway is downregulated in the loss function of *fpgs1* mutants, next cause reduced lignin contents (Srivastava, Chen et al. 2015). These complexities indicate that many

elements and environmental cues could affect plants' folate accumulation during their development.

## *2.5 OCM in mammal*

In mammals, THF can be converted to 5,10-$CH_2$-THF by SHMT2, and then GCS further catalyzes the product in mitochondria. Next, MTHFD2 and MTHFD2L, two enzymes with bifunctional dehydrogenase/cyclohydrolase activity, catalyze the reaction of 5,10-$CH_2$-THF to 10-F-THF (Tibbetts and Appling 2010, Ducker and Rabinowitz 2017) **(Figure 1-6)**. Subsequently, 10-F-THF is used to form formate by formate-THF ligase enzyme MTHFD1L. In cytosolic, 10-F-THF can be synthesized by a single trifunctional enzyme MTHFD1, then consumed to 5,10-$CH_2$-THF. The latter is used to produce THF by SHMT1 (Pike, Rajendra et al. 2010). Also, 10-F-THF can be converted to THF and $CO_2$ directly in an $NADP^+$-dependent reaction by 10-formyltetrahydrofolate dehydrogenase (FDH, ALDH1L) (Krupenko 2009). 5-M-THF is synthesized by the cytosolic NADPH-dependent activity of MTHFR and consumed by the cobalamin-dependent enzyme methionine synthase (MS, MTR) to form methionine. The 5-F-THF does not play a direct biosynthetic role but serves as a 1C reserve, and methenyltetrahydrofolate synthetase (MTHFS) catalyzes the ATP-dependent conversion of 5-F-THF to 5,10-CH=THF (Jolivet, Dayan et al. 1996).

The regulatory mechanism of OCM in mammal is also complex. Increasing the expression level of the DHFR promoter region can regulate the DHFR transcription, which is mediated through transcription factor Sp1 and E2F in the mouse (Farnham and Schimke 1985, Slansky and Farnham 1996). DHFR protein can be autoregulated by binding to DHFR mRNA and inhibits its own translation (Ercikan-Abali, Banerjee et al. 1997). The *TS* gene (named *TYMS* in mammals) is also controlled by transcriptional factors such as Late SV40 Factor, Forkhead Box M1 and E2F (Powell, Rudge et al. 2000). MTHFD1 is localized in the cytosol as a single trifunctional protein. MTHFD1L structurally resembles MTHFD1, it retains the size of its trifunctional precursor, but both the dehydrogenase and cyclohydrolase domains lack catalytic activity. The MTHFD1L is a monofunctional formyltetrahydrofolate synthetase (Christensen, Patel et al. 2005, Christensen and Mackenzie 2008). The promoter of MTHFD1 is bound by the transcription factor FOXM1, though the regulatory mechanism of MTHFD1 gene expression is still unknown (Grant, Brooks et al. 2013). MTHFD2 expression can be stimulated by the mechanistic target of rapamycin complex 1 (mTORC1) via transcription factor 4 in the mammalian cell, which proves the mTORC1 can control the mitochondrial THF cycle to induce purine synthesis (Ben-Sahra, Hoxhaj et al. 2016). Meanwhile, MTHFD performs the function of balancing the NADP+/NADPH ratio (Fan, Ye et al. 2014). Both SHMT and MTHFS are the critical enzyme of OCM that are inhibited by 5-F-THF (Girgis, Suh et al. 1997). The cytosolic FDH (ALDH1L1) and mitochondria FDH (ALDH1L2) are abundant folate enzymes. Both enzymes perform an essential role in converting folate-bound formate to $CO_2$. This conversion process is proposed to be a likely source of $CO_2$ production from 10-F-THF in the distribution of OCM between the cytosolic and mitochondrial compartments of the cell (Krupenko, Dubard et al. 2010). The SNPs in

ALDH1L1 are associated with serine to glycine ratio in serum (Dharuri, Henneman et al. 2013). The ALDH1L2 provides the reduced NADPH in mitochondria and impacts the ratio of GSH/GSSG (Ducker, Chen et al. 2016). Knocking out the *MTHFR* in murine, which has reduced MTHFR enzyme activity, results in homocystinuria (Schwahn, Laryea et al. 2004). Such similar phenotype is observed in humans as well (Goyette, Frosst et al. 1995). The methionine, formed by the MTR enzyme, is the substrate for SAM synthesis. The MTR is a vitamin B12-dependent enzyme that utilizes a methyl group from 5-M-THF to catalyze the homocysteine to methionine. The methionine concentration is associated with the ratio of SAM to SAH, which affect many methylation processes (Mentch, Mehrmohamadi et al. 2015). For example, histone methylation is catalyzed by histone methyltransferase, which transfers methyl groups from the substrate SAM. Folate deficiency is associated with the increased methylated lysine 4 of histone 3 in mice, suggesting that OCM is tightly connected to histone methylation in mammals (Garcia, Luka et al. 2016). On the other side, the regulation of MTHFR reaction is crucial for deciding whether DNA is to be synthesized by TS or methyl groups are to be supplied by MTR. SAM acts as a regulator of MTHFR by inhibiting its activity (Kutzbach and Stokstad 1971). In the case of folate insufficiency, the depression of SAM production eliminates the inhibition of MTHFR, which results in a decline in nucleotide synthesis (Kutzbach and Stokstad 1971). Moreover, arylamine N-acetyltransferases (NATs), which catalyze the acetylation of aromatic amine and hydrazines, have been identified to contribute to the regulation of folate metabolism in mice and humans. (Cao, Chau et al. 2005, Cao, Strnatka et al. 2010).

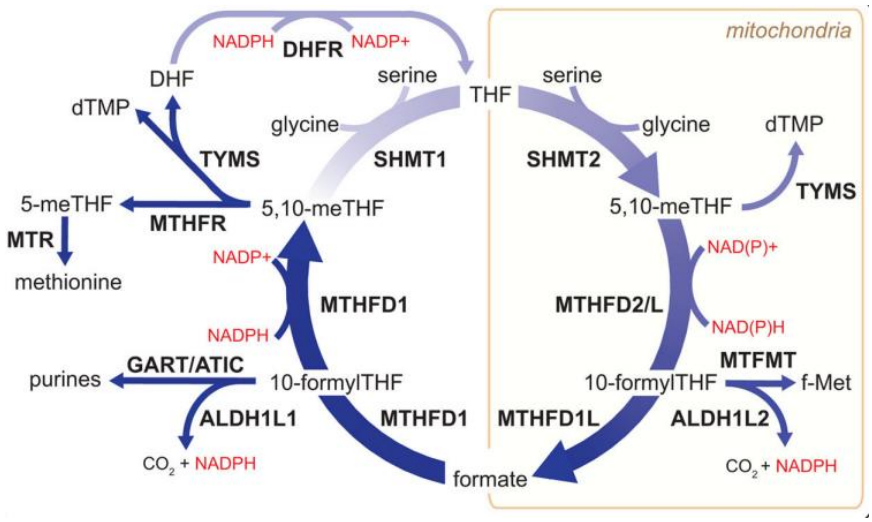



**Figure 1-6.** Compartmentation of folate-mediated one-carbon metabolism in mammal. SHMT1/2, serine hydroxymethyl transferase, cytosolic(1)/ mitochondrial (2); MTHFD1, methylenetetrahydrofolate dehydrogenase, cyclohydrolase, and formyltetrahydrofolate

synthetase 1; MTHFD2/L, methylenetetrahydrofolate dehydrogenase 2/ 2-like; MTHFD1L, monofunctional tetrahydrofolate synthase, mitochondrial; MTFMT, mitochondrial methionyl-tRNA formyltransferase; TYMS, thymidylate synthetase; MTHFR, methylenetetrahydrofolate reductase; MTR, methionine synthase; DHFR, dihydrofolate reductase; GART, phosphoribosylglycinamide formyltransferase; ATIC, 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase; ALDH1L1/2, cytosolic (1)/mitochondrial (2) 10-formyltetrahydrofolate dehydrogenase. (Ducker and Rabinowitz 2017)

## *2.6 Other metabolisms associated with OCM in mammal*

The amino acid metabolisms, formate metabolism and DNA synthesis are associated with OCM in mammals. SHMT can convert glycine into serine (Ducker, Chen et al. 2016), and also plays the function of breaking threonine into glycine via a retroaldol mechanism *in vitro* (Schirch and Szebenyi 2005). Glycine is a required precursor for many biosynthetic pathways, including glutathione, purine, creatine, and heme synthesis (Jain, Nilsson et al. 2012, di Salvo, Contestabile et al. 2013, Amelio, Cutruzzola et al. 2014). In mice, overexpressing *SHMT1* results in decreased SAM level (MacFarlane, Liu et al. 2008). SHMT2is involved in tRNA methylation, which is required for mitochondrial translation and oxidative phosphorylation (Morscher, Ducker et al. 2018). The loss function of SHMT2 in MYC-dependent cells leads to a lower NADPH/NADP+ ratio, demonstrating that the mitochondrial protein SHMT2 is essential for regulating the balance between NADPH production and redox balance (Ye, Fan et al. 2014). In addition, 5-M-THF is a crucial substrate involved in methionine synthesis *via* MS, and then methionine can be utilized for SAM biosynthesis. Active 5-MTHF acts as a methyl donor also participate in transmethylation reaction, regulation of homocysteine and biosynthesis of DNA/RNA, and these biological processes are essential for embryonic and fetal stages during body growth (Ferrazzi, Tiso et al. 2020). 10-F-THF, required for de novo synthesis of purines, can be hydrolyzed to formate and oxidized to $CO_2$.

Furthermore, many research reports have demonstrated that OCM is associated with choline and lipid metabolism in the mammal. The mice fed a choline- and folate-deficient diet for three months show significantly increased hepatic lipid accumulation and reduced expression levels of gene *MS*, *MTHFD1* and *MTHFR* (Pogribny, Kutanzi et al. 2013). The evidence supporting the connection between OCM and lipid metabolism have been well studied. (da Silva, Kelly et al. 2014). The loss function of G6PD, which involve balancing the ratio of NADPH/NADP, results in high NADP that inhibits DHFR activity, and then further affects the folate biosynthesis. Indeed, G6PD is required to maintain NADP and folate homeostasis (Chen, Zhang et al. 2019). Several experiments in humans and rats indicate that methionine synthase is copper-dependent, and the metabolism of homocysteine and folate is regulated in part by copper nutriture (Tamura, Hong et al. 1999, Tamura and Turnlund 2004). In particular, the formiminotransferase cyclodeaminase in the OCM catalyzes histidine degradation that connects folate metabolism to histidine metabolism (Kohls, Sulea et al. 2000). The cytosolic FDH can regulate the glycine metabolism in mouse liver, and the

decreased THF causing reduced generation of glycine can further impair histidine degradation (Krupenko, Sharma et al. 2019). Besides, the OCM enzyme cytosolic FDH can conjugate to MNSFbeta, a ubiquitously expressed member of the ubiquitin-like family. Double knockdown of MNSFbeta and FDH significantly reduced dexamethasone-induced apoptosis, suggesting that MNSFbeta/FDH complex formation may positively regulate apoptosis in thymocytes (Cooperman and Lopez 2002). The experimental results also prove that low folate intake leads to histidine depletion and further causes anaemia (Cooperman and Lopez 2002).

# 3 Folate deficiency and biofortification

Due to the lack of related functional genes, Humans cannot synthesize folate de novo and must absorb it from their daily diet. The required intake of folate is 400 μg/day for adults and 600 μg/d for a pregnant woman (recommended by FAO/WHO) (Bailey, Stover et al. 2015). However, many people still do not get enough folate due to lacking fresh fruit and vegetables in their diet. Given the importance of folate function in the organism, it is not surprising that folate deficiency is associated with some diseases.

## *3.1 folate deficiency in human*

As mentioned above, folate is necessary for methylation reactions that affect critical processes, especially gene expression and translation. Low natural intake of folates will increase the risk of cancers such as pharyngeal, esophageal and pancreatic cancer (Tio, Andrici et al. 2014, Galeone, Edefonti et al. 2015, Pieroth, Paver et al. 2018, Fu, Zeng et al. 2021). Besides, gene polymorphism has been studied to investigate the relationship between folate and diseases. For instance, MTHFR, which decides whether DNA is to be synthesized or methylated, its polymorphism is associated with an increased risk of stomach, esophageal and breast cancer (Tang, Wang et al. 2014, Zhang, Jia et al. 2017). Not only cancers, but MTHFR polymorphism also contribute to neural-tube defects and congenital heart defects (Wang, Wang et al. 2013, Yadav, Kumar et al. 2015). Nowadays, folate deficiency has become a global public health problem. Folic acid pills as folate supplementation are prohibitively expensive in poor regions of the world, as the pills frequently failed to reach the intended populations. The industrial fortification of folate products is limited in developing countries (Gorelova, Ambach et al. 2017). Hence, it is necessary to enhance the stable folate accumulation in the edible parts of crops by biofortification to alleviate this problem.

## *3.2 Bacteria as folate producers and folate biofortification*

Lactic acid bacteria (LAB), which naturally exist in the human and animal gastrointestinal tract, can synthesize folate as an industrially important folate producer (LeBlanc, Laino et al. 2011). The LAB can potentially enhance folate concentration in the milk. *Strep. thermophilus* strains are critical producers of folates in milk, which leads to yogurts having more than six times increased 5-MTHF content compared to the control after 12h of fermentation (LeBlanc, Laino et al. 2011). Specifically, the

*Lactococcus lactis*, *Streptococcus thermophilus*, *Bifidobacterium catenulatum* and the yeast are well-known due to their folate-producing capacities (Rossi, Amaretti et al. 2011). However, even though many researches support those bacterias' ability to produce folate *in vivo*, the application of lactobacilli as folate-producing probiotics still seems to be limited. For example, the strains *L. plantarum* must be cultured in the absence of ρ-ABA for its producing folate (Wegkamp, van Oorschot et al. 2007).

To elevate and improve folate production, the modulation of the bacteria genome provides technical support for the research of bacterial metabolic engineering. In *L. lactis*, overexpressing the *FolKE* gene (encode *folK* and *folE*) lead to a ten-fold increased extracellular folate and three-fold increased total folate, while overexpression of *folA* results in 50% reduced total folate (Wegkamp, van Oorschot et al. 2007). In addition, overexpressing the genes involved in folate biosynthesis and other genes associated with the biosynthesis pathway of related metabolites can increase folate production. For example, simultaneous overexpression of the ρ-ABA and the folate biosynthesis genes results in high folate accumulation, suggesting that both folate and ρ-ABA biosynthesis pathway is needed to balance carbon flux (Wegkamp, van Oorschot et al. 2007).

## *3.3 Crops as folate producers and folate biofortification*

Folates are mostly enriched in green leafy vegetables, beans and some fruit, which are the dietary sources of folate, while the edible part of most staple crops contains a low level of folate (Blancquaert, De Steur et al. 2014). Maize, rice (*Oryza sativa*) and wheat (*Triticum aestivuml*) are the major produced staple crop globally. Nutritionally, all of them contain a low level of folate, 19 μg per 100g FW in maize, 9 μg of folate per 100g FW in rice and 26 μg per 100g FW in wheat flour. Besides, the other staple crop such as potato and cassava also contain a poor source of folates: 19 μg per 100g FW in cultivated potato and 27 μg of folate per 100g FW in cassava. Those areas that consume these staple crops daily often suffer from a suboptimal folate intake. In contrast to Mung beans and soybean, both kinds of beans contain a high level of folate: 625 μg per 100g FW in mature mung beans and 165 μg per 100g FW in soybean. All above data come from the US Department of Agriculture, Agricultural Research Service (http://www.ars.usda.gov/ba/bhnrc/ndl.) and are reviewed in detail (Blancquaert, De Steur et al. 2014). A total of 21 fresh-cut vegetable and fruit packed products are analyzed for total folate to screen out the high folate species (Fajardo, Alonso-Aperte et al. 2015). Finally, spinach, rocket, watercress, chard and broccoli are found with a high level of folate, which ranges from 70.1μg/100g to 140.9 μg/100g (Fajardo, Alonso-Aperte et al. 2015).

There are two approaches to enhancing the folate in staple crops: conventional breeding and engineering breeding. Conventional breeding usually depends on the variation in folate content between different inbred lines. Different rice and wheat genotypes have 2-fold variational folate content that seems insufficient to enrich them by conventional breeding (Piironen, Edelmann et al. 2008, Riaz, Liang et al. 2019). Several studies have already introduced the strategy of enhancing folate levels in plants by metabolic engineering, including increasing the supply of the two folate

precursors: pterin and ρ-ABA. Single folate gene *E. coli GCHI* has been overexpressed in Arabidopsis, leading to a 1,250-fold increase in pterin but only two-fold increase in folate. This could be due to unknown factors limiting the conversion of pterin to folate, or restricted ρ-ABA pathway (Hossain, Rosenberg et al. 2004). Similarly, *GCHI* from Arabidopsis is introduced into common bean resulting in 150-fold increases in pteridine and three-fold enhancement in folate. However, these transgenic lines contain a general increase in ρ-ABA level, which is not observed in other engineering studies (Ramirez Rivera, Garcia-Salinas et al. 2016). Interestingly, the *ADCS* from Arabidopsis is introduced into rice, causing 49-fold increases in ρ-ABA but significantly reducing folate. It seems that a high level of ρ-ABA inhibits folate biosynthesis (Blancquaert, Storozhenko et al. 2013). Hence, the regulatory mechanism of folate metabolism appears to be different among species. In general, overexpressing Arabidopsis *ADCS*, *HPPK-DHPS*, *DHFS*, *GCHI* and *FPGS* genes individually in rice or potato results in a minor enhancement of folate level (Storozhenko, De Brouwer et al. 2007, Blancquaert, Storozhenko et al. 2013, Dong, Cheng et al. 2014), which indicate that single transgenic approaches using these genes may have insufficient for folate improvement. Indeed, gene *GCHI* and *ADCS* from Arabidopsis are transferred to rice, leading to significantly huge increased folate contents in rice seeds (Storozhenko, De Brouwer et al. 2007). Likewise, overexpression of *Mus musculus GCHI* and Arabidopsis *ADCS* in tomatoes also shows 25-fold enhancement of folate (Diaz de la Garza, Gregory et al. 2007). The above results demonstrate that two-genes strategies likely remain an essential effect on folate accumulation. Notably, the same two-genes transgenic strategy (first generation) applied in the biofortification of Arabidopsis or potato tubers has no significant change in folate levels. Despite the fact that the potato transgenic lines have a significant increase in pterin, there is no expression of the gene *ADCS* (Blancquaert, Storozhenko et al. 2013). In the second generation of two-genes strategy, the researchers used the new promoter to overexpress Arabidopsis *ADCS* and *Mus musculus GCHI* in potatoes. It significantly boosted the contents of pterin and ρ-ABA, but only led to a three-fold increase in folate (Blancquaert, Storozhenko et al. 2013). The failure biofortification cases indicate that biofortification strategies application in staple crop species may not be equally efficient (Gorelova, Ambach et al. 2017). Subsequently, additional introduction of HPPK/DHPS and/or FPGS enables augmentation of folates to satisfactory levels in potato (12-fold) (De Lepeleire, Strobbe et al. 2018). The above findings indicate the complexity of folate biofortification and accumulation in different species. It appears that the two-genes strategy (only overexpressing GCHI and ADCS) can work in some species, such as tomato and rice, to reach sufficient levels of folate improvement. Nevertheless, the three-genes (or multi-genes) approach would be required to boost folate accumulation in other plants like potatoes.

To develop an efficient biofortification strategy, it is necessary to understand the regulation of folate biosynthesis and OCM in the crops. The regulatory effect mostly depends on both enzyme activity and gene expression. The cofactors and transcription factors also play an essential role in regulating folate metabolism. Until now, these

regulatory mechanisms of crops are limited known, especially in maize. Maize (*Zea mays* ssp. mays L.) is one of the stable global crops, not only as a food, feed and source of diverse industrial products but also as a model system plant for general plant biology studies. The kernel of maize contains essential chemical components necessary for seed germination and seedling growth, and also produces the essential nutrients for human food and animal feed. To date, only a few studies on folate metabolism genes in maize are available. In our previous studies, genome-wide identification and transcriptional analysis of folate metabolism-related genes in maize kernel have been performed to identify the folate-related genes in maize (Lian, Guo et al. 2015). Previous results demonstrate that 5-M-THF is the major folate derivatives, and the regulatory mechanism of folate is complicated during kernel formation (Lian, Guo et al. 2015). The *GmGCHI* and *GmADCS* from soybean are co-expressed in maize to boost their folate metabolic flux showing a 4.2-fold increase of folate levels in transgenic maize seed (Liang, Wang et al. 2019). The brown *midrib4* (*bm4*) gene involves the accumulation and composition of lignin in maize is identified. It encodes a functional FPGS that connects the lignin metabolism and folate metabolism in a certain extent (Li, Hill-Skinner et al. 2015). In this study, we are looking for the candidate genes that may regulate folate metabolism in maize to figure out the regulatory mechanism of folate during maize kernel formation.

# References

Achari, A., D. O. Somers, J. N. Champness, P. K. Bryant, J. Rosemond and D. K. Stammers (1997). "Crystal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase." Nat Struct Biol 4(6): 490-497.

Adeyanju, A. O., S. E. Sattler, P. J. Rich, L. A. Rivera-Burgos, X. Xu and G. Ejeta (2021). "Sorghum Brown Midrib19 (Bmr19) Gene Links Lignin Biosynthesis to Folate Metabolism." Genes (Basel) 12(5).

Akhtar, T. A., R. P. McQuinn, V. Naponelli, J. F. Gregory, 3rd, J. J. Giovannoni and A. D. Hanson (2008). "Tomato gamma-glutamylhydrolases: expression, characterization, and evidence for heterodimer formation." Plant Physiol 148(2): 775-785.

Akhtar, T. A., G. Orsomando, P. Mehrshahi, A. Lara-Nunez, M. J. Bennett, J. F. Gregory, 3rd and A. D. Hanson (2010). "A central role for gamma-glutamyl hydrolases in plant folate homeostasis." Plant J 64(2): 256-266.

Amelio, I., F. Cutruzzola, A. Antonov, M. Agostini and G. Melino (2014). "Serine and glycine metabolism in cancer." Trends Biochem Sci 39(4): 191-198.

Asensi-Fabado, M. A. and S. Munne-Bosch (2010). "Vitamins in plants: occurrence, biosynthesis and antioxidant function." Trends Plant Sci 15(10): 582-592.

Ayala-Rodriguez, J. A., S. Barrera-Ortiz, L. F. Ruiz-Herrera and J. Lopez-Bucio (2017). "Folic acid orchestrates root development linking cell elongation with auxin response and acts independently of the TARGET OF RAPAMYCIN signaling in Arabidopsis thaliana." Plant Sci 264: 168-178.

Bailey, L. B., P. J. Stover, H. McNulty, M. F. Fenech, J. F. Gregory, 3rd, J. L. Mills, C. M. Pfeiffer, Z. Fazili, M. Zhang, P. M. Ueland, A. M. Molloy, M. A. Caudill, B. Shane, R. J. Berry, R. L. Bailey, D. B. Hausman, R. Raghavan and D. J. Raiten (2015). "Biomarkers of Nutrition for Development-Folate Review." J Nutr 145(7): 1636S-1680S.

Basset, G., E. P. Quinlivan, M. J. Ziemak, R. Diaz De La Garza, M. Fischer, S. Schiffmann, A. Bacher, J. F. Gregory, 3rd and A. D. Hanson (2002). "Folate synthesis in plants: the first step of the pterin branch is mediated by a unique bimodular GTP cyclohydrolase I." Proc Natl Acad Sci U S A 99(19): 12489-12494.

Basset, G. J., E. P. Quinlivan, S. Ravanel, F. Rebeille, B. P. Nichols, K. Shinozaki, M. Seki, L. C. Adams-Phillips, J. J. Giovannoni, J. F. Gregory, 3rd and A. D. Hanson (2004). "Folate synthesis in plants: the p-aminobenzoate branch is initiated by a bifunctional PabA-PabB protein that is targeted to plastids." Proc Natl Acad Sci U S A 101(6): 1496-1501.

Basset, G. J., S. Ravanel, E. P. Quinlivan, R. White, J. J. Giovannoni, F. Rebeille, B. P. Nichols, K. Shinozaki, M. Seki, J. F. Gregory, 3rd and A. D. Hanson (2004). "Folate synthesis in plants: the last step of the p-aminobenzoate branch is catalyzed by a plastidial aminodeoxychorismate lyase." Plant J 40(4): 453-461.

Batool, N., K. S. Ko, A. K. Chaurasia and K. K. Kim (2020). "Functional Identification of Serine Hydroxymethyltransferase as a Key Gene Involved in Lysostaphin Resistance and Virulence Potential of Staphylococcus aureus Strains." Int J Mol Sci 21(23).

Bauwe, H. and U. Kolukisaoglu (2003). "Genetic manipulation of glycine decarboxylation." J Exp Bot 54(387): 1523-1535.

Ben-Sahra, I., G. Hoxhaj, S. J. H. Ricoult, J. M. Asara and B. D. Manning (2016). "mTORC1 induces purine synthesis through control of the mitochondrial tetrahydrofolate cycle." Science 351(6274): 728-733.

Blancquaert, D., H. De Steur, X. Gellynck and D. Van Der Straeten (2014). "Present and future of folate biofortification of crop plants." J Exp Bot 65(4): 895-906.

Blancquaert, D., S. Storozhenko, J. Van Daele, C. Stove, R. G. Visser, W. Lambert and D. Van Der Straeten (2013). "Enhancing pterin and para-aminobenzoate content is not sufficient to successfully biofortify potato tubers and Arabidopsis thaliana plants with folate." J Exp Bot 64(12): 3899-3909.

Bognar, A. L., C. Osborne, B. Shane, S. C. Singer and R. Ferone (1985). "Folylpoly-gamma-glutamate synthetase-dihydrofolate synthetase. Cloning and high expression of the Escherichia coli folC gene and purification and properties of the gene product." J Biol Chem 260(9): 5625-5630.

Bozzo, G. G., G. J. Basset, V. Naponelli, A. Noiriel, J. F. Gregory, 3rd and A. D. Hanson (2008). "Characterization of the folate salvage enzyme p-aminobenzoylglutamate hydrolase in plants." Phytochemistry 69(1): 29-37.

Camara, D., C. Richefeu-Contesto, B. Gambonnet, R. Dumas and F. Rebeille (2011). "The synthesis of pABA: Coupling between the glutamine amidotransferase and aminodeoxychorismate synthase domains of the bifunctional aminodeoxychorismate synthase from Arabidopsis thaliana." Arch Biochem Biophys 505(1): 83-90.

Carreras, C. W. and D. V. Santi (1995). "The catalytic mechanism and structure of thymidylate synthase." Annu Rev Biochem 64: 721-762.

Chen, L., S. Y. Chan and E. A. Cossins (1997). "Distribution of Folate Derivatives and Enzymes for Synthesis of 10-Formyltetrahydrofolate in Cytosolic and Mitochondrial Fractions of Pea Leaves." Plant Physiol 115(1): 299-309.

Chen, L., Z. Zhang, A. Hoshino, H. D. Zheng, M. Morley, Z. Arany and J. D. Rabinowitz (2019). "NADPH production by the oxidative pentose-phosphate pathway supports folate metabolism." Nature Metabolism 1(3): 404-415.

Christensen, K. E. and R. E. Mackenzie (2008). "Mitochondrial methylenetetrahydrofolate dehydrogenase, methenyltetrahydrofolate cyclohydrolase, and formyltetrahydrofolate synthetases." Vitam Horm 79: 393-410.

Christensen, K. E., H. Patel, U. Kuzmanov, N. R. Mejia and R. E. MacKenzie (2005). "Disruption of the mthfd1 gene reveals a monofunctional 10-

formyltetrahydrofolate synthetase in mammalian mitochondria." J Biol Chem 280(9): 7597-7602.

Clark, J. E. and L. G. Ljungdahl (1982). "Purification and properties of 5,10-methenyltetrahydrofolate cyclohydrolase from Clostridium formicoaceticum." J Biol Chem 257(7): 3833-3836.

Collakova, E., A. Goyer, V. Naponelli, I. Krassovskaya, J. F. Gregory, 3rd, A. D. Hanson and Y. Shachar-Hill (2008). "Arabidopsis 10-formyl tetrahydrofolate deformylases are essential for photorespiration." Plant Cell 20(7): 1818-1832.

Cooperman, J. M. and R. Lopez (2002). "The role of histidine in the anemia of folate deficiency." Exp Biol Med (Maywood) 227(11): 998-1000.

Cox, K., D. Robertson and R. Fites (1999). "Mapping and expression of a bifunctional thymidylate synthase, dihydrofolate reductase gene from maize." Plant Mol Biol 41(6): 733-739.

D'Ari, L. and J. C. Rabinowitz (1991). "Purification, characterization, cloning, and amino acid sequence of the bifunctional enzyme 5,10-methylenetetrahydrofolate dehydrogenase/5,10-methenyltetrahydrofolate cyclohydrolase from Escherichia coli." J Biol Chem 266(35): 23953-23958.

da Silva, R. P., K. B. Kelly, A. Al Rajabi and R. L. Jacobs (2014). "Novel insights on interactions between folate and lipid metabolism." Biofactors 40(3): 277-283.

Dallas, W. S., J. E. Gowen, P. H. Ray, M. J. Cox and I. K. Dev (1992). "Cloning, sequencing, and enhanced expression of the dihydropteroate synthase gene of Escherichia coli MC4100." J Bacteriol 174(18): 5961-5970.

de Crecy-Lagard, V., B. El Yacoubi, R. D. de la Garza, A. Noiriel and A. D. Hanson (2007). "Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations." BMC Genomics 8: 245.

De Lepeleire, J., S. Strobbe, J. Verstraete, D. Blancquaert, L. Ambach, R. G. F. Visser, C. Stove and D. Van Der Straeten (2018). "Folate Biofortification of Potato by Tuber-Specific Expression of Four Folate Biosynthesis Genes." Mol Plant 11(1): 175-188.

DellaPenna, D. (2005). "Progress in the dissection and manipulation of vitamin E synthesis." Trends Plant Sci 10(12): 574-579.

Dev, I. K. and R. J. Harvey (1984). "Role of methionine in the regulation of the synthesis of serine hydroxymethyltransferase in Escherichia coli." J Biol Chem 259(13): 8402-8406.

di Salvo, M. L., R. Contestabile, A. Paiardini and B. Maras (2013). "Glycine consumption and mitochondrial serine hydroxymethyltransferase in cancer cells: the heme connection." Med Hypotheses 80(5): 633-636.

Diaz de la Garza, R. I., J. F. Gregory, 3rd and A. D. Hanson (2007). "Folate biofortification of tomato fruit." Proc Natl Acad Sci U S A 104(10): 4218-4222.

Dong, W., Z. J. Cheng, C. L. Lei, X. L. Wang, J. L. Wang, J. Wang, F. Q. Wu, X. Zhang, X. P. Guo, H. Q. Zhai and J. M. Wan (2014). "Overexpression of folate

biosynthesis genes in rice (Oryza sativa L.) and evaluation of their impact on seed folate content." Plant Foods Hum Nutr 69(4): 379-385.

Ducker, G. S., L. Chen, R. J. Morscher, J. M. Ghergurovich, M. Esposito, X. Teng, Y. Kang and J. D. Rabinowitz (2016). "Reversal of Cytosolic One-Carbon Flux Compensates for Loss of the Mitochondrial Folate Pathway." Cell Metab 24(4): 640-641.

Ducker, G. S. and J. D. Rabinowitz (2017). "One-Carbon Metabolism in Health and Disease." Cell Metab 25(1): 27-42.

Ercikan-Abali, E. A., D. Banerjee, M. C. Waltham, N. Skacel, K. W. Scotto and J. R. Bertino (1997). "Dihydrofolate reductase protein inhibits its own translation by binding to dihydrofolate reductase mRNA sequences within the coding region." Biochemistry 36(40): 12317-12322.

Fajardo, V., E. Alonso-Aperte and G. Varela-Moreiras (2015). "Folate content in fresh-cut vegetable packed products by 96-well microtiter plate microbiological assay." Food Chem 169: 283-288.

Fan, J., J. Ye, J. J. Kamphorst, T. Shlomi, C. B. Thompson and J. D. Rabinowitz (2014). "Quantitative flux analysis reveals folate-dependent NADPH production." Nature 510(7504): 298-302.

Farnham, P. J. and R. T. Schimke (1985). "Transcriptional regulation of mouse dihydrofolate reductase in the cell cycle." J Biol Chem 260(12): 7675-7680.

Faure, M., J. Bourguignon, M. Neuburger, D. MacHerel, L. Sieker, R. Ober, R. Kahn, C. Cohen-Addad and R. Douce (2000). "Interaction between the lipoamide-containing H-protein and the lipoamide dehydrogenase (L-protein) of the glycine decarboxylase multienzyme system 2. Crystal structures of H- and L-proteins." Eur J Biochem 267(10): 2890-2898.

Ferrazzi, E., G. Tiso and D. Di Martino (2020). "Folic acid versus 5- methyl tetrahydrofolate supplementation in pregnancy." Eur J Obstet Gynecol Reprod Biol 253: 312-319.

Fu, H., J. Zeng, C. Liu, Y. Gu, Y. Zou and H. Chang (2021). "Folate Intake and Risk of Pancreatic Cancer: A Systematic Review and Updated Meta-Analysis of Epidemiological Studies." Dig Dis Sci 66(7): 2368-2379.

Fujiwara, K. and Y. Motokawa (1983). "Mechanism of the glycine cleavage reaction. Steady state kinetic studies of the P-protein-catalyzed reaction." J Biol Chem 258(13): 8156-8162.

Fujiwara, K., K. Okamura-Ikeda and Y. Motokawa (1984). "Mechanism of the glycine cleavage reaction. Further characterization of the intermediate attached to H-protein and of the reaction catalyzed by T-protein." J Biol Chem 259(17): 10664-10668.

Galeone, C., V. Edefonti, M. Parpinel, E. Leoncini, K. Matsuo, R. Talamini, A. F. Olshan, J. P. Zevallos, D. M. Winn, V. Jayaprakash, K. Moysich, Z. F. Zhang, H. Morgenstern, F. Levi, C. Bosetti, K. Kelsey, M. McClean, S. Schantz, G. P. Yu, P. Boffetta, Y. C. Lee, M. Hashibe, C. La Vecchia and S. Boccia (2015).

"Folate intake and the risk of oral cavity and pharyngeal cancer: a pooled analysis within the International Head and Neck Cancer Epidemiology Consortium." Int J Cancer 136(4): 904-914.

Garcia, B. A., Z. Luka, L. V. Loukachevitch, N. V. Bhanu and C. Wagner (2016). "Folate deficiency affects histone methylation." Med Hypotheses 88: 63-67.

Gorelova, V., L. Ambach, F. Rebeille, C. Stove and D. Van Der Straeten (2017). "Folates in Plants: Research Advances and Progress in Crop Biofortification." Front Chem 5: 21.

Gorelova, V., J. De Lepeleire, J. Van Daele, D. Pluim, C. Mei, A. Cuypers, O. Leroux, F. Rebeille, J. H. M. Schellens, D. Blancquaert, C. P. Stove and D. Van Der Straeten (2017). "Dihydrofolate Reductase/Thymidylate Synthase Fine-Tunes the Folate Status and Controls Redox Homeostasis in Plants." Plant Cell 29(11): 2831-2853.

Goyer, A., V. Illarionova, S. Roje, M. Fischer, A. Bacher and A. D. Hanson (2004). "Folate biosynthesis in higher plants. cDNA cloning, heterologous expression, and characterization of dihydroneopterin aldolases." Plant Physiol 135(1): 103-111.

Goyette, P., P. Frosst, D. S. Rosenblatt and R. Rozen (1995). "Seven novel mutations in the methylenetetrahydrofolate reductase gene and genotype/phenotype correlations in severe methylenetetrahydrofolate reductase deficiency." Am J Hum Genet 56(5): 1052-1059.

Grant, G. D., L. Brooks, 3rd, X. Zhang, J. M. Mahoney, V. Martyanov, T. A. Wood, G. Sherlock, C. Cheng and M. L. Whitfield (2013). "Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors." Mol Biol Cell 24(23): 3634-3650.

Green, J. M., W. K. Merkel and B. P. Nichols (1992). "Characterization and sequence of Escherichia coli pabC, the gene encoding aminodeoxychorismate lyase, a pyridoxal phosphate-containing enzyme." J Bacteriol 174(16): 5317-5323.

Green, J. M. and B. P. Nichols (1991). "p-Aminobenzoate biosynthesis in Escherichia coli. Purification of aminodeoxychorismate lyase and cloning of pabC." J Biol Chem 266(20): 12971-12975.

Greene, R. C. and C. Radovich (1975). "Role of methionine in the regulation of serine hydroxymethyltransferase in Eschericia coli." J Bacteriol 124(1): 269-278.

Guillon, J. M., Y. Mechulam, J. M. Schmitter, S. Blanquet and G. Fayat (1992). "Disruption of the gene for Met-tRNA(fMet) formyltransferase severely impairs growth of Escherichia coli." J Bacteriol 174(13): 4294-4301.

Hamm-Alvarez, S. F., A. Sancar and K. V. Rajagopalan (1990). "The presence and distribution of reduced folates in Escherichia coli dihydrofolate reductase mutants." J Biol Chem 265(17): 9850-9856.

Hanson, A. D. and J. F. Gregory, 3rd (2011). "Folate biosynthesis, turnover, and transport in plants." Annu Rev Plant Biol 62: 105-125.

Hanson, A. D. and S. Roje (2001). "One-Carbon Metabolism in Higher Plants." Annu Rev Plant Physiol Plant Mol Biol 52: 119-137.

Haussmann, C., F. Rohdich, E. Schmidt, A. Bacher and G. Richter (1998). "Biosynthesis of pteridines in Escherichia coli. Structural and mechanistic similarity of dihydroneopterin-triphosphate epimerase and dihydroneopterin aldolase." J Biol Chem 273(28): 17418-17424.

Hossain, T., I. Rosenberg, J. Selhub, G. Kishore, R. Beachy and K. Schubert (2004). "Enhancement of folates in plants through metabolic engineering." Proc Natl Acad Sci U S A 101(14): 5158-5163.

Ishikawa, T., C. Machida, Y. Yoshioka, H. Kitano and Y. Machida (2003). "The GLOBULAR ARREST1 gene, which is involved in the biosynthesis of folates, is essential for embryogenesis in Arabidopsis thaliana." Plant J 33(2): 235-244.

Jain, M., R. Nilsson, S. Sharma, N. Madhusudhan, T. Kitami, A. L. Souza, R. Kafri, M. W. Kirschner, C. B. Clish and V. K. Mootha (2012). "Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation." Science 336(6084): 1040-1044.

Janave, M. T., N. K. Ramaswamy and P. M. Nair (1993). "Purification and characterization of glyoxylate synthetase from greening potato-tuber chloroplasts." Eur J Biochem 214(3): 889-896.

Jeanguenin, L., A. Lara-Nunez, A. Pribat, M. H. Mageroy, J. F. Gregory, 3rd, K. C. Rice, V. de Crecy-Lagard and A. D. Hanson (2010). "Moonlighting glutamate formiminotransferases can functionally replace 5-formyltetrahydrofolate cycloligase." J Biol Chem 285(53): 41557-41566.

Jiang, L., Y. Liu, H. Sun, Y. Han, J. Li, C. Li, W. Guo, H. Meng, S. Li, Y. Fan and C. Zhang (2013). "The mitochondrial folylpolyglutamate synthetase gene is required for nitrogen utilization during early seedling development in arabidopsis." Plant Physiol 161(2): 971-989.

Jolivet, J., A. Dayan, M. Beauchemin, D. Chahla, A. Mamo and R. Bertrand (1996). "Biochemical and molecular studies of human methenyltetrahydrofolate synthetase." Stem Cells 14(1): 33-40.

Kikuchi, G., Y. Motokawa, T. Yoshida and K. Hiraga (2008). "Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia." Proc Jpn Acad Ser B Phys Biol Sci 84(7): 246-263.

Kohls, D., T. Sulea, E. O. Purisima, R. E. MacKenzie and A. Vrielink (2000). "The crystal structure of the formiminotransferase domain of formiminotransferase-cyclodeaminase: implications for substrate channeling in a bifunctional enzyme." Structure 8(1): 35-46.

Krupenko, N. I., M. E. Dubard, K. C. Strickland, K. M. Moxley, N. V. Oleinik and S. A. Krupenko (2010). "ALDH1L2 is the mitochondrial homolog of 10-formyltetrahydrofolate dehydrogenase." J Biol Chem 285(30): 23056-23063.

Krupenko, N. I., J. Sharma, P. Pediaditakis, B. Fekry, K. L. Helke, X. Du, S. Sumner and S. A. Krupenko (2019). "Cytosolic 10-formyltetrahydrofolate

dehydrogenase regulates glycine metabolism in mouse liver." Sci Rep 9(1): 14937.

Krupenko, S. A. (2009). "FDH: an aldehyde dehydrogenase fusion enzyme in folate metabolism." Chem Biol Interact 178(1-3): 84-93.

Lazar, G., H. Zhang and H. M. Goodman (1993). "The origin of the bifunctional dihydrofolate reductase-thymidylate synthase isogenes of Arabidopsis thaliana." Plant J 3(5): 657-668.

LeBlanc, J. G., J. E. Laino, M. J. del Valle, V. Vannini, D. van Sinderen, M. P. Taranto, G. F. de Valdez, G. S. de Giori and F. Sesma (2011). "B-group vitamin production by lactic acid bacteria--current knowledge and potential applications." J Appl Microbiol 111(6): 1297-1309.

Leduc, D., F. Escartin, H. F. Nijhout, M. C. Reed, U. Liebl, S. Skouloubris and H. Myllykallio (2007). "Flavin-dependent thymidylate synthase ThyX activity: implications for the folate cycle in bacteria." J Bacteriol 189(23): 8537-8545.

Li, L., S. Hill-Skinner, S. Liu, D. Beuchle, H. M. Tang, C. T. Yeh, D. Nettleton and P. S. Schnable (2015). "The maize brown midrib4 (bm4) gene encodes a functional folylpolyglutamate synthase." Plant J 81(3): 493-504.

Li, R., M. Moore and J. King (2003). "Investigating the regulation of one-carbon metabolism in Arabidopsis thaliana." Plant Cell Physiol 44(3): 233-241.

Li, W., Q. Liang, R. C. Mishra, R. Sanchez-Munoz, H. Wang, X. Chen, D. Van Der Straeten, C. Zhang and Y. Xiao (2021). "The 5-formyl-tetrahydrofolate proteome links folates with C/N metabolism and reveals feedback regulation of folate biosynthesis." Plant Cell.

Lian, T., W. Guo, M. Chen, J. Li, Q. Liang, F. Liu, H. Meng, B. Xu, J. Chen, C. Zhang and L. Jiang (2015). "Genome-wide identification and transcriptional analysis of folate metabolism-related genes in maize kernels." BMC Plant Biol 15: 204.

Liang, Q., K. Wang, X. Liu, B. Riaz, L. Jiang, X. Wan, X. Ye and C. Zhang (2019). "Improved folate accumulation in genetically modified maize and wheat." J Exp Bot 70(5): 1539-1551.

Loizeau, K., V. De Brouwer, B. Gambonnet, A. Yu, J. P. Renou, D. Van Der Straeten, W. E. Lambert, F. Rebeille and S. Ravanel (2008). "A genome-wide and metabolic analysis determined the adaptive response of Arabidopsis cells to folate depletion induced by methotrexate." Plant Physiol 148(4): 2083-2095.

Luo, M., P. Piffanelli, L. Rastelli and R. Cella (1993). "Molecular cloning and analysis of a cDNA coding for the bifunctional dihydrofolate reductase-thymidylate synthase of Daucus carota." Plant Mol Biol 22(3): 427-435.

MacFarlane, A. J., X. Liu, C. A. Perry, P. Flodby, R. H. Allen, S. P. Stabler and P. J. Stover (2008). "Cytoplasmic serine hydroxymethyltransferase regulates the metabolic partitioning of methylenetetrahydrofolate but is not essential in mice." J Biol Chem 283(38): 25846-25853.

Mathieu, M., G. Debousker, S. Vincent, F. Viviani, N. Bamas-Jacques and V. Mikol (2005). "Escherichia coli FolC structure reveals an unexpected dihydrofolate

binding site providing an attractive target for anti-microbial therapy." J Biol Chem 280(19): 18916-18922.

Maynard, A. G. and N. Kanarek (2020). "NADH Ties One-Carbon Metabolism to Cellular Respiration." Cell Metab 31(4): 660-662.

McIntosh, S. R., D. Brushett and R. J. Henry (2008). "GTP cyclohydrolase 1 expression and folate accumulation in the developing wheat seed." J. Cereal Sci. 48: 503-512.

McIntosh, S. R. and R. J. Henry (2008). "Genes of folate biosynthesis in wheat." J. Cereal Sci. 48: 632-638.

Mehrshahi, P., S. Gonzalez-Jorge, T. A. Akhtar, J. L. Ward, A. Santoyo-Castelazo, S. E. Marcus, A. Lara-Nunez, S. Ravanel, N. D. Hawkins, M. H. Beale, D. A. Barrett, J. P. Knox, J. F. Gregory, 3rd, A. D. Hanson, M. J. Bennett and D. Dellapenna (2010). "Functional analysis of folate polyglutamylation and its essential role in plant metabolism and development." Plant J 64(2): 267-279.

Meng, H., B. Xu, C. Zhang and L. Jiang (2017). "Arabidopsis plastidial folylpolyglutamate synthetase is required for nitrogen metabolism under nitrate-limited condition in darkness." Biochem Biophys Res Commun 482(2): 277-281.

Mentch, S. J., M. Mehrmohamadi, L. Huang, X. Liu, D. Gupta, D. Mattocks, P. Gomez Padilla, G. Ables, M. M. Bamman, A. E. Thalacker-Mercer, S. N. Nichenametla and J. W. Locasale (2015). "Histone Methylation Dynamics and Gene Regulation Occur through the Sensing of One-Carbon Metabolism." Cell Metab 22(5): 861-873.

Mishra, P., A. Jain, T. Takabe, Y. Tanaka, M. Negi, N. Singh, N. Jain, V. Mishra, R. Maniraj, S. L. Krishnamurthy, R. Sreevathsa, N. K. Singh and V. Rai (2019). "Heterologous Expression of Serine Hydroxymethyltransferase-3 From Rice Confers Tolerance to Salinity Stress in E. coli and Arabidopsis." Front Plant Sci 10: 217.

Mordukhova, E. A. and J. G. Pan (2013). "Evolved cobalamin-independent methionine synthase (MetE) improves the acetate and thermal tolerance of Escherichia coli." Appl Environ Microbiol 79(24): 7905-7915.

Morscher, R. J., G. S. Ducker, S. H. Li, J. A. Mayer, Z. Gitai, W. Sperl and J. D. Rabinowitz (2018). "Mitochondrial translation requires folate-dependent tRNA methylation." Nature 554(7690): 128-132.

Mouillon, J. M., S. Aubert, J. Bourguignon, E. Gout, R. Douce and F. Rebeille (1999). "Glycine and serine catabolism in non-photosynthetic higher plant cells: their role in C1 metabolism." Plant J 20(2): 197-205.

Mouillon, J. M., S. Ravanel, R. Douce and F. Rebeille (2002). "Folate synthesis in higher-plant mitochondria: coupling between the dihydropterin pyrophosphokinase and the dihydropteroate synthase activities." Biochem J 363(Pt 2): 313-319.

Navarrete, O., J. Van Daele, C. Stove, W. Lambert, D. Van Der Straeten and S. Storozhenko (2012). "A folate independent role for cytosolic HPPK/DHPS upon stress in Arabidopsis thaliana." Phytochemistry 73(1): 23-33.

Neuburger, M., A. M. Polidori, E. Pietre, M. Faure, A. Jourdain, J. Bourguignon, B. Pucci and R. Douce (2000). "Interaction between the lipoamide-containing H-protein and the lipoamide dehydrogenase (L-protein) of the glycine decarboxylase multienzyme system. 1. Biochemical studies." Eur J Biochem 267(10): 2882-2889.

Neuburger, M., F. Rebeille, A. Jourdain, S. Nakamura and R. Douce (1996). "Mitochondria are a major site for folate and thymidylate synthesis in plants." J Biol Chem 271(16): 9466-9472.

Okamura-Ikeda, K., K. Fujiwara and Y. Motokawa (1999). "The amino-terminal region of the Escherichia coli T-protein of the glycine cleavage system is essential for proper association with H-protein." Eur J Biochem 264(2): 446-453.

Okamura-Ikeda, K., N. Kameoka, K. Fujiwara and Y. Motokawa (2003). "Probing the H-protein-induced conformational change and the function of the N-terminal region of Escherichia coli T-protein of the glycine cleavage system by limited proteolysis." J Biol Chem 278(12): 10067-10072.

Orsomando, G., G. G. Bozzo, R. D. de la Garza, G. J. Basset, E. P. Quinlivan, V. Naponelli, F. Rebeille, S. Ravanel, J. F. Gregory, 3rd and A. D. Hanson (2006). "Evidence for folate-salvage reactions in plants." Plant J 46(3): 426-435.

Orsomando, G., R. D. de la Garza, B. J. Green, M. Peng, P. A. Rea, T. J. Ryan, J. F. Gregory, 3rd and A. D. Hanson (2005). "Plant gamma-glutamyl hydrolases and folate polyglutamates: characterization, compartmentation, and co-occurrence in vacuoles." J Biol Chem 280(32): 28877-28884.

Paukert, J. L. and J. C. Rabinowitz (1980). "Formyl-methenyl-methylenetetrahydrofolate synthetase (combined): a multifunctional protein in eukaryotic folate metabolism." Methods Enzymol 66: 616-626.

Pieroth, R., S. Paver, S. Day and C. Lammersfeld (2018). "Folate and Its Impact on Cancer Risk." Curr Nutr Rep 7(3): 70-84.

Piironen, V., M. Edelmann, S. Kariluoto and Z. Bedo (2008). "Folate in wheat genotypes in the HEALTHGRAIN Diversity Screen." J Agric Food Chem 56(21): 9726-9731.

Pike, S. T., R. Rajendra, K. Artzt and D. R. Appling (2010). "Mitochondrial C1-tetrahydrofolate synthase (MTHFD1L) supports the flow of mitochondrial one-carbon units into the methyl cycle in embryos." J Biol Chem 285(7): 4612-4620.

Piper, M. D., S. P. Hong, G. E. Ball and I. W. Dawes (2000). "Regulation of the balance of one-carbon metabolism in Saccharomyces cerevisiae." J Biol Chem 275(40): 30987-30995.

Pogribny, I. P., K. Kutanzi, S. Melnyk, A. de Conti, V. Tryndyak, B. Montgomery, M. Pogribna, L. Muskhelishvili, J. R. Latendresse, S. J. James, F. A. Beland and I. Rusyn (2013). "Strain-dependent dysregulation of one-carbon metabolism in

male mice is associated with choline- and folate-deficient diet-induced liver injury." FASEB J 27(6): 2233-2243.

Powell, C. M., T. L. Rudge, Q. Zhu, L. F. Johnson and U. Hansen (2000). "Inhibition of the mammalian transcription factor LSF induces S-phase-dependent apoptosis by downregulating thymidylate synthase expression." EMBO J 19(17): 4665-4675.

Pribat, A., L. Jeanguenin, A. Lara-Nunez, M. J. Ziemak, J. E. Hyde, V. de Crecy-Lagard and A. D. Hanson (2009). "6-pyruvoyltetrahydropterin synthase paralogs replace the folate synthesis enzyme dihydroneopterin aldolase in diverse bacteria." J Bacteriol 191(13): 4158-4165.

Puthusseri, B., P. Divya, V. Lokesh, G. Kumar, M. A. Savanur and B. Neelwarne (2018). "Novel Folate Binding Protein in Arabidopsis Expressed during Salicylic Acid-Induced Folate Accumulation." J Agric Food Chem 66(2): 505-511.

Puthusseri, B., P. Divya, V. Lokesh and B. Neelwarne (2012). "Enhancement of folate content and its stability using food grade elicitors in coriander (Coriandrum sativum L.)." Plant Foods Hum Nutr 67(2): 162-170.

Ramirez Rivera, N. G., C. Garcia-Salinas, F. J. Aragao and R. I. Diaz de la Garza (2016). "Metabolic engineering of folate and its precursors in Mexican common bean (Phaseolus vulgaris L.)." Plant Biotechnol J 14(10): 2021-2032.

Ravanel, S., M. A. Block, P. Rippert, S. Jabrin, G. Curien, F. Rebeille and R. Douce (2004). "Methionine metabolism in plants: chloroplasts are autonomous for de novo methionine synthesis and can import S-adenosylmethionine from the cytosol." J Biol Chem 279(21): 22548-22557.

Ravanel, S., H. Cherest, S. Jabrin, D. Grunwald, Y. Surdin-Kerjan, R. Douce and F. Rebeille (2001). "Tetrahydrofolate biosynthesis in plants: molecular and functional characterization of dihydrofolate synthetase and three isoforms of folylpolyglutamate synthetase in Arabidopsis thaliana." Proc Natl Acad Sci U S A 98(26): 15360-15365.

Rebeille, F., D. Macherel, J. M. Mouillon, J. Garin and R. Douce (1997). "Folate biosynthesis in higher plants: purification and molecular cloning of a bifunctional 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase/7,8-dihydropteroate synthase localized in mitochondria." EMBO J 16(5): 947-957.

Rebelo, J., G. Auerbach, G. Bader, A. Bracher, H. Nar, C. Hosl, N. Schramek, J. Kaiser, A. Bacher, R. Huber and M. Fischer (2003). "Biosynthesis of pteridines. Reaction mechanism of GTP cyclohydrolase I." J Mol Biol 326(2): 503-516.

Reyes-Hernandez, B. J., A. C. Srivastava, Y. Ugartechea-Chirino, S. Shishkova, P. A. Ramos-Parra, V. Lira-Ruan, R. I. Diaz de la Garza, G. Dong, J. C. Moon, E. B. Blancaflor and J. G. Dubrovsky (2014). "The root indeterminacy-to-determinacy developmental switch is operated through a folate-dependent pathway in Arabidopsis thaliana." New Phytol 202(4): 1223-1236.

Riaz, B., Q. Liang, X. Wan, K. Wang, C. Zhang and X. Ye (2019). "Folate content analysis of wheat cultivars developed in the North China Plain." Food Chem 289: 377-383.

Robinson, B. R., C. Garcia Salinas, P. Ramos Parra, J. Bamberg, R. I. Diaz de la Garza and A. Goyer (2019). "Expression Levels of the γ-Glutamyl Hydrolase I Gene Predict Vitamin B9 Content in Potato Tubers." Agronomy 9(11).

Rodionova, I. A., N. Goodacre, J. Do, A. Hosseinnia, M. Babu, P. Uetz and M. H. Saier, Jr. (2018). "The uridylyltransferase GlnD and tRNA modification GTPase MnmE allosterically control Escherichia coli folylpoly-gamma-glutamate synthase FolC." J Biol Chem 293(40): 15725-15732.

Roje, S. (2007). "Vitamin B biosynthesis in plants." Phytochemistry 68(14): 1904-1921.

Roje, S., M. T. Janave, M. J. Ziemak and A. D. Hanson (2002). "Cloning and characterization of mitochondrial 5-formyltetrahydrofolate cycloligase from higher plants." J Biol Chem 277(45): 42748-42754.

Roje, S., H. Wang, S. D. McNeil, R. K. Raymond, D. R. Appling, Y. Shachar-Hill, H. J. Bohnert and A. D. Hanson (1999). "Isolation, characterization, and functional expression of cDNAs encoding NADH-dependent methylenetetrahydrofolate reductase from higher plants." J Biol Chem 274(51): 36089-36096.

Rossi, M., A. Amaretti and S. Raimondi (2011). "Folate production by probiotic bacteria." Nutrients 3(1): 118-134.

Roux, B. and C. T. Walsh (1992). "p-aminobenzoate synthesis in Escherichia coli: kinetic and mechanistic characterization of the amidotransferase PabA." Biochemistry 31(30): 6904-6910.

Sah, S., S. Aluri, K. Rex and U. Varshney (2015). "One-carbon metabolic pathway rewiring in Escherichia coli reveals an evolutionary advantage of 10-formyltetrahydrofolate synthetase (Fhs) in survival under hypoxia." J Bacteriol 197(4): 717-726.

Sahr, T., S. Ravanel, G. Basset, B. P. Nichols, A. D. Hanson and F. Rebeille (2006). "Folate synthesis in plants: purification, kinetic properties, and inhibition of aminodeoxychorismate synthase." Biochem J 396(1): 157-162.

Schirch, V., S. Hopkins, E. Villar and S. Angelaccio (1985). "Serine hydroxymethyltransferase from Escherichia coli: purification and properties." J Bacteriol 163(1): 1-7.

Schirch, V. and D. M. Szebenyi (2005). "Serine hydroxymethyltransferase revisited." Curr Opin Chem Biol 9(5): 482-487.

Schoedon, G., U. Redweik, G. Frank, R. G. Cotton and N. Blau (1992). "Allosteric characteristics of GTP cyclohydrolase I from Escherichia coli." Eur J Biochem 210(2): 561-568.

Schwahn, B. C., M. D. Laryea, Z. Chen, S. Melnyk, I. Pogribny, T. Garrow, S. J. James and R. Rozen (2004). "Betaine rescue of an animal model with methylenetetrahydrofolate reductase deficiency." Biochem J 382(Pt 3): 831-840.

Shetty, S. and U. Varshney (2021). "Regulation of translation by one-carbon metabolism in bacteria and eukaryotic organelles." J Biol Chem 296: 100088.

Slansky, J. E. and P. J. Farnham (1996). "Transcriptional regulation of the dihydrofolate reductase gene." Bioessays 18(1): 55-62.

Srivastava, A. C., F. Chen, T. Ray, S. Pattathil, M. J. Pena, U. Avci, H. Li, D. V. Huhman, J. Backe, B. Urbanowicz, J. S. Miller, M. Bedair, C. E. Wyman, L. W. Sumner, W. S. York, M. G. Hahn, R. A. Dixon, E. B. Blancaflor and Y. Tang (2015). "Loss of function of folylpolyglutamate synthetase 1 reduces lignin content and improves cell wall digestibility in Arabidopsis." Biotechnol Biofuels 8: 224.

Srivastava, A. C., P. A. Ramos-Parra, M. Bedair, A. L. Robledo-Hernandez, Y. Tang, L. W. Sumner, R. I. Diaz de la Garza and E. B. Blancaflor (2011). "The folylpolyglutamate synthetase plastidial isoform is required for postembryonic root development in Arabidopsis." Plant Physiol 155(3): 1237-1251.

Storozhenko, S., V. De Brouwer, M. Volckaert, O. Navarrete, D. Blancquaert, G. F. Zhang, W. Lambert and D. Van Der Straeten (2007). "Folate fortification of rice by metabolic engineering." Nat Biotechnol 25(11): 1277-1279.

Stover, P. and V. Schirch (1990). "Serine hydroxymethyltransferase catalyzes the hydrolysis of 5,10-methenyltetrahydrofolate to 5-formyltetrahydrofolate." J Biol Chem 265(24): 14227-14233.

Sybesma, W., E. Van Den Born, M. Starrenburg, I. Mierau, M. Kleerebezem, W. M. De Vos and J. Hugenholtz (2003). "Controlled modulation of folate polyglutamyl tail length by metabolic engineering of Lactococcus lactis." Appl Environ Microbiol 69(12): 7101-7107.

Tamura, T., K. H. Hong, Y. Mizuno, K. E. Johnston and C. L. Keen (1999). "Folate and homocysteine metabolism in copper-deficient rats." Biochim Biophys Acta 1427(3): 351-356.

Tamura, T. and J. R. Turnlund (2004). "Effect of long-term, high-copper intake on the concentrations of plasma homocysteine and B vitamins in young men." Nutrition 20(9): 757-759.

Tang, M., S. Q. Wang, B. J. Liu, Q. Cao, B. J. Li, P. C. Li, Y. F. Li, C. Qin and W. Zhang (2014). "The methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism and tumor risk: evidence from 134 case-control studies." Mol Biol Rep 41(7): 4659-4673.

Tibbetts, A. S. and D. R. Appling (2010). "Compartmentalization of Mammalian folate-mediated one-carbon metabolism." Annu Rev Nutr 30: 57-81.

Tio, M., J. Andrici, M. R. Cox and G. D. Eslick (2014). "Folate intake and the risk of upper gastrointestinal cancers: a systematic review and meta-analysis." J Gastroenterol Hepatol 29(2): 250-258.

Trimmer, E. E., D. P. Ballou, L. J. Galloway, S. A. Scannell, D. R. Brinker and K. R. Casas (2005). "Aspartate 120 of Escherichia coli methylenetetrahydrofolate

reductase: evidence for major roles in folate binding and catalysis and a minor role in flavin reactivity." Biochemistry 44(18): 6809-6822.

Verberne, M. C., K. Sansuk, J. F. Bol, H. J. Linthorst and R. Verpoorte (2007). "Vitamin K1 accumulation in tobacco plants overexpressing bacterial genes involved in the biosynthesis of salicylic acid." J Biotechnol 128(1): 72-79.

Viswanathan, V. K., J. M. Green and B. P. Nichols (1995). "Kinetic characterization of 4-amino 4-deoxychorismate synthase from Escherichia coli." J Bacteriol 177(20): 5918-5923.

Voll, L. M., A. Jamai, P. Renne, H. Voll, C. R. McClung and A. P. Weber (2006). "The photorespiratory Arabidopsis shm1 mutant is deficient in SHM1." Plant Physiol 140(1): 59-66.

Waller, J. C., T. A. Akhtar, A. Lara-Nunez, J. F. Gregory, 3rd, R. P. McQuinn, J. J. Giovannoni and A. D. Hanson (2010). "Developmental and feedforward control of the expression of folate biosynthesis genes in tomato fruit." Mol Plant 3(1): 66-77.

Wang, W., Y. Wang, F. Gong, W. Zhu and S. Fu (2013). "MTHFR C677T polymorphism and risk of congenital heart defects: evidence from 29 case-control and TDT studies." PLoS One 8(3): e58041.

Wang, Y., Y. Li, Y. Wu and H. Yan (2007). "Mechanism of dihydroneopterin aldolase. NMR, equilibrium and transient kinetic studies of the Staphylococcus aureus and Escherichia coli enzymes." FEBS J 274(9): 2240-2252.

Wegkamp, A., W. van Oorschot, W. M. de Vos and E. J. Smid (2007). "Characterization of the role of para-aminobenzoic acid biosynthesis in folate production by Lactococcus lactis." Appl Environ Microbiol 73(8): 2673-2681.

Wohlfarth, G., G. Geerligs and G. Diekert (1991). "Purification and characterization of NADP(+)-dependent 5,10-methylenetetrahydrofolate dehydrogenase from Peptostreptococcus productus marburg." J Bacteriol 173(4): 1414-1419.

Xiang, N., J. Hu, T. Wen, M. A. Brennan, C. S. Brennan and X. Guo (2020). "Effects of temperature stress on the accumulation of ascorbic acid and folates in sweet corn (Zea mays L.) seedlings." J Sci Food Agric 100(4): 1694-1701.

Yadav, U., P. Kumar, S. K. Yadav, O. P. Mishra and V. Rai (2015). ""Polymorphisms in folate metabolism genes as maternal risk factor for neural tube defects: an updated meta-analysis"." Metab Brain Dis 30(1): 7-24.

Yan, H. and X. Ji (2011). "Role of protein conformational dynamics in the catalysis by 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase." Protein Pept Lett 18(4): 328-335.

Yan, X., L. Ma, H. Pang, P. Wang, L. Liu, Y. Cheng, J. Cheng, Y. Guo and Q. Li (2019). "METHIONINE SYNTHASE1 Is Involved in Chromatin Silencing by Maintaining DNA and Histone Methylation." Plant Physiol 181(1): 249-261.

Ye, J., J. Fan, S. Venneti, Y. W. Wan, B. R. Pawel, J. Zhang, L. W. Finley, C. Lu, T. Lindsten, J. R. Cross, G. Qing, Z. Liu, M. C. Simon, J. D. Rabinowitz and C. B.

Thompson (2014). "Serine catabolism regulates mitochondrial redox control during hypoxia." Cancer Discov 4(12): 1406-1417.

Ye, Q. Z., J. Liu and C. T. Walsh (1990). "p-Aminobenzoate synthesis in Escherichia coli: purification and characterization of PabB as aminodeoxychorismate synthase and enzyme X as aminodeoxychorismate lyase." Proc Natl Acad Sci U S A 87(23): 9391-9395.

Zhang, H., X. Deng, D. Miki, S. Cutler, H. La, Y. J. Hou, J. Oh and J. K. Zhu (2012). "Sulfamethazine suppresses epigenetic silencing in Arabidopsis by impairing folate synthesis." Plant Cell 24(3): 1230-1241.

Zhang, Y., H. Jia, S. Wang and D. Jiang (2017). "Cumulative review and meta-analyses on the association between MTHFR rs1801133 polymorphism and breast cancer risk: a pooled analysis of 83 studies with 74,019 participants." Minerva Med 108(1): 57-73.

Zhou, H. R., F. F. Zhang, Z. Y. Ma, H. W. Huang, L. Jiang, T. Cai, J. K. Zhu, C. Zhang and X. J. He (2013). "Folate polyglutamylation is involved in chromatin silencing by maintaining global DNA methylation and histone H3K9 dimethylation in Arabidopsis." Plant Cell 25(7): 2545-2559.

# 2

# Objectives of research

Folate is a group of water-soluble B vitamins. Humans cannot synthesize folates de novo, therefore depending on daily diet. Folate deficiencies increase the risk of cancers, neural-tube defects, and congenital heart defects. Improving folate accumulation in the edible parts of crops and vegetables by biofortification has proven to be an efficient strategy to alleviate folate deficiency. Maize is one of the world's most important cereal grains which provides an estimated 15% of the global protein and 20% of the global calories (Shiferaw, Prasanna et al. 2011). The wet-milling maize kernel can be used for producing feed product, corn oil, ethanol and starch, and dry-milling corn usually for animal feed, liquor products, arepa flour, and so on (Nuss and Tanumihardjo 2010). Nutritionally, maize contains high levels of carbohydrate (74.26g per 100 g FW) but low levels of folates (19 μg per 100g FW). The two-gene strategy of folate biofortification successfully increased total folate levels by 150 folds in rice and tomato, but only 4 folds in maize (Diaz de la Garza, Gregory et al. 2007, Blancquaert, Van Daele et al. 2015, Liang, Wang et al. 2019). Another study for simultaneous increase of multiple vitamins in transgenic maize showed a 169-fold increase of β-carotene, 6-fold increase of ascorbate and only 2-fold increae of folates as compared to wild type (Naqvi et al. 2009). It appears that the two-gene approach has effects on folate accumulation, while the outcome may be species-dependent. So far, folate enhancement in potato and Arabidopsis is unsuccessful in the two-gene strategy. The activity of the endogenous downstream enzymes (such as HPPK-DHPS) is insufficient for folate enhancement in those aforementioned cases. Whether it is caused by the species-specific regulatory mechanism of folate biosynthesis is still unclear. Maize is an important model plant for the research of plant genetics, genomics and molecular biology. Investigating the regulatory mechanism of folate in maize will provide insight into folate accumulation that help us to carry out the biofortification efficiently in those folate-insufficient species.

Even though genetic modification of crops can lead to a significant improvement of nutrition and yield, there exist concerns about human health (Kramkowska, Grzelak et al. 2013). To date, the transgenic folate-fortified crops are still not approved for commercial release due to its safety issue. Thus, molecular marker-assisted breeding for high folate crops will be an alternative for sufficient daily folate intake. Folic acid metabolism in humans is influenced by genetic polymorphism. The MTHFR C677T polymorphism reduces the activity of MTHFR, resulting in the inability of the MTHFR enzyme to catalyze the conversion of 5,10-$CH_2$-THF to 5-M-THF. This allele frequency is reported to be more than 10% in many countries like Italy, Japan, America, and Canada, even up to 25% in Colombian (Hispanic Whites) (Liew and Gupta 2015). Hence, the folic acid from industrial fortification is usually utilized inefficiently. On the other hand, high folic acid consumption sometimes masks symptoms of vitamin B12 deficiency (Oakley 2002). However, this is not problematic for the 5-M-THF supplement. The natural 5-M-THF can be absorbed directly into the human body and perform its health-promoting function (Gutstein, Bernstein et al. 1973). Meanwhile, 5-M-THF is the most abundant folate derivative in two-gene transgenic plants, especially accounting for up to 89% of total folate in biofortified rice (Storozhenko, De Brouwer et al. 2007). Therefore, folate biofortification in crops

is a healthier strategy for folate supplementation. However, the transgenic strategy of folate biofortification suggests some unknown factors limiting the conversion of pterin to folate. Meanwhile, the folate biofortified crops usually contain a very high level of pteridine, which has potential health risks because the effects of high pteridine intake are unknown (Diaz de la Garza, Quinlivan et al. 2004). A better understanding of the regulatory mechanism of folate accumulation is required to address the above problem.

Our previous work demonstrates that the folate regulatory mechanism in maize kernel is complicated (Lian, Guo et al. 2015). Even though we have already identified the folate-related gene in maize using bioinformatics analysis. However, genes and metabolism pathways controlling folate accumulation during kernel development have not yet been well studied. To uncover the metabolic pathways/genes that affect folate accumulation during kernel formation through transcriptional regulation, two maize inbred lines with different patterns or levels of folate accumulation in mature kernels were subjected to comparative transcriptome analysis. The weighted gene co-expression network analysis was performed to identify the rate-limiting genes whose transcriptional level had positive or negative correlation with folate accumulation levels. Besides, to obtain the quantitative trait loci (QTLs) which are associated with folate contents in maize kernel, we used a segregated population crossed by two maize lines for QTLs mapping, and the candidate genes related to folate were identified through bioinformatic analysis and functional analysis. In conclusion, the genes participating in folate metabolism have been identified, thus helping develop new strategies for folate biofortification.

# References

Blancquaert, D., H. De Steur, X. Gellynck and D. Van Der Straeten (2014). "Present and future of folate biofortification of crop plants." J Exp Bot 65(4): 895-906.

Diaz de la Garza, R., E. P. Quinlivan, S. M. Klaus, G. J. Basset, J. F. Gregory, 3rd and A. D. Hanson (2004). "Folate biofortification in tomatoes by engineering the pteridine branch of folate synthesis." Proc Natl Acad Sci U S A 101(38): 13720-13725.

Gutstein, S., L. H. Bernstein, L. Levy and G. Wagner (1973). "Failure of response to N 5 -methyltetrahydrofolate in combined folate and B 12 deficiency. Evidence in support of the "folate trap" hypothesis." Am J Dig Dis 18(2): 142-146.

Kramkowska, M., T. Grzelak and K. Czyzewska (2013). "Benefits and risks associated with genetically modified food products." Ann Agric Environ Med 20(3): 413-419.

Lian, T., W. Guo, M. Chen, J. Li, Q. Liang, F. Liu, H. Meng, B. Xu, J. Chen, C. Zhang and L. Jiang (2015). "Genome-wide identification and transcriptional analysis of folate metabolism-related genes in maize kernels." BMC Plant Biol 15: 204.

Liang, Q., K. Wang, X. Liu, B. Riaz, L. Jiang, X. Wan, X. Ye and C. Zhang (2019). "Improved folate accumulation in genetically modified maize and wheat." J Exp Bot 70(5): 1539-1551.

Naqvi, S., C. Zhu, G. Farre, K. Ramessar, L. Bassie, J. Breitenbach, D. Perez Conesa, G. Ros, G. Sandmann, T. Capell and P. Christou (2009). "Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways." Proc Natl Acad Sci U S A 106(19): 7762-7767.

Nuss, E. T. and S. A. Tanumihardjo (2010). "Maize: A Paramount Staple Crop in the Context of Global Nutrition." Compr Rev Food Sci Food Saf 9(4): 417-436.

Oakley, G. P., Jr. (2002). "Inertia on folic acid fortification: public health malpractice." Teratology 66(1): 44-54.

Shiferaw, B., B. M. Prasanna, J. Hellin and M. Bänziger (2011). "Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security." Food Security 3(3): 307.

# 3

# Comparative transcriptome analysis reveals mechanisms of folate accumulation in maize grains

*In this chapter, two maize inbred lines with different patterns of folate accumulation and different total folate contents in mature kernels were used to investigate the transcriptional regulation of folate metabolism during kernel development. The folate accumulation from DAP 24 to mature kernels might be regulated by the circumjacent pathways of folate biosynthesis such as pyruvate metabolism, glutamate metabolism and glycine/serine metabolism; The folate variation of inbred-specific might be controlled by those genes involved in the two branches of folate biosynthesis, serine/THF/5-M-THF cycle and conversion of THF monoglutamate into THF polyglutamate. The results provide insight into the mechanism of folate accumulation, and will help the futural breeding work of folate biofortification*

From Tong Lian*, Xuxia Wang* et al. Comparative transcriptome analysis reveals mechanisms of folate accumulation in maize grains. Int J of Mol Sci. 2022, 23(3): 1708 Doi: 10.3390/ijms23031708

**Abstract:** Previously, the complexity of folate accumulation in the early stages of maize kernel development has been reported, but the mechanisms of folate accumulation remain unclear. Two maize inbred lines, DAN3130 and JI63, with different patterns of folate accumulation and different folate levels in mature kernels were used to investigate the transcriptional regulation of folate metabolism during late stages of kernel formation by comparative transcriptome analysis. The folate accumulation during DAP 24 to mature kernels could be controlled by circumjacent pathways of folate biosynthesis, such as pyruvate metabolism, glutamate metabolism, and serine/glycine metabolism. In addition, the folate variation between these two inbred lines was related to those genes among folate metabolism, such as the genes in pteridine branch, para-aminobenzoate branch, serine/tetrahydrofolate (THF)/5-methyltetrahydrofolate cycle, and the conversion of THF monoglutamate to THF polyglutamate. The findings provided insight into folate accumulation mechanisms during maize kernel formation to promote folate biofortification.

**Key words**: folate accumulation; transcriptome; WGCNA; maize kernel

# 1 Introduction

Folate, a water-soluble vitamin B9, acts as the one-carbon acceptor and donor involved in the biosynthesis of amino acids, nucleic acids, proteins, and lipids (Gorelova, Ambach et al. 2017). In plants, folate can be synthesized de novo, which makes plants serve as a good folate source for humans. However, folate contents in edible parts of crops, such as maize (*Zea mays*), rice (*Oryza sativa*), and potato (*Solanum tuberosum*), are at low levels compared with green vegetables (Bekaert, Storozhenko et al. 2008, Blancquaert, Van Daele et al. 2015, Strobbe and Van Der Straeten 2017). Folate deficiency increases the risk of neural tube defects and always is a global public health issue (Ratan, Rattan et al. 2008, Palchetti, Steluti et al. 2020). Folate biofortification by metabolic engineering of crops, e.g., maize, rice, potato, tomato (*Solanum lycopersicum*) and wheat (*Triticum aestivum*), have been investigated to obtain folate-enriched food to alleviate this problem (Bekaert, Storozhenko et al. 2008). In plants, the connection between folate metabolism and other metabolites has been explored to understand the mechanism of folate accumulation. Most vitamins have the same upstream precursors from the pentose and triose pool, which produce chorismate for the biosynthesis of tocochromanols, phylloquinone, and folates, and ribose 5-P for the biosynthesis of thiamine, riboflavin, and folates (DellaPenna 2005, Sahr, Ravanel et al. 2006, Roje 2007, Verberne, Sansuk et al. 2007, Asensi-Fabado and Munne-Bosch 2010). Furthermore, glutamine and glutamate, which are involved in a set of reactions within the specific and shared biosynthetic pathways of B vitamins and purine, connects to thiamine, riboflavin, and folates (Asensi-Fabado and Munne-Bosch 2010). In addition, the dysfunctions of folate metabolism impact the balance of NADPH/NADP ratio, nitrogen metabolism, and photorespiration in plants (Collakova, Goyer et al. 2008, Jiang, Liu et al. 2013, Fan, Ye et al. 2014, Gorelova, De Lepeleire et al. 2017, Meng, Xu et al. 2017, Chen, Zhang et al. 2019). These complexities indicate that plant folate accumulation in plants could be affected by multiple elements and environmental cues.

The folate biosynthesis pathway has been well described in plants (Hanson and Gregory 2011). The upstream precursors for folate mainly come from carbohydrate metabolism, which provides erythrose-P and phosphoenolpyruvate for chorismate synthesis, and ribose 5-P for GTP synthesis (Roje 2007, Asensi-Fabado and Munne-Bosch 2010). Chorismate can be converted to para-aminobenzoate (ρ-ABA) by aminodeoxychorismate synthase (ADCS) and aminodeoxychorismate lyase (ADCL) in plastids. Moreover, GTP can be converted to 6-hydroxymethyldihydropterin (HMDHP) by GTP cyclohydrolase I (GCHI) and dihydroneopterin aldolase (DHNA) in the cytosol. Subsequently, the HMDHP and ρ-ABA are assembled into dihydrofolate (DHF) by HMDHP pyrophosphokinase/dihydropteroate synthase (HPPK/DHPS) and dihydrofolate synthetase (DHFS) in the mitochondrion. Later, bifunctional dihydrofolate reductase-thymidylate synthase (DHFR-TS, DRTS) catalyzes the DHF to tetrahydrofolate (THF). Then, THF can be polyglutamylated through folylpolyglutamate synthetase (FPGS). Afterward, the glutamate tail can be removed by γ-glutamyl hydrolase (GGH) (Orsomando, de la Garza et al. 2005). In one-carbon metabolism, one-carbon units, which derive from polyglutamylated THF,

can be generated directly within the conversion of THF and serine to glycine and 5,10-CH2-THF by serine hydroxymethyltransferase (SHMT) and glycine cleavage system (GCS). Later, the conversion among different folate derivatives responsible for the biosynthesis of purine, thymidylate (dTMP), and methionine by dehydrogenase/5,10-CH=THF cyclohydrolase (DHC), 10-formyl THF deformylase (10-FDF), 10-formyltetrahydrofolate synthetase (FTHS), methylenetetrahydrofolate reductase (MTHFR), methionine synthesis (MS), and 5-formyl THF cycloligase (5-FCL) (De Lepeleire, Strobbe et al. 2018, Danchin, Sekowska et al. 2020). In plants, several folate biosynthetic genes are developmentally regulated. For example, leaves, roots, and stems express the HPPK-DHPS enzyme during seedling development in pea (*Pisum sativum*) (Jabrin, Ravanel et al. 2003). In tomatoes, GCHI transcript and protein level are expressed strongly in unripe fruit but not ripe fruit. The *ADCS* and *ADCL* transcript levels in mature green fruits were approximately 10% in leaves; then, they decreased to nothing after the breaker stage (Basset, Quinlivan et al. 2002, Basset, Ravanel et al. 2004). Otherwise, some others are subject to feedforward control by pathway intermediates (Waller, Akhtar et al. 2010). For instance, both plant ADCS and DHPS are inhibited by the downstream products DHP and DHF *in vitro* (Mouillon, Ravanel et al. 2002, Sahr, Ravanel et al. 2006) and overexpressing exogenous *GCHI* and *ADCS* in tomato fruit; the transcription levels of the respective objective genes do not change, whereas those of the downstream genes *ADCL*, *DHNA*, and mitochondrial *FPGS* are increased (Waller, Akhtar et al. 2010). Indeed, the transcript expression levels of several folate biosynthesis genes were associated with the folate contents, such as *GGH1* in potato (Robinson, Garcia Salinas et al. 2019) as well as *GCHI* and the *ADCS* in *pak choi* (Shohag, Wei et al. 2020). Nevertheless, the transcriptional regulation of plant folate metabolism still needs more exploration.

Maize is a significant source of food, feed, and industrial products (Guo, Lian et al. 2019). The complexity of folate accumulation in the early stages of maize kernel development has been reported, but mechanisms of folate accumulation have not been well revealed (Lian, Guo et al. 2015). In this study, transcriptional levels of the folate-related genes in low- and high folate maize inbred lines during the late stages of kernel formation were investigated by RNA-seq analysis. Comparative transcriptome analysis and weighted gene co-expression network analysis (WGCNA) were performed to reveal how folates accumulate in maize kernel. The results showed that folate accumulation during the late stages of kernel development might be mostly related to pyruvate metabolism, glutamate metabolism, and serine/lysine metabolism. Additionally, the variation of folate concentration between different inbred lines may be relevant to those genes belonging to the pteridine branch, ρ-ABA branch, serine/THF/5-methyltetrahydrofolate (5-M-THF) cycle, and conversion of monoglutamate THF to polyglutamate THF. The findings provided valuable knowledge of the folate accumulation mechanism during maize kernel formation for further folate biofortification and breeding in the future.

# 2 Results

## *2.1 Folate metabolic profiling during late developmental stages of kernels*

Two maize inbred lines DAN3130 and JI63 had a significant difference in total folate content in mature kernels. Four kinds of folate derivatives were detected by liquid chromatography-tandem mass spectroscopy (LC/MS), including 5-M-THF, 5-formyltetrahydrofolate (5-F-THF), 5,10-CH=THF, and THF. At the same harvest stage (mature kernel), the total folate content of DAN3130 was 2.25-fold that of JI63 (0.99 ± 0.02 nmol/g DW versus 0.44 ± 0.02 nmol/g DW), and their differences were attributed to the inbred-specific variation (**Figure 3-1A**, **Table S3-1**). In contrast, they shared a similar proportion of folate derivatives (5-M-THF, 55% versus 51%; 5-F-THF, 34% versus 36%; 5,10-CH=THF, 6.7% versus 7.5%; THF, 4.2% versus 5.5%) (**Figure 3-1B**), indicating that 5-M-THF and 5-F-THF were the major folate derivatives in mature kernels.

To investigate the folate accumulation pattern during the late stage of kernel formation in maize, the immature kernels on the day after pollination (DAP) 24 and DAP 35 were collected for folate analysis. Interestingly, even though DAN3130 had higher levels of folates in mature kernels compared to JI63, its folate contents on DAP 24 and DAP 35 were significantly lower than JI63 (**Figure 3-1A**). Considering the patterns of folate accumulation from DAP 24 to DAP 35, both inbred lines had a similar accumulation pattern. The level of 5-F-THF was stable, but 5-M-THF decreased significantly (**Figure 3-1A**). Whereas from DAP 35 to mature kernels, in DAN3130, the levels of 5-F-THF sharply increased while 5-M-THF was unchanged, resulting as the increase in total folates to 0.99 ± 0.05 nmol/g. In JI63, levels of 5-M-THF decreased significantly, and 5-F-THF was stable, resulting as the decrease in total folate to 0.44 ± 0.02 nmol/g (**Figure 3-1A**). From DAP 24 to mature kernels, the proportion of THF and 5-F-THF increased significantly while 5-M-THF decreased in both lines (**Figure 3-1B**). Then, the dynamic patterns of folate derivatives from DAP 24 to mature kernel were regarded as the development-specific variation. Thus, the development-specific variation and inbred-specific variation were further analyzed by using comparative transcriptomes to investigate the critical roles associated with folate accumulation during kernel formation.

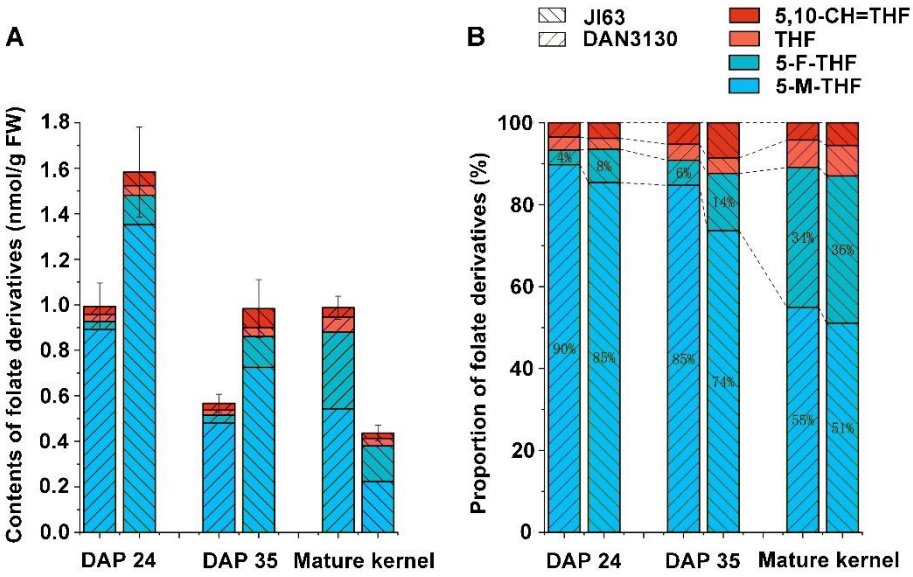



**Figure 3-1.** Folate profiling of maize kernel during late stages of development. A, contents of folate derivatives on DAP 24, DAP 35 and mature kernels in DAN3130 and JI63. B, the proportion of different folate derivatives. DAP, the day after pollination; 5-M-THF, 5-methyl-tetrahydrofolate; 5-F-THF, 5-formyl-tetrahydrofolate; 5,10-CH=THF, 5,10-Methenyltetrahydrofolate; THF, tetrahydrofolate.

## *2.2 Transcriptome profiling of DAN3130 and JI63 during kernel development by RNA-Seq*

DAN3130 and JI63 inbred lines were grown in Langfang, China. The fresh kernel samples of each inbred line were collected on DAP 24 and DAP 35, respectively, and then frozen in liquid nitrogen immediately. Mature kernel samples were harvested after all the plants turned yellow. Three biological replicates of each sample were collected, and total RNA with high quality was pooled and sent for sequencing. Total RNA of high quality was pooled for transcriptome analysis, and raw RNA-seq data of DAP 24, DAP35, and mature kernels for both two inbred lines were obtained (**Table S3-2**). All reads were aligned to the B73 genome, and most of the reads matched the maize genome with a unique location. The quality and accuracy of the sequencing data were sufficient for further analysis. All assembled transcripts were compared to genes in the NCBI non-redundant nucleotide (nt) database using BLASTn, and 50,810 unigenes with similarity to known genes were annotated. Then, those unannotated unigenes were searched against the Swiss-Prot and NCBI non-redundant protein (nr) databases using Trinoate with an E-value cutoff of $1 \times 10^{-3}$. Among the remaining 33,590 unigenes, 19,485 were annotated to known genes/proteins, and 14,105 were identified as non-protein-coding genes (**Figure 3-2**).

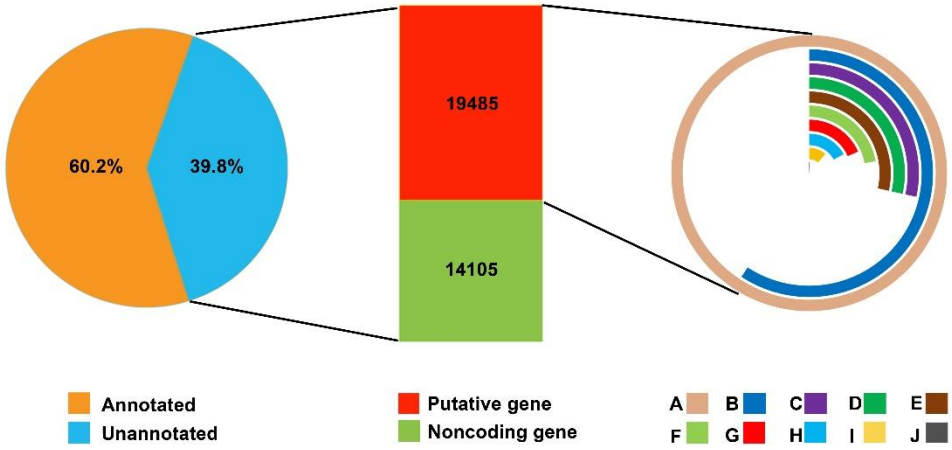



**Figure 3-2**. RNA-seq analysis of two inbred lines and annotation source. A: total numbers of genes; B: homologous analysis of DNA sequence (nr); C: SignalP prediction; D: TMHMM prediction; E: amino acid sequence prediction; F: homologous analysis of amino acid sequence; G: homologous analysis of DNA sequence (UniProt); H: PFAM prediction; I: homologous analysis of amino acid sequence (UniProt); J: RNA MMER identification.

## 2.3 Screening of Differentially Expressed Genes (DEGs) in the Development Specific Variation

To figure out the relationship between the folate accumulation and other metabolism pathways in maize, comparisons between transcriptomes of DAN3130 and JI63 were performed, and seven different comparisons were analyzed: for the development-specific variation during development in the same line (I: DAN3130 DAP 24 versus DAN3130 DAP 35, II: JI63 DAP 24 versus JI63 DAP 35, III: DAN3130 DAP 35 versus DAN3130 mature kernels, IV: JI63 DAP 35 versus JI63 mature kernels) and for the inbred-specific variation between two lines at the same time point (DAN3130 DAP 24 versus JI63 DAP 24, DAN3130 DAP 35 versus JI63 DAP 35, DAN3130 mature kernels versus JI63 mature kernels).

Since different folate derivatives had different accumulation patterns during late stages of kernel development, the comparisons of DAP 24 versus DAP 35 and DAP 35 versus mature kernels in the same inbred line were analyzed to figure out the transcriptional variation with folate change during kernel development. The results showed that the variation level in gene expression of DAP 35 versus mature kernels was more observable than those in DAP 24 versus DAP 35 (**Figure 3-3A**). The numbers of DEGs during DAP 35 to mature kernels (13,542 (III) and 12,575 (IV) DEGs) was much more than DAP 24 to DAP 35 (3,931 (I) and 4,345 (II) DEGs) (**Figure 3-3B**), indicating that the variational gene expression of DAP 35 versus mature kernels was more complicated than DAP 24 versus DAP 35. Furthermore, 770

common genes were differentially expressed among the comparisons of I, II, III, and IV (**Figure 3-3C**).

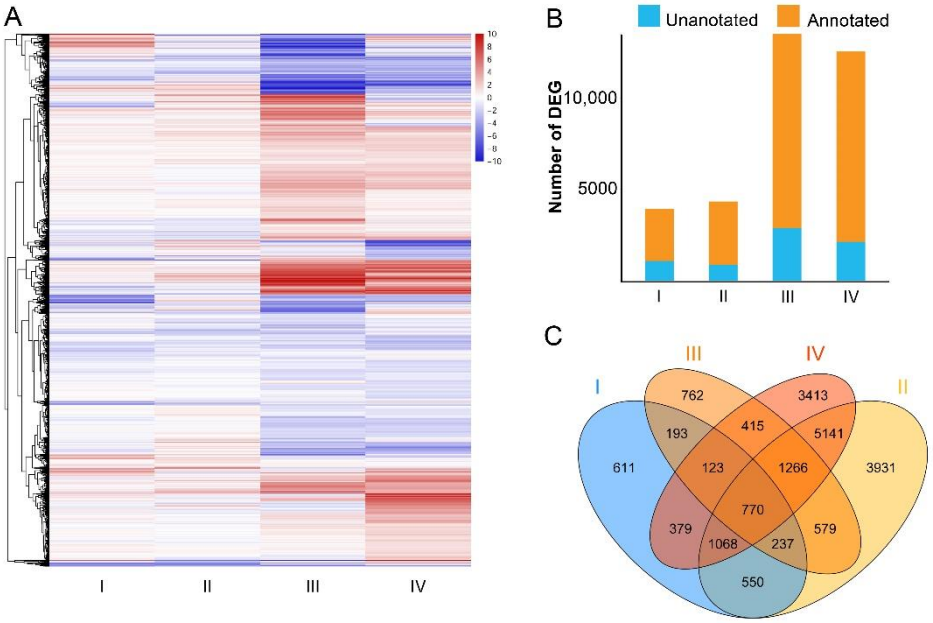



**Figure 3-3.** Screening of DEGs in development-specific variation; A, heatmap of DEGs in comparison be-tween different stages; B number of DEGs in comparison between different stages; C, Venn diagram of unique and common DEGs in comparison between different stages. Ⅰ, DAN3130 DAP 24 versus DAP 35; Ⅱ, JI63 DAP 24 versus DAP 35; Ⅲ, DAN3130 DAP 35 versus mature kernel; Ⅳ, JI63 DAP 35 versus mature kernel.

## 2.4 *Gene ontology (GO) enrichment and kyoto encyclopedia of genes and genomes (KEGG) analysis of DEGs in development-specific variation*

GO and KEGG enrichment analyses were carried out further to study the function of DEGs in development-specific variation. The DEGs of Comparison I and Comparison III enriched in DAN3130 were assigned to 343 and 552 GO terms ($p$-value < 0.05), respectively. For JI63, the total 333 GO terms ($p$-value < 0.05) were observed in Comparison II and 533 GO terms ($p$-value < 0.05) were assigned to Comparison IV. The top 20 GO terms were screened out based on enriched gene numbers (**Figure 3-4**). Considering the similar accumulation pattern that occurred from DAP 24 to DAP 35, it is noteworthy that common GO terms such as catalytic activity, hydrolase activity, oxidoreductase activity, metabolic process, response to stimulus, and nitrogen metabolic process were significantly enriched from DAP 24 to

DAP 35 in DAN3130 and JI63 (**Figure 3-4A**). Since the folate phenotype results showed different accumulation patterns of 5-M-THF and 5-F-THF from DAP 35 to mature kernel, we focused on the specific terms of *N*-methyltransferase activity and acyltransferase activity in JI63 and DAN3130, respectively (**Figure 3-4B**).

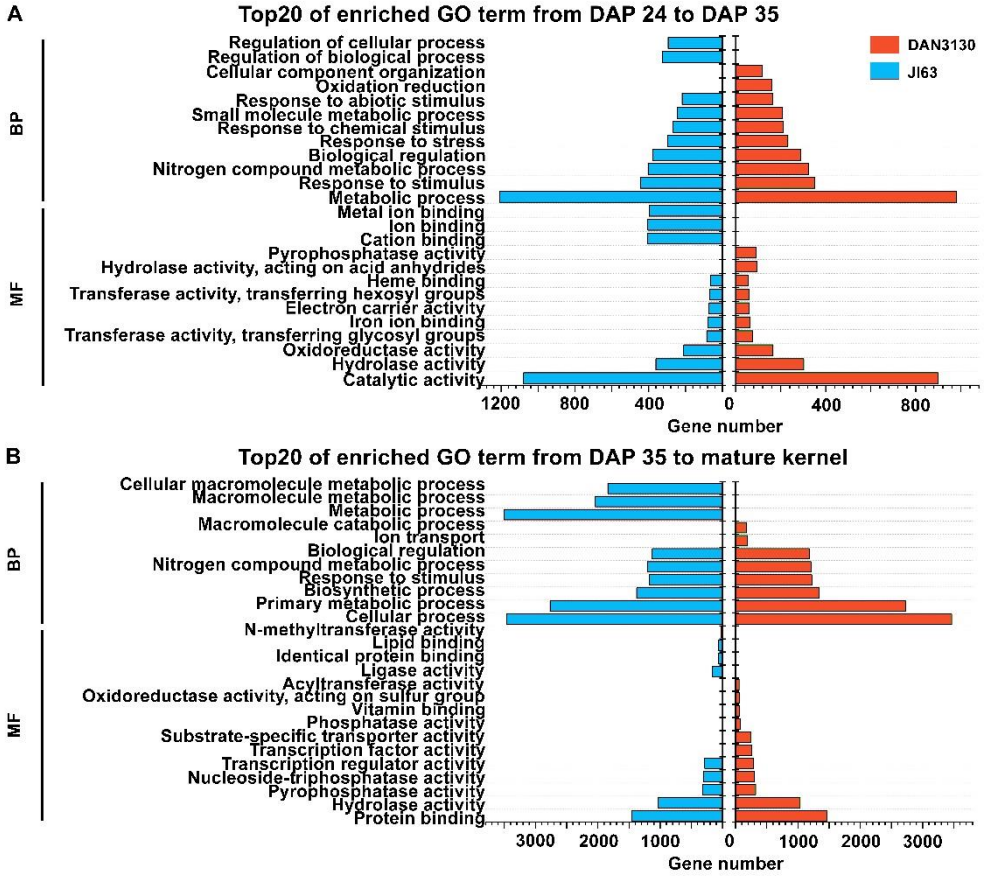



**Figure 3-4.** Bioinformatics analysis of DEGs in development-specific variation. A, GO function analysis of DEGs during DAP 24 to DAP 35; B GO function analysis of DEGs during DAP 35 to mature kernel. MF, molecular function; BP, biological process.

KEGG enrichment results showed DEGs from Comparison I and Comparison III assigned to 43 and 49 KEGG pathways ($p$-value < 0.05) in DAN3130, respectively. For JI63, a total of 31 and 52 KEGG pathways ($p$-value < 0.05) were identified from Comparison II and Comparison IV, respectively. The top 20 KEGG pathways based on the rich factor among the comparisons were screened out (**Figure 3-5**). The common pathways such as biosynthesis of secondary metabolites, nitrogen metabolism, glutamate metabolism and pyruvate metabolism were obviously enriched

from DAP 24 to DAP 35, while one-carbon pool by folate pathway was explicitly found in DAN3130 (**Figure 3-5A**). From DAP 35 to mature kernel, the pathway of glutamate metabolism and pyruvate metabolism were identified again in both lines, while the glycine/serine metabolism pathway, which might be relevant to folate metabolism, was only found on DAN3130 (**Figure 3-5B**). Therefore, the folate variation during DAP 24 to mature kernel might be affected by the circumjacent pathways of folate biosynthesis such as pyruvate metabolism, glutamate metabolism, and glycine/serine metabolism, and also, those genes serve a molecular function for methyltransferase activity and acyltransferase activity.

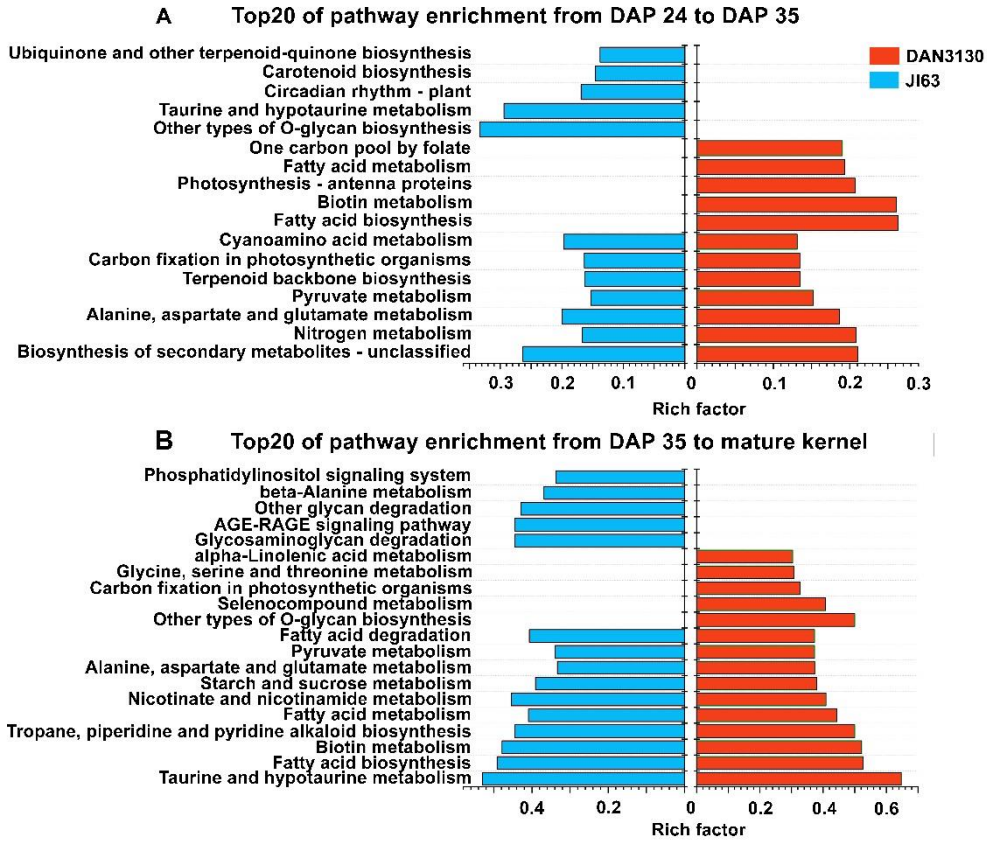



**Figure 3-5.** Bioinformatics analysis of DEGs in development-specific variation. A. KEGG enrichment of DEGs during DAP 24 to DAP 35; B. KEGG enrichment of DEGs during DAP 35 to mature kernel.

## *2.5 Screening of DEGs in inbred-specific variation*

Folate contents varied according to not only the development stages but also genotypes within species. Considering that significant differences in total folate contents were observed at the same harvest time point of these two lines, the comparison between two lines at the same harvest was summarized. A heatmap of DEGs in these comparisons showed that some genes had variational expression between two lines on mature kernels compared to DAP 24 and DAP 35 (**Figure 3-6A**). A total of 8395, 8193, and 10,608 DEGs on DAP 24, DAP 35, and the mature kernel were identified, respectively, indicating the variation of transcriptional levels in mature kernels was more complex than in the immature stage between two lines (**Figure 3-6B**). In all, 3463 DEGs were commonly regulated between these two inbred lines from the Venn diagram (**Figure 3-6C**).

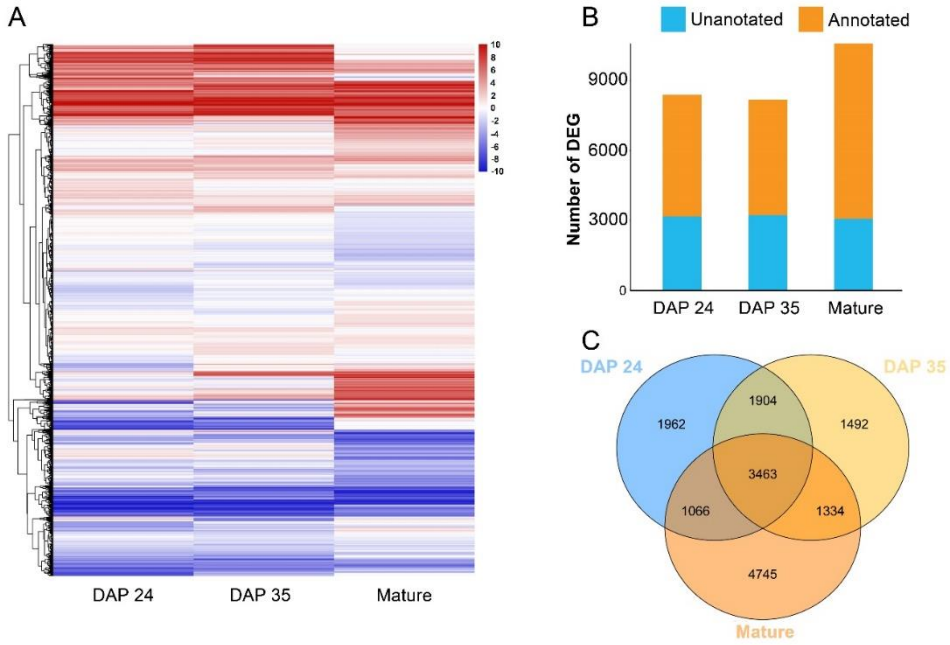



**Figure 3-6.** Screening of DEGs in the inbred-specific variation; A, Heatmap of DEGs in comparison between line DAN3130 and JI63 on DAP 24, DAP 35 and mature kernels; B number of DEGs in comparison between line DAN3130 and JI63 on DAP 24, DAP 35 and mature kernels; C, Venn diagram of unique and common DEGs in the comparison between line DAN3130 and JI63 on DAP 24, DAP 35 and mature kernels.

## *2.6 GO enrichment and KEGG analysis of DEGs in the inbred-specific variation*

Based on the GO results, the DEGs of DAN3130 versus JI63 enrichment were assigned 251 GO terms ($p$-value < 0.05), 225 GO terms ($p$-value < 0.05), and 358 GO terms ($p$-value < 0.05) in DAP 24, DAP 35, and mature kernels, respectively. The top

20 GO terms were screened out based on enriched gene numbers, and the results displayed that the metabolic process was significantly found in all three stages in the biological process. From molecular function, the largest numbers of genes involved in catalytic activity were obviously found in DAP 35 and mature kernels while not in DAP 24 (**Figure 3-7A, B, C**).

The DEGs from DAN3130 versus JI63 on DAP 24, DAP 35, and mature kernels belonged to 36, 28, and 51 pathways ($p$-value < 0.05), respectively. The top 20 KEGG pathways based on rich factor were screened out in **Figure 3-7D, E, F**. Consistent with the KEGG result of development-specific variation, glutamate metabolism and pyruvate metabolism were identified again in inbred-specific variation. Unlike development-specific variation, the KEGG results of the inbred-specific variation at DAP 24, DAP 35, and mature kernel, respectively, all contain folate metabolism pathways, including folate biosynthesis and one-carbon pool by folate, indicating that the folate-related metabolism pathway might have contributions to folate variation of different inbred lines. In addition, several pathways indirectly involved in folate, such as vitamin B6 metabolism and thiamine metabolism, were also found. These results further demonstrated that folate biosynthesis and the one-carbon metabolism pathway could be tightly associated with folate variation of different inbred lines.


**Figure 3-7. Bioinformatics analysis of DEGs in inbred-specific variation.** A, B, C: GO function analysis of DEGs on DAP 24, DAP 35 and mature kernel; D, E, F KEGG enrichment of DEGs on DAP 24, DAP 35 and mature kernel.

## *2.7 WGCNA analysis*

To systematically understand the gene sets related to the biological processes of folate accumulation, the WGCNA method was used to evaluate the gene expression network during kernel formation. The sample dendrogram and trait heatmap are shown in Figure S3-1A. The soft thresholding powers with $\beta = 7$ (scale free $R2 = 0.82$) was selected for screening network topology and further cluster dendrogram analysis

(Figure S3-1B,C). Finally, a total of 13,955 genes were classified into 15 modules with similar expression trends, and ten modules were either positively (correlation ≥ 0.5, $p < 0.05$) or negatively (correlation ≤−0.5, $p < 0.05$) associated with different folate derivatives (**Figure 3-8A**). The gene expression trends of the modules related to 5-M-THF and 5-F-THF were presented, and the turquoise module showed significant expression variation from DAP 35 to mature kernel (**Figure 3-8B, C**). The gene expression trends of modules, which had no correlation with folate, are shown in the supplementary materials (**Figure 3-S2**).

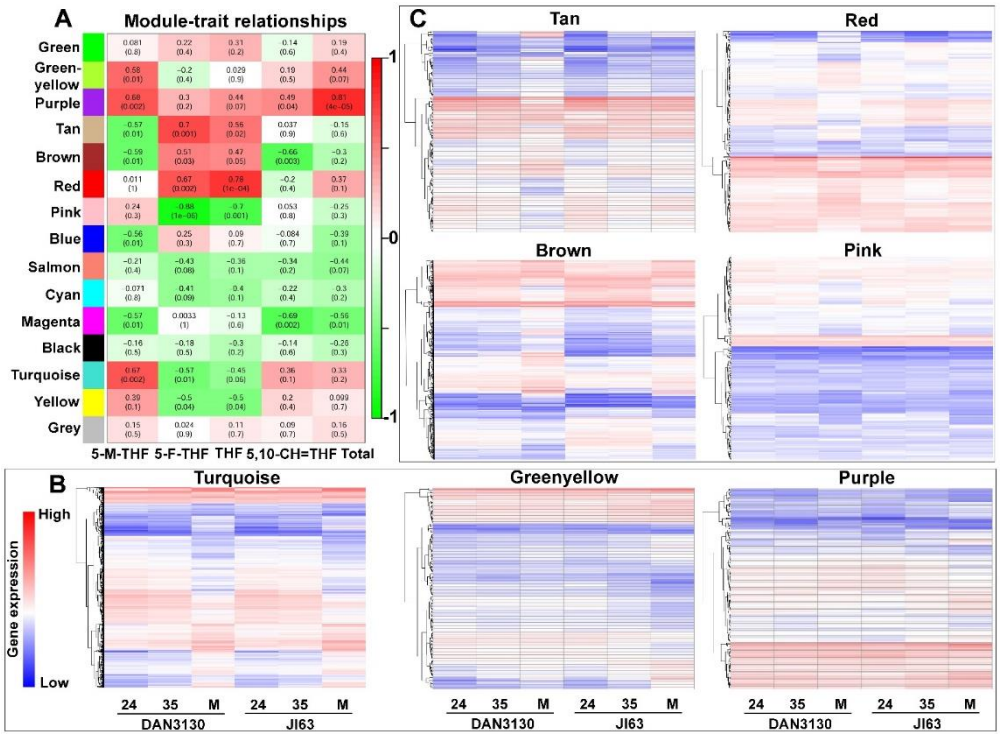



**Figure 3-8.** WGCNA analysis of kernel formation. A. Heatmap of the correlation between modules and folate derivatives (Each cell contained the correlation coefficient and corresponding P-value). B, Gene expression trends of modules related to 5-M-THF; C, Gene expression trends of modules related to 5-F-THF; 24, DAP 24; 35, DAP 35; M, mature kernel.

## *2.8 Bioinformatic analysis of the co-expressed modules*

KEGG analysis of genes from concerned modules was performed to reveal the specific function played by each co-expressed module, and the top 20 KEGG pathways were screened out. In the turquoise module, “folate biosynthesis” and “one-carbon pool by folate” pathways were significantly enriched, especially “one-carbon pool by folate” had the highest rich factor (**Figure 3-9A**). Meanwhile, similarly to the

KEGG results of development-specific and inbred-specific variation, pyruvate metabolism was identified again here with a high rich factor, which further confirms that this pathway might be related to folate accumulation (**Figure 3-5**, **3-7**, **3-9A**). The gene significance (GS) and module membership analysis results were plotted for the turquoise module, which indicated that this module was significantly relevant with 5-M-THF (Figure 3-9B). Moreover, the genes involved in the turquoise module were highly down-regulated in mature kernels according to the heatmap and bar plot with module eigengene (**Figure 3-9C**). In addition, green-yellow, purple, tan, brown, red, pink, blue, magenta, and yellow modules also showed more or less relationships with different folate derivatives. The "folate biosynthesis" pathway was also identified in the KEGG enrichment of brown and blue modules (data not shown).

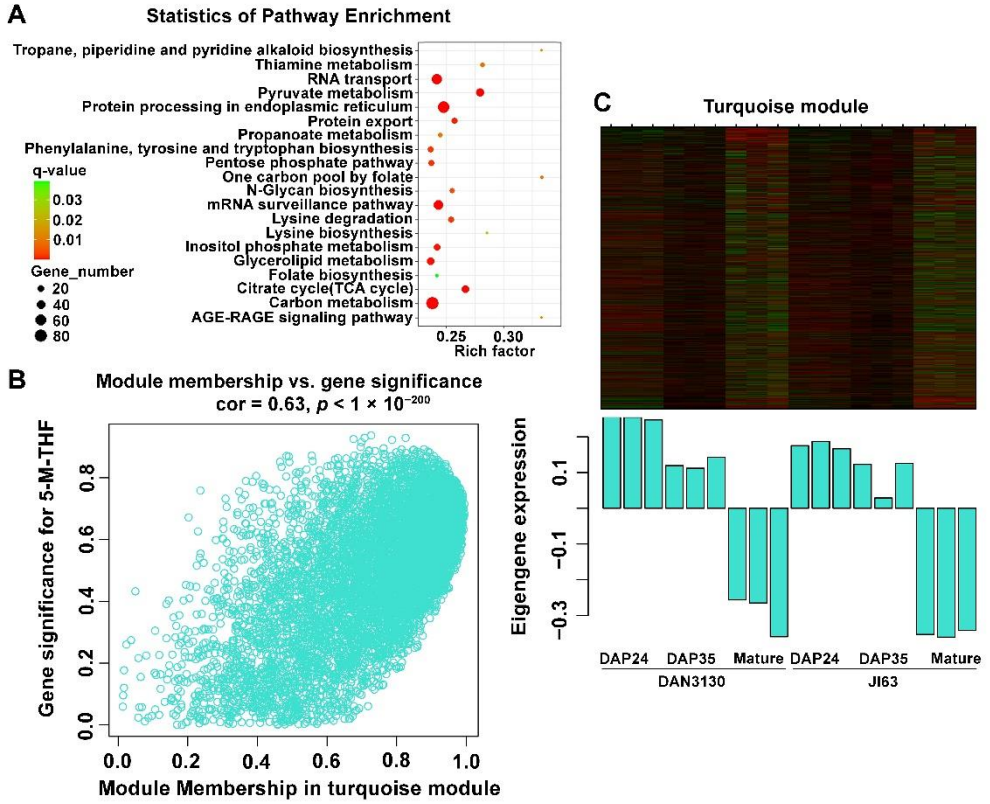



**Figure 3-9.** Bioinformatic Analysis of the turquoise module. A. KEGG enrichment of genes in the turquoise module; B. The gene significance for 5-M-THF in the turquoise module (One dot represents one gene in the turquoise module); C Expression pattern of the genes and eigengenes of the turquoise module.

After matching known folate metabolism genes with genes in each concerned module, 17 folate-related genes were found in the turquoise module, 5 were found in

the blue module, and 3 were found in the brown module (**Table S3-3**). Some of the genes showed either positive (GS ≥ 0.5, p.GS ≤ 0.05) or negative (GS ≤ −0.5, p.GS ≤ 0.05) correlation with folate phenotype, which were considered as putative key genes in folate metabolism. Firstly, most genes involved in the one-carbon metabolism process showed the opposite correlation between 5-M-THF and 5-F-THF, suggesting the complexity of folate metabolism. Secondly, these genes encode protein ADCS, ADCL, GCHI, DHNA, DHFR-TS, FPGS, GGH, GCS, SHMT, FPGS, MTHFR, MS, and DHC, which were almost involved in the entire metabolism chain from folate biosynthesis to one-carbon metabolism. Genes that participate in the folate metabolic process from THF to 5-M-THF, such as *GCST*, *SHMT7-2*, *MTHFR1*, and *MS2*, as well as those that control the polyglutamate tail length (such as *GGH* and *FPGS2*) showed the positive correlation with 5-M-THF (**Figure 3-10**, **Table S3-3**). These genes consistently showed downregulated expression levels from DAP 35 to mature kernel. For 5-F-THF, most genes had a negative correlation. In contrast, gene *ADCS* in the ρ-ABA branch, gene *DHNA2* in the pteridine branch, and gene *SHMT7-1*, which might be involved in 5-F-THF biosynthesis, were positively associated, and these three genes obviously had expressional variation between the two inbred lines (**Figure 3-10**, **Table S3-3**). For the other modules, those with the absolute value of correlation ≥ 0.7, p-value ≤ 0.05, such as the purple, tan, red, and pink modules, were screened out for analyzing hub genes. The larger weighted value indicates the stronger proof that a gene is a trait-associated hub gene (Langfelder and Horvath 2008, van Dam, Vosa et al. 2018). The networks of the top 200 edges ranked with weighted values were sorted; then, those genes with the edge numbers ≥ 15 were selected as highly connected intramodular hub genes (**Table S3-4**). Four hub genes were identified in the purple module, two of them encoding the 60S ribosomal protein, which functions in rRNA binding. In the tan module, three hub genes were identified, one belonging to the folate transporter family and another with GTPase activity. Five hub genes were found in the red module, three of which function in metal ion binding. In the pink module, six hub genes were identified: two of them related to the ribosome, one had GTPase activity, one with RNA binding activity, and one with serine/threonine kinase activity. Above all, genes related to ribosomal proteins and folate/biopterin transporters might be related to the folate concentration.

## *2.9 The expression analysis of folate metabolism genes*

As mentioned above, those genes associated with 5-M-THF and 5-F-THF from module analysis results are listed in **Table S3-5**. JI63 had higher contents of 5-M-THF and 5-F-THF than DAN3130 on DAP 24, while it was the opposite on mature kernels. For immature DAP 24, 5-M-THF was the major folate derivatives that accounts for 85%–90% of total folate, *GGH* and *FPGS2*, which showed the highest positive correlation (GS = 0.62) with 5-M-THF, which had significantly higher gene expression in JI63 than DAN3130. For mature kernels, the proportion of 5-F-THF had sharply increased from 4%–8% to 34%–36%; *DHNA2* and *SHMT7-1*, which both showed a highly positive correlation with 5-F-THF (GS = 0.63 and 0.64, respectively), had higher FPKM value in DAN3130 than JI63.

The RNA-seq data were verified by qPCR using eight representative DEGs from the folate metabolism pathway. The expression level profiles of these eight genes were similar as determined by RNA-seq and qPCR, and the two methods showed high consistency (**Figure S3-3**).

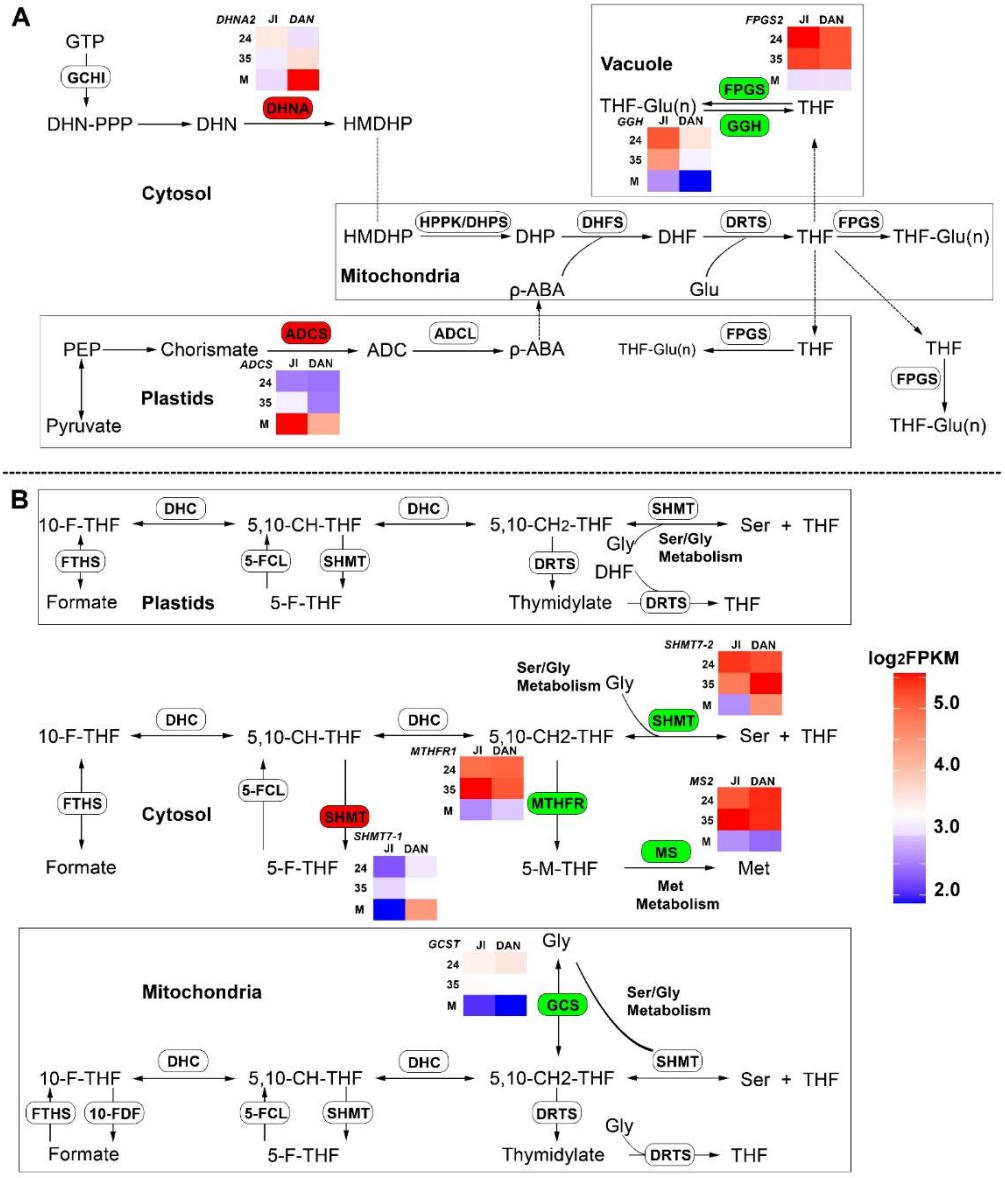



**Figure 3-10.** Schematic representation of the folate and one-carbon metabolic reactions in plants. A. Folate biosyntheis pathway; B. One-carbon metabolism pathway. Green, gene

shows positive correlation with 5-M-THF; Red, gene shows positive correlation with 5-F-THF. The panel of heatmaps show the gene expression variation. JI, JI63; DAN, DAN3130; 24, DAP 24; 35, DAP35; M, mature kernel. Chemical compounds: ADC: Aminodeoxychorismate; DHF: Dihydrofolate; DHN: Dihydroneopterin; DHP: Dihydropteroate; GTP: Guanosine-5'-triphosphate; HMDHP: 6-Hydroxymethyldihydropterin; PEP: Phosphoenolpyruvate; ρ-ABA: Para-aminobenzoate; THF: Tetrahydrofolate; THF-Glu(n): Tetrahydrofolate polyglutamate; 5,10-$CH_2$-THF: 5,10-Methylene-THF; 5,10-CH=THF: 5,10-Methenyl-THF; 10-F-THF: 10-Formyl tetrahydrofolate; Ser: Serine; Gly: Glycine; Met: Methionine. Enzymes: DHC: 5,10-Methenyl-THF cyclohydrolase/5,10-methylene-THF dehydrogenase; DHFS: Dihydrofolate synthase; DHNA: Dihydro-neopterin aldolase; DRTS: Dihydrofolate reductase-thymidylate synthase; FPGS: Folylpolyglutamate synthetase; FTHS: 10-Formyltetrahydrofolate synthetase; GCHI: GTP cyclohydrolase I; GCS: Glycine cleavage system; GGH: γ-Glutamyl hydrolase; HPPK/DHPS: Pyrophosphokinase/dihydropteroate synthase; MS: Methionine synthase; MTHFR: Methylenetetrahydrofolate reductase; SHMT: Serine hydroxymethyltransferase; 5-FCL: 5-formyl THF cycloligase; 10-FDF: 10-formyl THF deformylase.

# 3 Discussion

The development of maize kernels includes early development, differentiation, periods of mitosis and endoreduplication, accumulation of storage compounds, and maturation (Sabelli and Larkins 2009). It spans the early stage (DAP 1 to 11), filling stage (DAP 12 to 39), and desiccation stage (DAP 40 to 70). During the filling stage, sucrose and amino acids are converted to starch and storage proteins. The genes related to amino acid biosynthesis accumulate to their optimal level from DAP 10 (Lee, Williams et al. 2002), starch accumulation peaks at DAP 15 and then remains steady (Liu, Fu et al. 2008, Yan, Pan et al. 2009), and oil concentration reaches its peak level on DAP 30 (Zhang, Hong et al. 2019). At the end of seed filling, seed maturation starts with the progressive loss of water (Leprince, Pellizzaro et al. 2017). The pyruvate metabolism and glutamate metabolism were identified from the development-specific folate variation in kernel filling stage (DAP 24 to DAP 35) and desiccation stage (DAP 35 to mature kernel) in DAN3130 and JI63 (**Figure 3-4A**). Pyruvate is the end-product of glycolysis and a major substrate for oxidative metabolism, and it is converted into phosphoenolpyruvate, an upstream precursor for chorismate biosynthesis (Herrmann and Weaver 1999, Wulfert, Schilasky et al. 2020). In Arabidopsis (*Arabidopsis thaliana*) and potato, pyruvate might be degraded to formate by pyruvate formate-lyase to provide the source of one-carbon source for folate regulation (Bykova, Stensballe et al. 2003). Glutamate is one of the most abundant amino acids that play an essential role in folate biosynthesis as glutamate tail, and the number of glutamate tails has been known as a key factor of folate stability (Shane 1989). The Venn diagram of common genes, which are found in alanine/aspartate/glutamate metabolism from development-variation analysis and WGCNA analysis, showed that three of five common genes are involved in glutamate metabolism (Figure S3-4). The transcription level of these three genes increased from DAP 24 to DAP 35. We hypothesized that the more active transcription of these genes

causes a reduction in the glutamate level, resulting in decreased folate from DAP 24 to DAP 35.

The accumulation pattern of folate derivatives differed from DAP 35 to mature kernels in the two lines, suggesting that some specific pathways could affect 5-M-THF and 5-F-THF separately during this stage. As shown in **Figure 3-4B**, acyltransferase activity term and methyltransferase activity term were found in DAN3130 and JI63, respectively, during DAP 35 to mature kernel, which refers to the 5-F-THF increase in DAN3130 and 5-M-THF decrease in JI63. A total of 55 genes were identified in terms of acyltransferase activity. Among them, 11 genes belonged to the Acyl-CoA N-acetyltransferase (NAT) superfamily protein/GNAT-transcription factor family (data not shown). In plants, folates are degraded into ρ-aminobenzoylglutamate (ρ-ABAGlu) in vivo, ρ-ABAGlu and its polyglutamates are hydrolyzed into glutamate and ρ-ABA (Bozzo, Basset et al. 2008). NAT type 1 (NAT1) in humans and Nat2 in mice reportedly use the ρ-ABAGlu as the substrate (Upton, Smelt et al. 2000). Increasing the ability to acetylate ρ-ABAGlu enhances folate metabolism and demonstrates that NAT regulates folate levels by acetylating ρ-ABAGlu in animals (Cao, Strnatka et al. 2010), while whether NAT in plants is related to ρ-ABAGlu remains unknown. Aminoacyl-tRNA, which is formylated to form THF to participate in one-carbon metabolism (Fox and Stover 2008), also is acetylated by GNAT toxins in Arabidopsis (Ovchinnikov, Bikmetov et al. 2020). Here, genes belonging to the GNAT-transcription factor family were significantly enriched, while more evidence is required to validate if GNAT contributes to folate accumulation in maize. There are a total of 28 genes in terms of N-methyltransferase activity, among which 17 genes encoded histone-lysine N-methyltransferase (data not shown). In plants, histone-lysine methyltransferases catalyze the transfer of the methyl group from S-adenosylmethionine (SAM), which is generated from folate and methionine cycles (Couture, Hauk et al. 2006). In Arabidopsis, the loss function of ATMS1 causes an accumulation of Hcy and SAH, leading to reduced SAM accessibility to histone methyltransferases (Yan, Ma et al. 2019). In animals, the loss of histone H3 lysine 4 trimethylation at promoters has been linked to decreases in Transcription Factor II D (TFIID) (Mentch and Locasale 2016). Taken together, the results demonstrate that the variation in folate accumulation during the late stage of maize kernel development could be mainly controlled by circumjacent pathways of folate, especially pyruvate metabolism and glutamate metabolism. Genes belonging to the GNAT-transcription factor family and histone-lysine N-methyltransferase family might also be associated with folate metabolism.

The folate biosynthesis pathway and one-carbon pool of folate were specially found in the KEGG results of the inbred-specific folate variation. WGCNA analysis showed that those genes belonging to folate biosynthesis (pteridine branch and ρ-ABA branch) and the one-carbon metabolism process had highly positive/negative correlations to folate phenotype. In folate biosynthesis, *ADCS* and *DHNA2* showed a positive correlation with 5-F-THF and THF, confirming the plant folate biofortification strategy by engineering of pteridine and ρ-ABA branches (Diaz de la Garza, Quinlivan et al. 2004, Bekaert, Storozhenko et al. 2008, Strobbe and Van Der Straeten 2017).

For example, ADCS has been applied in Arabidopsis, rice, potato, tomato, maize, and Mexican common bean (*Phaseolus vulgaris*) to increase folate accumulation (Diaz de la Garza, Quinlivan et al. 2004, Hossain, Rosenberg et al. 2004, Diaz de la Garza, Gregory et al. 2007, Storozhenko, De Brouwer et al. 2007, Nunes, Kalkmann et al. 2009, Blancquaert, Storozhenko et al. 2013, Liang, Wang et al. 2019). Overexpression of the heterologous *Escherichia. coli* DHNA (*EcDHNA*) in tobacco (*Nicotiana tabacum*) enhances the folate accumulation (Pandey, Kumar et al. 2017). DRTS-3 and DRTS-4 showed negative correlations with 5-F-THF and THF, which is consistent with the findings in Arabidopsis: *THY3* (*DRTS-3*) inhibits the expression of its two homolog genes and fine-tunes the folate concentration during development. Overexpression of the *THY3* results in a reduced folate concentration (Gorelova, De Lepeleire et al. 2017). Furthermore, genes involved in the pteridine and ρ-ABA branches, such as *ADCS*, *GCHI-1* and *DHNA1*, were negatively correlated with 5-M-THF, suggesting that the negative feedback control might exist between 5-M-THF and folate biosynthesis genes. The ADCS activity is inhibited by DHF *in vitro*, but whether this is also the case in vivo is unclear (Sahr, Ravanel et al. 2006), and the expression of GCHI and ADCS varies significantly in tomato fruit (Basset, Quinlivan et al. 2002, Basset, Quinlivan et al. 2004). Notably, genes involved in the interconversion of monoglutamate-folate and polyglutamate-folate, such as *FPGS2* and *GGH*, had a significant correlation with folate content. In Arabidopsis and tomato, FPGS and GGH play an essential function in folate homeostasis. Indeed, knock-out of the mitochondrial FPGS or overexpression of the GGH could reduce 40–45% of the total folate in Arabidopsis leaves, and the reduction of Arabidopsis GGH enzyme activity resulted in 30% increase of total folate (Akhtar, Orsomando et al. 2010, Mehrshahi, Gonzalez-Jorge et al. 2010). The transcript level of *GGH1*, which removes the polyglutamyl tail from folates, is an indicator of folate content in potato tubers (Robinson, Garcia Salinas et al. 2019). The *GGH* and *FPGS2*, which are related to glutamate utilization, identified in WGCNA analysis, were consistent with the glutamate pathway identified in the KEGG results.

In addition, the expression level of these genes that participate in one-carbon metabolism are classed into serine/THF/5-M-THF cycle (*SHMT*, *GCS*, *MTHFR*, and *MS*) that show a positive correlation with folate accumulation. SHMT is involved in glycine oxidation, photorespiration, and one-carbon metabolism that produces 5,10-CH=THF from THF. In Saccharomyces cerevisiae, the stimulation of SHMT2 expression leads to folate accumulation (Saint-Marc, Hurlimann et al. 2015). The glycine cleavage system catalyzes the THF-dependent catabolism of glycine into 5,10-CH=THF to join the folate cycle, and the P protein (GCSP) binds the alpha-amino group of glycine via its pyridoxal-5'-phosphate cofactor (Kikuchi, Motokawa et al. 2008). A novel rice tetrahydrofolate cycle mutant hpa1 in which *HPA1* (encoding a tetrahydrofolate deformylase) is mutated shows the variation in THF, 5-F-THF, and 10-F-THF as well as the downregulation of *SHMT* and *GCS* (Xiong, Dong et al. 2020). *MTHFR* and *MS* directly participate in 5-M-THF metabolism. Notably, these two genes were identified in this study **(Figure 3-10**, **Table S3-3**), reflecting the connection between folate accumulation and conversion of homocysteine into

methionine. The importance of folate polyglutamylation in methionine biosynthesis has been confirmed. The metabolic profiling of folate-deficient cells shows that the SAM and methionine pools decline during the initial period of folate starvation in Arabidopsis, indicating that methionine synthesis is mainly associated with the perturbation of the folates pool (Loizeau, Gambonnet et al. 2007). *SHMT* and *GCS* also participate in the interconversion of serine and glycine, which is consistent with the identification of the Ser/Gly metabolism pathway in the KEGG analysis (**Figure 3-5**, **Figure 3-10**, **Table S3-3**). Moreover, one of the genes encoding the folate/biopterin transporter (FBT) protein was screened out as a hub gene from modules analysis (**Table S3-4**). In Leishmania, FBT family members are shown experimentally to transport folate, but their functions in maize are unknown (Ouameur, Girard et al. 2008).

The disturbance in metabolisms of either the pyruvate or amino acids may result from the varied folate metabolism at transcript levels, which is in accordance with the previous studies. For example, pyruvate orthophosphate dikinase (PPDK) is an important enzyme participating in pyruvate metabolism that catalyzes phosphoenolpyruvate regeneration. The dysfunction of PPDK in maize resulted in a dramatical increase in expression of cytosolic glutamine synthetase (*GS1*) and asparagine synthetase (*ASN1*), and eight folate-related genes showed differential expression patterns in leaves compared with wild type (**Table S3-6**) (Zhang, Giuliani et al. 2018). The *GS1*, *ASN*, and seven of eight folate-related genes were also identified in this study, and they showed strong correlation (either positive or negative) with folate derivatives, indicating the connection between pyruvate and folate metabolisms (**Figure S3-4**, **Table S3-3**, **Table S3-6**). Similar connections were also identified between amino acid and folate metabolisms. For instance, *O2* and *O16* encode the transcription factors of the basic leucine zipper family, and they have been used to improve the amino acid quality of maize grains (Wang, Dai et al. 2019). The kernels of double recessive mutant *o2o2o16o16* contained increased Gly as well as reduced Glu and Met. Coincidently, a total of 24 folate-related genes showed significant variation of expression between mutant and wild type, and 21 of them were also found in our WGCNA module analysis, including the key folate genes *ADCS*, *DHNA2*, *GGH*, *SHMT7-1*, *SHMT7-2*, and *GCST* (**Table S3-3**, **Table S3-7**) (Wang, Dai et al. 2019). Similarly, studies in tomato and millet (*Setaria italica*) also demonstrated that folate accumulation in different cultivars was closely related to the genes involved in pteridine branch, ρ-ABA branch, folate degradation, and serine/THF/5-M-THF cycle, such as *SiDHNA2*, *SiGGH*, *SiSHMT2*, *SiSHMT3*, *SiMTHFR*, *SiAMT* (known as *GCST*), *SlDHNA*, and *SlADCS* (Waller, Akhtar et al. 2010, Hou, Man et al. 2022).

In conclusion, the comparative transcriptome analysis in this study provides the complexity of folate metabolism and creates awareness of the interaction among metabolism mentation above during the futural maize breeding.

# 4 Materials and methods

## *4.1 Plant material and folate measurement*

DAN3130 and JI63 inbred lines are originated from China, belonging to the NSS subpopulation, JI63 with pedigree being (127-32 × Tie84) × (Wei24 × Wei20) and DAN3130 with pedigree being American hybrid P78599 (Yang, Gao et al. 2011). They were grown at Langfang, Hebei province, in the summer of 2018. Fresh kernel samples were collected 24 and 35 days after pollination and then frozen in liquid nitrogen immediately. Mature kernel samples were harvested after all the plants turned yellow. Three biological replicates of each stage were harvested. The folates exaction and measurement were repeated three times in each replicate. The methods of folate extraction and measurement were the same as in the previous study (Liang, Wang et al. 2019).

## *4.2 RNA extraction and library preparation*

The total RNA of DAP 24, DAP 35, and mature kernels were extracted using a BAIXU Maize Kernel Total RNA Extraction Kit (Beijing, China). To eliminate any re-sidual genomic DNA, the total RNA was treated with RNase-free DNase I (New Eng-land Biolabs, Ipswich, Massachusetts, USA), and the concentration of RNA was deter-mined using a Nanodrop-2000. The total RNA was used to synthesize first-strand complementary DNA (cDNA) using the RevertAid First Strand cDNA Synthesis kit (Fermentas, Waltham, Massachusetts, USA).

## *4.3 Raw read filtering and assembly*

RNA-Seq library preparation and sequencing were carried out by BEIJING BAIXU BIOTECH CO., LTD, Beijing, China. All libraries were sequenced using the Sanger method by Illumina 1.9 Platform. The adapters, low-quality reads, and the reads containing poly-N were removed to obtain clean reads. Then, the clean reads were mapped to the Zea mays B73 genome by using Hisat software (v2.1.0, http://daehwankimlab.github.io/hisat2/) with setting non-strand-specific and other de-fault parameters. Afterwards, the clean reads were assembled using Samtools software (http://www.htslib.org/), Cufflink software (v2.2.1, http://cole-trapnell-lab.github.io/cufflinks/) and Cuffcompare software (v2.2.1, http://cole-trapnell-lab.github.io/cufflinks/cuffcompare/).

## *4.4 Gene annotation and DEG analysis*

The sequences were aligned to *zea mays* B73 genome v4.40 (data download from: ftp://ftp.ensemblgenomes.org/pub/plants/release-40/gtf/zea_mays/Zea_mays.AGPv4.40.chr.gtf.gz Accessed date: 10.02.2020). Genes were annotated based on the maize B73 genome v4.40 annotation in Ensemblgenomes (data download from: ftp://ftp.ensemblgenomes.org/pub/plants/release-40/gff3/zea_mays/Zea_mays.AGPv4.40.chr.gff3.gz Accessed date: 10.02.2020). The rest of the unmatched transcripts were annotated using Trinotate software (v3.2.0, https://rnabio.org/module-07-trinotate/0007/02/01/Trinotate/, including TransDecoder v5.5.0, HMMER v3.1b2, Blast v2.9.0, SignalP v4.1, TMHMM v2.0c, and RNAmmer v1.2), Uniprot database (https://www.uniprot.org/ Accessed date: 29.03.2020), and NCBI Ref-Seq non-redundant proteins databases (data download

from ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz Accessed date: 29.03.2020). Htseq-count software (v0.11.2) was used for counting reads, and DESeq-2 software was used for standardization and DEG analysis. The FPKM (Fragments Per Kilobase of exon model per Million mapped reads) of each gene was calculated according to its length and the mapped read numbers. |Log2FC (Fold Change)| ≥ 1 and p-value <0.05 were used as the threshold to screen DEGs.

## *4.5 Bioinformatics analysis*

GO (GO, http://www.geneontology.org / Accessed date: 21.04.2020) and functional enrichment analysis were conducted on all DEGs using Agrigod(v2) software (http://systemsbiology.cau.edu.cn/agriGOv2/index.php Accessed date: 21.04.2020). Then, all DEGs were mapped to a pathway in the KEGG database (http://www.genome.jp/kegg/pathway.html Accessed date: 21. 04. 2020) using KOBAS 3.0 software (http://kobas.cbi.pku.edu.cn/kobas3 Accessed date: 21. 04. 2020). The p-value was set as ≤ 0.05 as the threshold to judge the significance of the GO and KEGG pathway enrichment analyses. Gene co-expression networks were performed using the WGCNA package in the R software. The steps were followed as below:
https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/ Accessed date: 05, 05, 2021 (Zhang and Horvath 2005). A total of 13955 genes were filtered by mean expression (mean normalized counts > 10) and then log-transformed using log2(x + 1) for WGCNA analysis.

## *4.6 Real-time quantitative PCR analysis*

qPCR was performed in ABI QuantStudio 6 real-time PCR system using TransStart Top Green qPCR SuperMix (TransGen Biotech, Beijing, China). The cDNAs were made from three samples, and all reactions were performed in quadruplicate. PCR conditions were as follows: 95 °C for 30 s, 40 cycles of 95 °C for 5 s, and 60 °C for 34 s. The primers of selected genes for qPCR are listed in **Table S8**. *GAPDH* (*Zm00001d049641* encodes the glyceraldehyde-3-phosphate dehydrogenase) was used as the reference gene to normalize the target gene expression, which was calculated using the relative quantization method ($2^{-\Delta\Delta}$CT).

# 5 Conclusions

Two maize inbred lines, DAN3130 and JI63, showed significant folate variation and different folate accumulation patterns during the late stage of kernel formation, indicating the complexity of mechanism of folate metabolism. The results demonstrated that the folate development-specific variation from DAP 24 to mature kernels might be regulated by the circumjacent pathways of folate biosynthesis such as pyruvate metabolism, glutamate metabolism, and glycine/serine metabolism, as well as by genes belonging to the GNAT-transcription factor family and histone-lysine N-methyltransferase family. Genes involved in the two branches of folate biosynthesis, serine/THF/5-M-THF cycle and conversion of THF monoglutamate into THF

polyglutamate, could affect the folate levels between different lines. The results provide insight into the mechanism of folate accumulation and will help the futural breeding work of folate biofortification.

**Author Contributions:** Conceptualization, H.W., C.Z. and L.J.; methodology, H.W. and L.J.; software, T.L., X.W.; validation, T.L.; formal analysis, T.L., X.W; investigation, T.L., X.W.; resources, C.Z. and L.J; data curation, X.W. and H.W.; writing—original draft preparation, T.L.; writing—review and editing, T.L., X.W., S.L., H.W. H.J, C.Z. and L.J.; visualization, T.L., X.W.; supervision, H.W., C.Z. and L.J.; project administration, H.W., C.Z. and L.J; funding acquisition, C.Z., H.W. and L.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China (2021YFF1000300 to C.Z.), The National Key Research and Development Program of China (2018YFD1000702/2018YFD1000700 to H.W.), The National Natural Science Foundation of China (31870283 to L.J.), and The Agricultural Science and Technology Innovation Program (ASTIP; CAAS-ZDRW202004 to C.Z. and H.W.).

**Data Availability Statement:** The raw data were uploaded and deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) under the accession number: GSE195815

**Acknowledgments:** We thank the public laboratory of the Biotechnology Research Institute, Chinese Academy of Agricultural Sciences for providing us with access to the HPLC and triple-quadrupole MS/MS instruments and for providing technical assistance. We would like to express our gratitude to Prof. Hervé Vanderschuren of the University of Liège for his thoughtful suggestions and comments on this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

# References

Akhtar, T. A., G. Orsomando, P. Mehrshahi, A. Lara-Nunez, M. J. Bennett, J. F. Gregory, 3rd and A. D. Hanson (2010). "A central role for gamma-glutamyl hydrolases in plant folate homeostasis." Plant J 64(2): 256-266.

Asensi-Fabado, M. A. and S. Munne-Bosch (2010). "Vitamins in plants: occurrence, biosynthesis and antioxidant function." Trends Plant Sci 15(10): 582-592.

Basset, G., E. P. Quinlivan, M. J. Ziemak, R. Diaz De La Garza, M. Fischer, S. Schiffmann, A. Bacher, J. F. Gregory, 3rd and A. D. Hanson (2002). "Folate synthesis in plants: the first step of the pterin branch is mediated by a unique bimodular GTP cyclohydrolase I." Proc Natl Acad Sci U S A 99(19): 12489-12494.

Basset, G. J., E. P. Quinlivan, S. Ravanel, F. Rebeille, B. P. Nichols, K. Shinozaki, M. Seki, L. C. Adams-Phillips, J. J. Giovannoni, J. F. Gregory, 3rd and A. D. Hanson (2004). "Folate synthesis in plants: the p-aminobenzoate branch is initiated by a bifunctional PabA-PabB protein that is targeted to plastids." Proc Natl Acad Sci U S A 101(6): 1496-1501.

Basset, G. J., S. Ravanel, E. P. Quinlivan, R. White, J. J. Giovannoni, F. Rebeille, B. P. Nichols, K. Shinozaki, M. Seki, J. F. Gregory, 3rd and A. D. Hanson (2004). "Folate synthesis in plants: the last step of the p-aminobenzoate branch is catalyzed by a plastidial aminodeoxychorismate lyase." Plant J 40(4): 453-461.

Bekaert, S., S. Storozhenko, P. Mehrshahi, M. J. Bennett, W. Lambert, J. F. Gregory, 3rd, K. Schubert, J. Hugenholtz, D. Van Der Straeten and A. D. Hanson (2008). "Folate biofortification in food plants." Trends Plant Sci 13(1): 28-35.

Blancquaert, D., S. Storozhenko, J. Van Daele, C. Stove, R. G. Visser, W. Lambert and D. Van Der Straeten (2013). "Enhancing pterin and para-aminobenzoate content is not sufficient to successfully biofortify potato tubers and Arabidopsis thaliana plants with folate." J Exp Bot 64(12): 3899-3909.

Blancquaert, D., J. Van Daele, S. Strobbe, F. Kiekens, S. Storozhenko, H. De Steur, X. Gellynck, W. Lambert, C. Stove and D. Van Der Straeten (2015). "Improving folate (vitamin B9) stability in biofortified rice through metabolic engineering." Nat Biotechnol 33(10): 1076-1078.

Bozzo, G. G., G. J. Basset, V. Naponelli, A. Noiriel, J. F. Gregory, 3rd and A. D. Hanson (2008). "Characterization of the folate salvage enzyme p-aminobenzoylglutamate hydrolase in plants." Phytochemistry 69(1): 29-37.

Bykova, N. V., A. Stensballe, H. Egsgaard, O. N. Jensen and I. M. Moller (2003). "Phosphorylation of formate dehydrogenase in potato tuber mitochondria." J Biol Chem 278(28): 26021-26030.

Cao, W., D. Strnatka, C. A. McQueen, R. J. Hunter and R. P. Erickson (2010). "N-acetyltransferase 2 activity and folate levels." Life Sci 86(3-4): 103-106.

Chen, L., Z. Zhang, A. Hoshino, H. D. Zheng, M. Morley, Z. Arany and J. D. Rabinowitz (2019). "NADPH production by the oxidative pentose-phosphate pathway supports folate metabolism." Nature Metabolism 1(3): 404-415.

Collakova, E., A. Goyer, V. Naponelli, I. Krassovskaya, J. F. Gregory, 3rd, A. D. Hanson and Y. Shachar-Hill (2008). "Arabidopsis 10-formyl tetrahydrofolate deformylases are essential for photorespiration." Plant Cell 20(7): 1818-1832.

Couture, J. F., G. Hauk, M. J. Thompson, G. M. Blackburn and R. C. Trievel (2006). "Catalytic roles for carbon-oxygen hydrogen bonding in SET domain lysine methyltransferases." J Biol Chem 281(28): 19280-19287.

Danchin, A., A. Sekowska and C. You (2020). "One-carbon metabolism, folate, zinc and translation." Microb Biotechnol 13(4): 899-925.

De Lepeleire, J., S. Strobbe, J. Verstraete, D. Blancquaert, L. Ambach, R. G. F. Visser, C. Stove and D. Van Der Straeten (2018). "Folate Biofortification of Potato by Tuber-Specific Expression of Four Folate Biosynthesis Genes." Mol Plant 11(1): 175-188.

DellaPenna, D. (2005). "Progress in the dissection and manipulation of vitamin E synthesis." Trends Plant Sci 10(12): 574-579.

Diaz de la Garza, R., E. P. Quinlivan, S. M. Klaus, G. J. Basset, J. F. Gregory, 3rd and A. D. Hanson (2004). "Folate biofortification in tomatoes by engineering the pteridine branch of folate synthesis." Proc Natl Acad Sci U S A 101(38): 13720-13725.

Diaz de la Garza, R. I., J. F. Gregory, 3rd and A. D. Hanson (2007). "Folate biofortification of tomato fruit." Proc Natl Acad Sci U S A 104(10): 4218-4222.

Fan, J., J. Ye, J. J. Kamphorst, T. Shlomi, C. B. Thompson and J. D. Rabinowitz (2014). "Quantitative flux analysis reveals folate-dependent NADPH production." Nature 510(7504): 298-302.

Fox, J. T. and P. J. Stover (2008). "Folate-mediated one-carbon metabolism." Vitam Horm 79: 1-44.

Gorelova, V., L. Ambach, F. Rebeille, C. Stove and D. Van Der Straeten (2017). "Folates in Plants: Research Advances and Progress in Crop Biofortification." Front Chem 5: 21.

Gorelova, V., J. De Lepeleire, J. Van Daele, D. Pluim, C. Mei, A. Cuypers, O. Leroux, F. Rebeille, J. H. M. Schellens, D. Blancquaert, C. P. Stove and D. Van Der Straeten (2017). "Dihydrofolate Reductase/Thymidylate Synthase Fine-Tunes the Folate Status and Controls Redox Homeostasis in Plants." Plant Cell 29(11): 2831-2853.

Guo, W., T. Lian, B. Wang, J. Guan, D. Yuan, H. Wang, F. M. Safiul Azam, X. Wan, W. Wang, Q. Liang, H. Wang, J. Tu, C. Zhang and L. Jiang (2019). "Genetic mapping of folate QTLs using a segregated population in maize." J Integr Plant Biol 61(6): 675-690.

Hanson, A. D. and J. F. Gregory, 3rd (2011). "Folate biosynthesis, turnover, and transport in plants." Annu Rev Plant Biol 62: 105-125.

Herrmann, K. M. and L. M. Weaver (1999). "The Shikimate Pathway." Annu Rev Plant Physiol Plant Mol Biol 50: 473-503.

Hossain, T., I. Rosenberg, J. Selhub, G. Kishore, R. Beachy and K. Schubert (2004). "Enhancement of folates in plants through metabolic engineering." Proc Natl Acad Sci U S A 101(14): 5158-5163.

Hou, S., X. Man, B. Lian, G. Ma, Z. Sun, L. Han, L. Yan, H. Gao, W. Du, X. Wang, Y. Zhang, H. Li and Y. Han (2022). "Folate metabolic profiling and expression of folate metabolism-related genes during panicle development in foxtail millet (Setaria italica (L.) P. Beauv)." J Sci Food Agric 102(1): 268-279.

Jabrin, S., S. Ravanel, B. Gambonnet, R. Douce and F. Rebeille (2003). "One-carbon metabolism in plants. Regulation of tetrahydrofolate synthesis during germination and seedling development." Plant Physiol 131(3): 1431-1439.

Jiang, L., Y. Liu, H. Sun, Y. Han, J. Li, C. Li, W. Guo, H. Meng, S. Li, Y. Fan and C. Zhang (2013). "The mitochondrial folylpolyglutamate synthetase gene is required for nitrogen utilization during early seedling development in arabidopsis." Plant Physiol 161(2): 971-989.

Kikuchi, G., Y. Motokawa, T. Yoshida and K. Hiraga (2008). "Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia." Proc Jpn Acad Ser B Phys Biol Sci 84(7): 246-263.

Langfelder, P. and S. Horvath (2008). "WGCNA: an R package for weighted correlation network analysis." BMC Bioinformatics 9: 559.

Lee, J. M., M. E. Williams, S. V. Tingey and J. A. Rafalski (2002). "DNA array profiling of gene expression changes during maize embryo development." Funct Integr Genomics 2(1-2): 13-27.

Leprince, O., A. Pellizzaro, S. Berriri and J. Buitink (2017). "Late seed maturation: drying without dying." J Exp Bot 68(4): 827-841.

Lian, T., W. Guo, M. Chen, J. Li, Q. Liang, F. Liu, H. Meng, B. Xu, J. Chen, C. Zhang and L. Jiang (2015). "Genome-wide identification and transcriptional analysis of folate metabolism-related genes in maize kernels." BMC Plant Biol 15: 204.

Liang, Q., K. Wang, X. Liu, B. Riaz, L. Jiang, X. Wan, X. Ye and C. Zhang (2019). "Improved folate accumulation in genetically modified maize and wheat." J Exp Bot 70(5): 1539-1551.

Liu, X., J. Fu, D. Gu, W. Liu, T. Liu, Y. Peng, J. Wang and G. Wang (2008). "Genome-wide analysis of gene expression profiles during the kernel development of maize (Zea mays L.)." Genomics 91(4): 378-387.

Loizeau, K., B. Gambonnet, G. F. Zhang, G. Curien, S. Jabrin, D. Van Der Straeten, W. E. Lambert, F. Rebeille and S. Ravanel (2007). "Regulation of one-carbon metabolism in Arabidopsis: the N-terminal regulatory domain of cystathionine gamma-synthase is cleaved in response to folate starvation." Plant Physiol 145(2): 491-503.

Mehrshahi, P., S. Gonzalez-Jorge, T. A. Akhtar, J. L. Ward, A. Santoyo-Castelazo, S. E. Marcus, A. Lara-Nunez, S. Ravanel, N. D. Hawkins, M. H. Beale, D. A.

Barrett, J. P. Knox, J. F. Gregory, 3rd, A. D. Hanson, M. J. Bennett and D. Dellapenna (2010). "Functional analysis of folate polyglutamylation and its essential role in plant metabolism and development." Plant J 64(2): 267-279.

Meng, H., B. Xu, C. Zhang and L. Jiang (2017). "Arabidopsis plastidial folylpolyglutamate synthetase is required for nitrogen metabolism under nitrate-limited condition in darkness." Biochem Biophys Res Commun 482(2): 277-281.

Mentch, S. J. and J. W. Locasale (2016). "One-carbon metabolism and epigenetics: understanding the specificity." Ann N Y Acad Sci 1363: 91-98.

Mouillon, J. M., S. Ravanel, R. Douce and F. Rebeille (2002). "Folate synthesis in higher-plant mitochondria: coupling between the dihydropterin pyrophosphokinase and the dihydropteroate synthase activities." Biochem J 363(Pt 2): 313-319.

Nunes, A. C., D. C. Kalkmann and F. J. Aragao (2009). "Folate biofortification of lettuce by expression of a codon optimized chicken GTP cyclohydrolase I gene." Transgenic Res 18(5): 661-667.

Orsomando, G., R. D. de la Garza, B. J. Green, M. Peng, P. A. Rea, T. J. Ryan, J. F. Gregory, 3rd and A. D. Hanson (2005). "Plant gamma-glutamyl hydrolases and folate polyglutamates: characterization, compartmentation, and co-occurrence in vacuoles." J Biol Chem 280(32): 28877-28884.

Ouameur, A. A., I. Girard, D. Legare and M. Ouellette (2008). "Functional analysis and complex gene rearrangements of the folate/biopterin transporter (FBT) gene family in the protozoan parasite Leishmania." Mol Biochem Parasitol 162(2): 155-164.

Ovchinnikov, S. V., D. Bikmetov, A. Livenskyi, M. Serebryakova, B. Wilcox, K. Mangano, D. I. Shiriaev, I. A. Osterman, P. V. Sergiev, S. Borukhov, N. Vazquez-Laslop, A. S. Mankin, K. Severinov and S. Dubiley (2020). "Mechanism of translation inhibition by type II GNAT toxin AtaT2." Nucleic Acids Res 48(15): 8617-8625.

Palchetti, C. Z., J. Steluti, E. Verly, Jr., R. A. Pereira, R. Sichieri and D. M. L. Marchioni (2020). "Prevalence of inadequate intake of folate after mandatory fortification: results from the first National Dietary Survey in Brazil." Eur J Nutr 59(6): 2793-2803.

Pandey, D. K., A. Kumar, J. S. Rathore, N. Singh and B. Chaudhary (2017). "Recombinant overexpression of dihydroneopterin aldolase catalyst potentially regulates folate-biofortification." J Basic Microbiol 57(6): 517-524.

Ratan, S. K., K. N. Rattan, R. M. Pandey, S. Singhal, S. Kharab, M. Bala, V. Singh and A. Jhanwar (2008). "Evaluation of the levels of folate, vitamin B12, homocysteine and fluoride in the parents and the affected neonates with neural tube defect and their matched controls." Pediatr Surg Int 24(7): 803-808.

Robinson, B. R., C. Garcia Salinas, P. Ramos Parra, J. Bamberg, R. I. Diaz de la Garza and A. Goyer (2019). "Expression Levels of the γ-Glutamyl Hydrolase I Gene Predict Vitamin B9 Content in Potato Tubers." Agronomy 9(11).

Roje, S. (2007). "Vitamin B biosynthesis in plants." Phytochemistry 68(14): 1904-1921.

Sabelli, P. A. and B. A. Larkins (2009). "The contribution of cell cycle regulation to endosperm development." Sex Plant Reprod 22(4): 207-219.

Sahr, T., S. Ravanel, G. Basset, B. P. Nichols, A. D. Hanson and F. Rebeille (2006). "Folate synthesis in plants: purification, kinetic properties, and inhibition of aminodeoxychorismate synthase." Biochem J 396(1): 157-162.

Saint-Marc, C., H. C. Hurlimann, B. Daignan-Fornier and B. Pinson (2015). "Serine hydroxymethyltransferase: a key player connecting purine, folate and methionine metabolism in Saccharomyces cerevisiae." Curr Genet 61(4): 633-640.

Shane, B. (1989). "Folylpolyglutamate synthesis and role in the regulation of one-carbon metabolism." Vitam Horm 45: 263-335.

Shohag, M. J. I., Y. Wei, J. Zhang, Y. Feng, M. Rychlik, Z. He and X. Yang (2020). "Genetic and physiological regulation of folate in pak choi (Brassica rapa subsp. Chinensis) germplasm." J Exp Bot 71(16): 4914-4929.

Storozhenko, S., V. De Brouwer, M. Volckaert, O. Navarrete, D. Blancquaert, G. F. Zhang, W. Lambert and D. Van Der Straeten (2007). "Folate fortification of rice by metabolic engineering." Nat Biotechnol 25(11): 1277-1279.

Strobbe, S. and D. Van Der Straeten (2017). "Folate biofortification in food crops." Curr Opin Biotechnol 44: 202-211.

Upton, A., V. Smelt, A. Mushtaq, R. Aplin, N. Johnson, H. Mardon and E. Sim (2000). "Placental arylamine N-acetyltransferase type 1: potential contributory source of urinary folate catabolite p-acetamidobenzoylglutamate during pregnancy." Biochim Biophys Acta 1524(2-3): 143-148.

van Dam, S., U. Vosa, A. van der Graaf, L. Franke and J. P. de Magalhaes (2018). "Gene co-expression analysis for functional classification and gene-disease predictions." Brief Bioinform 19(4): 575-592.

Verberne, M. C., K. Sansuk, J. F. Bol, H. J. Linthorst and R. Verpoorte (2007). "Vitamin K1 accumulation in tobacco plants overexpressing bacterial genes involved in the biosynthesis of salicylic acid." J Biotechnol 128(1): 72-79.

Waller, J. C., T. A. Akhtar, A. Lara-Nunez, J. F. Gregory, 3rd, R. P. McQuinn, J. J. Giovannoni and A. D. Hanson (2010). "Developmental and feedforward control of the expression of folate biosynthesis genes in tomato fruit." Mol Plant 3(1): 66-77.

Wang, W., Y. Dai, M. Wang, W. Yang and D. Zhao (2019). "Transcriptome Dynamics of Double Recessive Mutant, o2o2o16o16, Reveals the Transcriptional Mechanisms in the Increase of Its Lysine and Tryptophan Content in Maize." Genes (Basel) 10(4).

Wulfert, S., S. Schilasky and S. Krueger (2020). "Transcriptional and Biochemical Characterization of Cytosolic Pyruvate Kinases in Arabidopsis thaliana." Plants (Basel) 9(3).

Xiong, E., G. Dong, F. Chen, C. Zhang, S. Li, Y. Zhang, J. I. Shohag, X. Yang, Y. Zhou, Q. Qian, L. Wu and Y. Yu (2020). "Formyl tetrahydrofolate deformylase affects hydrogen peroxide accumulation and leaf senescence by regulating the folate status and redox homeostasis in rice." Sci China Life Sci.

Yan, H. B., X. X. Pan, H. W. Jiang and G. J. Wu (2009). "Comparison of the starch synthesis genes between maize and rice: copies, chromosome location and expression divergence." Theor Appl Genet 119(5): 815-825.

Yan, X., L. Ma, H. Pang, P. Wang, L. Liu, Y. Cheng, J. Cheng, Y. Guo and Q. Li (2019). "METHIONINE SYNTHASE1 Is Involved in Chromatin Silencing by Maintaining DNA and Histone Methylation." Plant Physiol 181(1): 249-261.

Yang, X., S. Gao, S. Xu, Z. Zhang, B. M. Prasanna, L. Li, J. Li and J. Yan (2011). "Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize." Molecular Breeding 28(4): 511-526.

Zhang, B. and S. Horvath (2005). "A general framework for weighted gene co-expression network analysis." Stat Appl Genet Mol Biol 4: Article17.

Zhang, X., M. Hong, H. Wan, L. Luo, Z. Yu and R. Guo (2019). "Identification of Key Genes Involved in Embryo Development and Differential Oil Accumulation in Two Contrasting Maize Genotypes." Genes (Basel) 10(12).

Zhang, Y., R. Giuliani, Y. Zhang, Y. Zhang, W. L. Araujo, B. Wang, P. Liu, Q. Sun, A. Cousins, G. Edwards, A. Fernie, T. P. Brutnell and P. Li (2018). "Characterization of maize leaf pyruvate orthophosphate dikinase using high throughput sequencing." J Integr Plant Biol 60(8): 670-690.

# *Supplementary Materials*

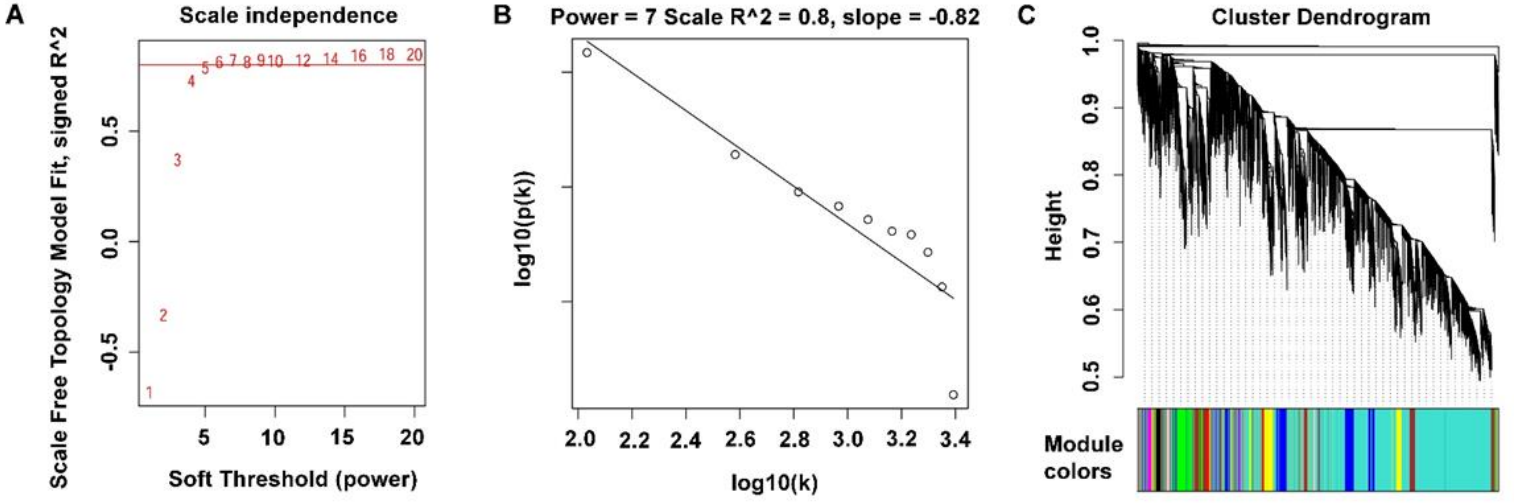



**Figure S3-1.** WGCNA analysis. A. Soft-thresholding power analysis; B. Scale free topology when soft-thresholding power β = 7; C. Clustering dendrogram of genes, with dissimilarity based on the topological overlap, together with assigned module colors

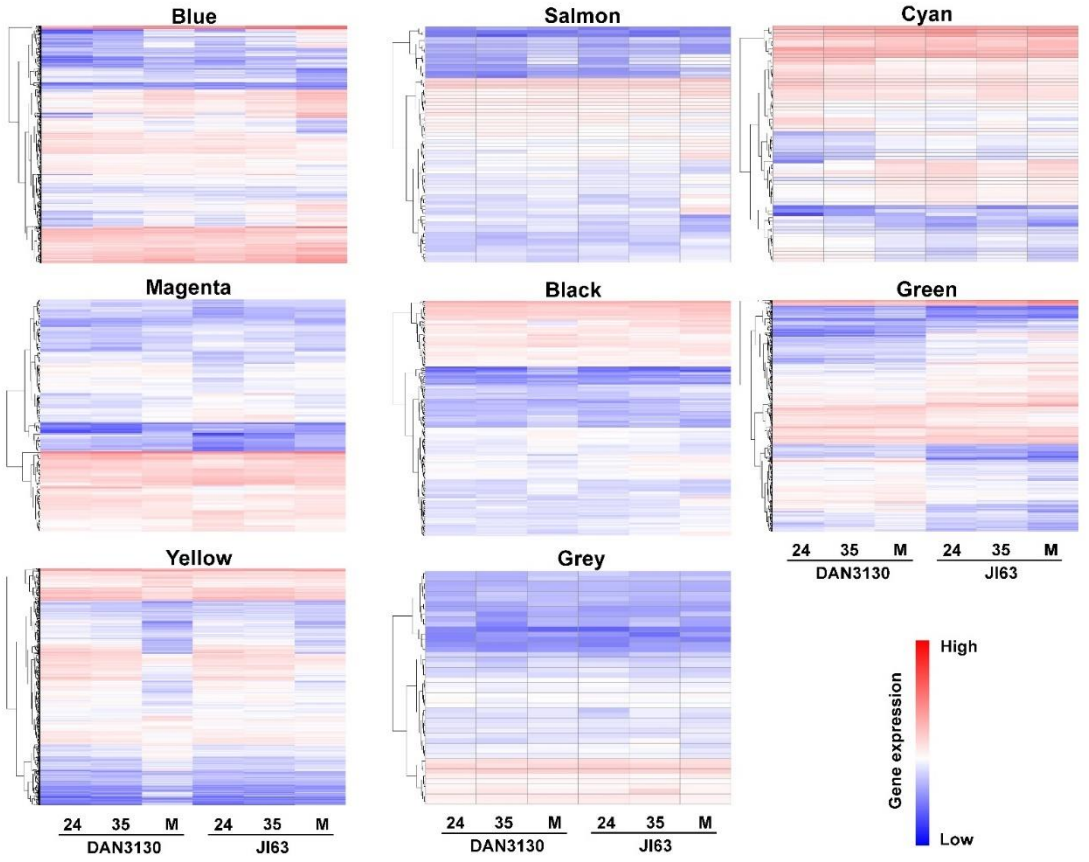



**Figure S3-2.** The gene expression heatmap of modules, which has no correlation with folate, from WGCNA results.

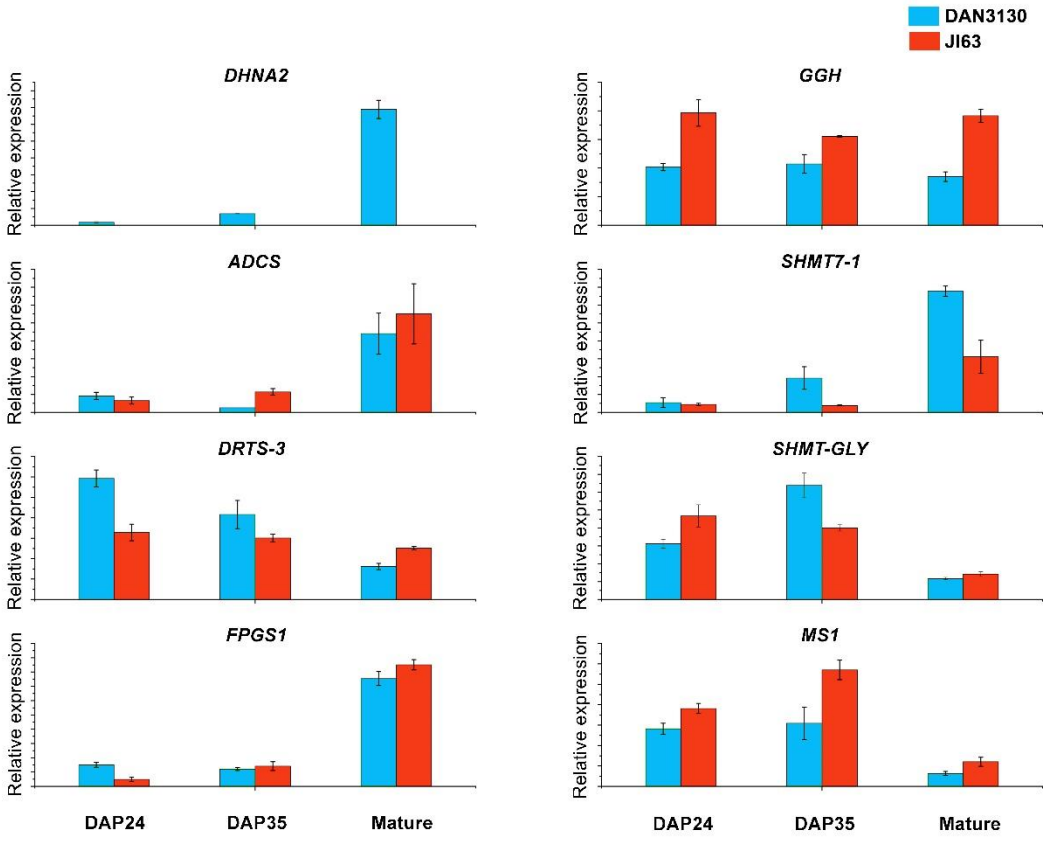
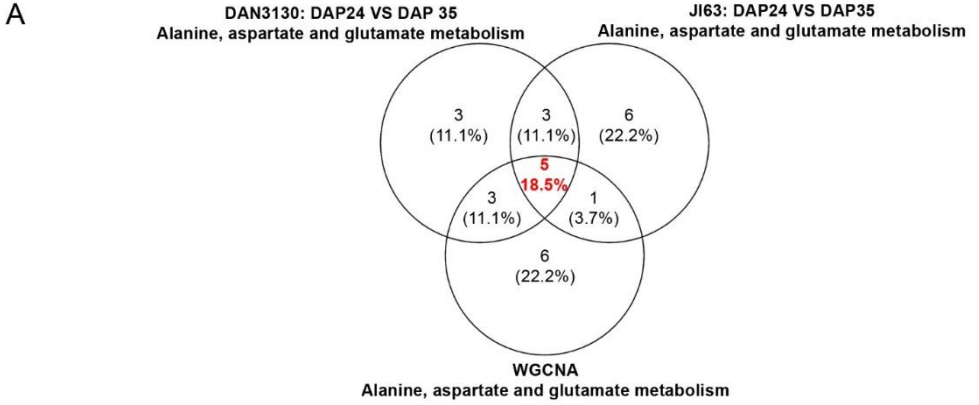



**Figure S3-3.** qPCR verification of representative DEGs from the folate metabolism pathway

A




B FPKM value of five common genes in alanine, aspartate and glutamate metabolism pathway

| Gene | Description | DAN3130 | | | JI63 | | |
|---|---|---|---|---|---|---|---|
| | | DAP 24 | DAP 35 | Matural kernel | DAP 24 | DAP 35 | Matural kernel |
| *Zm00001d044608* | Asparagine synthetase 2 | 33.88±7.37 | 235.40±36.28 | 453.36±140.30 | 61.68±9.85 | 217.41±37.32 | 373.29±14.96 |
| *Zm00001d028750* | Asparagine synthetase 3 | 3.47±1.51 | 8.30±9.49 | 83.87±19.03 | 4.83±1.46 | 19.34±2.19 | 154.75±22.47 |
| *Zm00001d031749* | Glutamate decarboxylase | 28.09±0.49 | 115.13±11.43 | 276.49±85.74 | 56.77±3.52 | 235.72±44.64 | 494.40±41.46 |
| *Zm00001d051804* | Glutamine synthetase5 | 4.43±0.24 | 11.18±2.35 | 26.75±4.31 | 7.72±1.27 | 17.93±5.61 | 29.09±8.15 |
| *Zm00001d010197* | Glutamine 5-phosphoribosylpyrophosphate amidotransferase | 4.52±0.90 | 11.26±0.54 | 115.59±45.78 | 4.75±1.63 | 14.38±6.91 | 120.32±4.62 |

**Figure S3-4.** The venn diagram of common genes in glutamate metabolism. A. The venn diagram of common genes in glutamate metabolism in alanine/aspartate/glutamate metabolism from development-variation analysis and WGCNA analysis. B. The FPKM value of five common genes in alanine/aspartate/glutamate metabolism.

**Table S3-1.** Folate profiling of maize kernel during late stages of development

| | DAP 24 (nmol/g FW) | | DAP 35 (nmol/g FW) | | Mature kernel (nmol/g DW) | |
|---|---|---|---|---|---|---|
| Folate derivatives | DAN3130 | JI63 | DAN3130 | JI63 | DAN3130 | JI63 |
| 5-M-THF | 0.89±0.08 | 1.35±0.15 | 0.48±0.03 | 0.72±0.08 | 0.54±0.02 | 0.22±0.02 |
| 5-F-THF | 0.04±0.01 | 0.12±0.02 | 0.03±0.01 | 0.16±0.01 | 0.34±0.02 | 0.16±0.001 |
| THF | 0.03±0.005 | 0.04±0.01 | 0.02±0.002 | 0.04±0.007 | 0.07±0.004 | 0.03±0.004 |
| 5,10-CH=THF | 0.03±0.003 | 0.06±0.01 | 0.03±0.005 | 0.08±0.001 | 0.04±0.005 | 0.02±0.005 |

**Table S3-2.** Overview of transcriptome sequencing data

| Sample ID | Rep | Clean reads | Q20 bases | Q30 bases | GC content | Mapping reads | Mapped reads with unique loci | Mapping ratio |
|---|---|---|---|---|---|---|---|---|
| DAN3130 DAP 24 | 1 | 74,987,086 | 95.05% | 88.88% | 51.51% | 64,521,241 | 54,258,276 | 86.04% |
| | 2 | 174,991,970 | 95.04% | 88.88% | 51.24% | 149,238,661 | 123,619,968 | 85.28% |
| | 3 | 48,358,824 | 93.64% | 87.34% | 45.67% | 36,727,718 | 30,061,383 | 75.95% |
| DAN3130 DAP 35 | 1 | 46,343,382 | 92.68% | 86.46% | 46.32% | 31,747,792 | 26,817,158 | 68.51% |
| | 2 | 44,861,502 | 93.64% | 87.37% | 47.87% | 33,448,957 | 27,666,652 | 74.56% |
| | 3 | 87,136,232 | 94.91% | 89.10% | 51.86% | 71,739,096 | 60,279,115 | 82.33% |
| DAN3130 Mature kernels | 1 | 43,319,674 | 93.62% | 87.93% | 48.30% | 31,086,099 | 17,759,180 | 71.76% |
| | 2 | 76,918,438 | 92.27% | 85.37% | 46.30% | 52,517,640 | 17,921,475 | 68.28% |
| | 3 | 117,938,336 | 96.78% | 92.45% | 59.38% | 94,704,697 | 56,161,081 | 80.30% |
| JI63 DAP 24 | 1 | 90,251,840 | 94.90% | 89.11% | 50.23% | 72,548,253 | 63,163,421 | 80.38% |
| | 2 | 46,572,962 | 93.29% | 87.92% | 43.57% | 31,705,696 | 26,162,933 | 68.08% |
| | 3 | 153,045,578 | 93.29% | 86.88% | 46.12% | 109,131,672 | 91,159,663 | 71.31% |
| JI63 DAP 35 | 1 | 236,947,674 | 94.58% | 88.46% | 52.47% | 191,990,396 | 172,220,593 | 81.03% |
| | 2 | 71,881,146 | 93.60% | 86.93% | 58.45% | 51,008,354 | 47,576,576 | 70.96% |
| | 3 | 175,373,372 | 94.87% | 89.08% | 50.63% | 140,232,312 | 126,527,577 | 79.96% |
| JI63 Mature kernel | 1 | 73,863,982 | 96.16% | 91.10% | 60.54% | 57,660,477 | 30,866,985 | 78.06% |
| | 2 | 64,518,056 | 96.64% | 92.33% | 58.13% | 50,827,949 | 30,464,526 | 78.78% |
| | 3 | 70,535,482 | 95.86% | 90.55% | 60.11% | 53,851,285 | 29,096,977 | 76.35% |

**Table S3-3.** Folate metabolism genes in concerned modules

| Gene | ID | M.C. | S.L. | MM | p.MM | 5MTHF | | 5FTHF | | THF | | 5,10-CH=THF | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GS | p.GS | GS | p.GS | GS | p.GS | GS | p.GS |
| *ADCS* | Zm00001d036484 | turquoise | C | -0.929 | 2.46E-08 | -0.676 | 2.09E-03 | 0.491 | 3.87E-02 | 0.335 | 1.74E-01 | -0.324 | 1.89E-01 |
| *ADCL1* | Zm00001d029231 | green | O | -0.820 | 3.07E-05 | 0.158 | 5.31E-01 | -0.415 | 8.68E-02 | -0.506 | 3.21E-02 | 0.307 | 2.16E-01 |
| *ADCL2* | Zm00001d039264 | turquoise | O | -0.415 | 8.71E-02 | -0.474 | 4.69E-02 | -0.272 | 2.75E-01 | -0.272 | 2.74E-01 | -0.550 | 1.80E-02 |
| *GCHI-1* | Zm00001d001959 | blue | O | 0.838 | 1.40E-05 | -0.771 | 1.82E-04 | 0.352 | 1.52E-01 | 0.102 | 6.87E-01 | -0.044 | 8.62E-01 |
| *GCHI-2* | Zm00001d026531 | blue | O | 0.777 | 1.51E-04 | -0.235 | 3.48E-01 | -0.147 | 5.61E-01 | -0.185 | 4.63E-01 | 0.138 | 5.86E-01 |
| *DHNA1* | Zm00001d031979 | turquoise | O | -0.865 | 3.53E-06 | -0.608 | 7.40E-03 | 0.252 | 3.13E-01 | 0.227 | 3.65E-01 | -0.458 | 5.57E-02 |
| *DHNA2* | Zm00001d031995 | turquoise | O | -0.604 | 7.95E-03 | -0.242 | 3.33E-01 | 0.627 | 5.35E-03 | 0.667 | 2.48E-03 | -0.102 | 6.88E-01 |
| *HPPK/DHPS* | Zm00001d021338 | turquoise | O/M | 0.823 | 2.78E-05 | 0.399 | 1.01E-01 | -0.333 | 1.77E-01 | -0.337 | 1.72E-01 | 0.363 | 1.39E-01 |
| *DHFS1* | Zm00001d023817 | blue | M | -0.166 | 5.10E-01 | 0.078 | 7.57E-01 | 0.035 | 8.90E-01 | 0.222 | 3.76E-01 | -0.270 | 2.78E-01 |
| *DHFS2* | Zm00001d041625 | magenta | M | -0.555 | 1.68E-02 | 0.551 | 1.77E-02 | -0.233 | 3.53E-01 | 0.033 | 8.97E-01 | 0.222 | 3.76E-01 |
| *DRTS3* | Zm00001d007318 | turquoise | O | 0.848 | 8.69E-06 | 0.406 | 9.42E-02 | -0.516 | 2.84E-02 | -0.463 | 5.29E-02 | 0.303 | 2.21E-01 |
| *DRTS4* | Zm00001d049188 | turquoise | M | 0.668 | 2.44E-03 | 0.436 | 7.03E-02 | -0.698 | 1.26E-03 | -0.630 | 5.06E-03 | 0.187 | 4.57E-01 |
| *FPGS1* | Zm00001d032529 | turquoise | C/O | -0.743 | 4.14E-04 | -0.745 | 3.92E-04 | 0.436 | 7.04E-02 | 0.308 | 2.14E-01 | -0.357 | 1.46E-01 |
| *FPGS2* | Zm00001d048514 | brown | O | -0.847 | 9.25E-06 | 0.618 | 6.30E-03 | -0.633 | 4.81E-03 | -0.563 | 1.49E-02 | 0.469 | 4.98E-02 |
| *GGH* | Zm00001d010744 | brown | SP | -0.927 | 3.18E-08 | 0.554 | 1.71E-02 | -0.412 | 8.91E-02 | -0.419 | 8.39E-02 | 0.551 | 1.78E-02 |
| *GCSP* | Zm00001d023437 | turquoise | O | 0.828 | 2.24E-05 | 0.437 | 6.96E-02 | -0.841 | 1.21E-05 | -0.774 | 1.62E-04 | 0.265 | 2.88E-01 |
| *GCST* | Zm00001d002258 | turquoise | M | 0.787 | 1.05E-04 | 0.520 | 2.71E-02 | -0.823 | 2.73E-05 | -0.696 | 1.34E-03 | 0.216 | 3.90E-01 |
| *GCSH2* | Zm00001d015378 | turquoise | M | 0.827 | 2.28E-05 | 0.411 | 8.98E-02 | -0.705 | 1.09E-03 | -0.698 | 1.29E-03 | 0.285 | 2.51E-01 |
| *SHMT-GLY1* | Zm00001d049234 | turquoise | C | 0.757 | 2.76E-04 | 0.378 | 1.22E-01 | -0.858 | 5.41E-06 | -0.821 | 2.95E-05 | 0.214 | 3.94E-01 |
| *SHMT7-1* | Zm00001d042661 | brown | O | 0.818 | 3.30E-05 | -0.526 | 2.49E-02 | 0.639 | 4.34E-03 | 0.569 | 1.37E-02 | -0.241 | 3.35E-01 |
| *SHMT7-2* | Zm00001d012247 | blue | O | -0.847 | 9.15E-06 | 0.541 | 2.04E-02 | -0.057 | 8.24E-01 | 0.059 | 8.15E-01 | 0.306 | 2.17E-01 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *MTHFR1* | Zm00001d034602 | turquoise | O | 0.931 | 2.05E-08 | 0.540 | 2.07E-02 | -0.537 | 2.17E-02 | -0.483 | 4.25E-02 | 0.534 | 2.25E-02 |
| *MS1* | Zm00001d031129 | turquoise | O | 0.846 | 9.90E-06 | 0.384 | 1.16E-01 | -0.673 | 2.19E-03 | -0.642 | 4.09E-03 | 0.363 | 1.39E-01 |
| *MS2* | Zm00001d013644 | turquoise | O | 0.934 | 1.44E-08 | 0.500 | 3.47E-02 | -0.705 | 1.07E-03 | -0.665 | 2.61E-03 | 0.368 | 1.33E-01 |
| *DHC1* | Zm00001d005614 | turquoise | O/M | -0.865 | 3.51E-06 | -0.836 | 1.53E-05 | 0.276 | 2.68E-01 | 0.059 | 8.16E-01 | -0.464 | 5.26E-02 |
| *DHC2* | Zm00001d010867 | blue | O | 0.939 | 8.08E-09 | -0.547 | 1.88E-02 | 0.198 | 4.30E-01 | 0.076 | 7.64E-01 | -0.022 | 9.30E-01 |
| *DHC3* | Zm00001d053900 | turquoise | C | 0.589 | 1.00E-02 | 0.024 | 9.24E-01 | -0.773 | 1.69E-04 | -0.778 | 1.42E-04 | -0.066 | 7.95E-01 |
| *5-FCL-1* | Zm00001d006676 | greenyellow | C | -0.710 | 9.70E-04 | -0.188 | 4.54E-01 | 0.145 | 5.65E-01 | 0.032 | 9.00E-01 | 0.289 | 2.45E-01 |
| *5-FCL-2* | Zm00001d023534 | brown | C | 0.645 | 3.82E-03 | -0.161 | 5.22E-01 | 0.245 | 3.28E-01 | 0.301 | 2.25E-01 | -0.478 | 4.50E-02 |
| *FTHS* | Zm00001d020669 | cyan | O | 0.793 | 8.50E-05 | -0.036 | 8.86E-01 | -0.278 | 2.65E-01 | -0.195 | 4.38E-01 | -0.028 | 9.13E-01 |

MM: module membership; p.MM: p-value of module membership; GS: gene significance; p.GS p-value of gene significance. M.C.: module color; S.L.: subcellular localization by TargetP 2.0; M: mitochondrial transit peptide, C: chloroplast transit peptide, SP: signal peptide, O: other.

**Table S3-4.** Hub genes in purple, tan, red and pink module

| Module | Gene ID | Annotation | Function |
|---|---|---|---|
| Purple | Zm00001d049830 | Nascent polypeptide-associated complex subunit alpha-like protein 3 | Unfolded protein binding |
| | Zm00001d037376 | Early nodulin-like protein 7 | Electron transfer activity |
| | Zm00001d035201 | 60S acidic ribosomal protein P0 | Large ribosomal subunit / rRNA binding |
| | Zm00001d003127 | 60S ribosomal protein L23a-1 | rRNA binding |
| Tan | Zm00001d033850 | Tubulin alpha chain | GTPase activity |
| | Zm00001d048584 | Thioesterase family protein, mRNA | Acyl-CoA hydrolase activity |
| | Zm00001d019510 | Folate/biopterin transporter family protein | Transport |
| Red | Zm00001d052905 | Squamosa promoter-binding protein-like (SBP domain) transcription factor | DNA binding / metal ion binding |
| | Zm00001d038776 | Putative mitochondrial-processing peptidase subunit alpha-2 | Metal ion binding |
| | Zm00001d028307 | Putative inactive poly [ADP-ribose] polymerase SRO1 | NAD+ ADP-ribosyltransferase activity |
| | Zm00001d007009 | DNAJ heat shock N-terminal domain-containing protein | Stress response |
| | Zm00001d042475 | Putative thimet oligopeptidase | Metal ion binding |
| Pink | Zm00001d038272 | Beta-glucosidase | beta-glucosidase activity |
| | Zm00001d022611 | La-related protein 1C | RNA binding |

| | | |
|---|---|---|
| Zm00001d033726 | Putative serine/threonine protein kinase IREH1 | Protein serine/threonine kinase activity |
| Zm00001d037972 | 60S ribosomal protein L29 | Structural constituent of ribosome |
| Zm00001d039211 | Notchless-like protein isoform 1 | Ribosomal large subunit assembly |
| Zm00001d046996 | Tubulin beta chain | GTPase activity |

**Table S3-5.** The FPKM value of folate genes identified by WGCNA analysis

| | DAP 24 | | DAP 35 | | Mature kernel | |
|---|---|---|---|---|---|---|
| Gene | DAN3130 | JI63 | DAN3130 | JI63 | DAN3130 | JI63 |
| ADCS | 5.39±0.69 | 5.61±0.78 | 5.56±1.05 | 8.32±2.66 | 12.13±3.29 | 18.64±4.79 |
| ADCL2 | 26.92±0.73* | 16.05±4.01* | 25.30±2.25 | 23.57±12.32 | 17.39±9.70 | 21.05±4.27 |
| GCH-1 | 19.05±0.39 | 24.20±3.71 | 46.71±9.19 | 60.07±4.60 | 41.40±13.85 | 44.18±3.58 |
| DHNA1 | 14.50±1.32 | 13.38±3.95 | 18.26±0.72 | 15.47±3.74 | 26.08±8.83 | 19.70±1.71 |
| DHNA2 | 6.09±2.32 | 12.09±7.61 | 13.99±2.30** | 6.87±1.07** | 118.57±52.95* | 5.72±0.56* |
| DRTS3 | 24.82±1.97* | 17.07±2.59* | 18.90±2.00 | 21.56±5.99 | 7.65±2.12 | 5.78±1.94 |
| DRTS4 | 22.13±2.98 | 21.95±2.25 | 16.22±3.87 | 20.03±6.01 | 6.06±1.29 | 8.86±1.74 |
| FPGS1 | 7.05±0.35** | 2.40±0.32** | 3.97±0.32* | 7.20±1.83* | 9.43±3.61 | 7.37±1.36 |
| FPGS2 | 35.29±1.46** | 46.43±2.97** | 34.96±6.70 | 39.21±10.93 | 6.99±3.21 | 6.84±2.20 |
| GGH | 10.18±2.02** | 20.82±3.68** | 8.25±1.27* | 15.53±3.43* | 2.67±1.40 | 5.14±1.90 |
| GCSP | 32.94±2.32 | 20.59±3.39 | 18.63±3.28 | 29.94±0.92 | 0.79±0.29 | 3.08±0.48 |
| GCST | 10.15±0.69 | 9.51±1.52 | 8.74±1.63 | 8.83±1.08 | 2.43±0.27 | 2.87±0.22 |
| GCSH2 | 62.70±6.70 | 61.08±7.50 | 77.92±9.45 | 64.71±10.62 | 10.36±10.12 | 10.91±3.06 |
| SHMT-GLY1 | 66.25±4.43 | 83.02±17.49 | 98.71±4.32 | 79.73±13.00 | 14.09±3.03* | 21.58±2.08* |
| SHMT7-1 | 8.15±1.49* | 5.04±1.26* | 8.72±0.78 | 7.70±0.71 | 12.77±3.44* | 4.09±0.40* |
| SHMT7-2 | 22.84±1.40 | 24.62±2.82 | 26.81±3.19* | 18.55±2.20* | 16.62±6.01* | 4.90±0.74* |
| MTHFR1 | 42.65±3.50 | 38.89±7.28 | 46.46±3.05 | 66.41±17.36 | 5.32±1.62 | 2.71±0.31 |
| MS1 | 0.19±0.1* | 2.91±0.68* | 0.84±0.35 | 0.71±0.11 | 3.18±2.80 | 2.08±0.23 |
| MS2 | 71.04±7.90 | 53.13±10.36 | 70.30±5.78 | 78.50±19.02 | 1.53±0.51* | 2.58±0.23* |
| DHC1 | 12.52±1.42 | 10.65±0.73 | 16.27±1.06 | 15.85±1.99 | 16.33±5.49 | 20.03±1.41 |
| DHC2 | 38.73±1.28* | 60.69±7.93* | 58.62±4.45 | 102.51±20.11 | 64.53±9.72 | 81.86±10.98 |
| DHC3 | 23.24±1.67* | 15.64±2.94* | 25.44±1.97 | 21.67±3.27 | 7.33±0.69 | 8.89±2.25 |

*: T-test value less than 0.05; **: T-test value less than 0.01.

**Table S3-6.** DEGs related to folate metabolism in *ppdk* mutant and wild type (from the published transcriptome data of Zhang *et al*, 2018).

| Gene | ID | HL_WT_ Base | HL_homo_ Base | Ratio of homo/WT | HL_WT_ 4cm | HL_homo_ 4cm | Ratio of homo/WT |
|---|---|---|---|---|---|---|---|
| *ADCL2* | Zm00001d039264 | 0.42 | 0.24 | 0.57 | 0.21 | 0.13 | 0.62 |
| *GCH-1* | Zm00001d001959 | 0.72 | 0.81 | 1.13 | 0.78 | 2.12 | 2.72 |
| *DHNA1* | Zm00001d031979 | 0.56 | 0.37 | 0.66 | 0.18 | 0.49 | 2.72 |
| *DHNA2* | Zm00001d031995 | 0.41 | 0.59 | 1.44 | 0.28 | 0.47 | 1.68 |
| *GCSP* | Zm00001d023437 | 4.14 | 4.59 | 1.11 | 17.07 | 14.17 | 0.83 |
| *GCST* | Zm00001d002258 | 9.44 | 9.86 | 1.04 | 21.58 | 4.51 | 0.21 |
| *GCSH1* | Zm00001d023699 | 2.98 | 5.47 | 1.84 | 19.85 | 11.54 | 0.58 |
| *SHMT-GLY1* | Zm00001d049234 | 9.97 | 8.52 | 0.85 | 3.11 | 1.46 | 0.47 |
| *SHMT7-1* | Zm00001d042661 | 0.31 | 0.48 | 1.55 | 0.37 | 1.31 | 3.54 |

HL, hight light; WT, wild type; homo, homozygous mutant; Tissue was collected from leaf three at 9 DAP 3 h into the light period from sections: base (1 cm above the leaf three ligule); 4 cm (4 cm above the leaf two ligule); RPKM (reads per kilobase per million mapped reads) was used to estimate the expression level of individual genes.

**Table S3-7.** FPKM of genes related to folate metabolism in o2o2o16o16 mutant and the wild type (from 'the published transcriptome data from Wang *et al*, 2019).

| Gene | ID | CML530_ 18DAP_A | CML530_ 18DAP_B | QCL8011_2_ 18DAP_A | QCL8011_2_ 18DAP_B | Ratio of average QCL8011/CML530 |
|---|---|---|---|---|---|---|
| *ADCS* | Zm00001d036484 | 3.23 | 3.8 | 1.38* | 1.03* | 0.35 |
| *ADCL1* | Zm00001d029231 | 11.88 | 12.97 | 8.31* | 5.83* | 0.57 |
| *GCH-2* | Zm00001d026531 | 1.04 | 0.92 | 2.27** | 2.03** | 2.20 |
| *DHNA2* | Zm00001d031995 | 7.96 | 7.54 | 32.63 | 25.79 | 3.77 |
| *HPPK/DHPS* | Zm00001d021338 | 10.14 | 11.81 | 4.12** | 2.83** | 0.32 |
| *DHFS1* | Zm00001d023817 | 2.78 | 2.72 | 5.52** | 5.4** | 1.99 |
| *DHFS2* | Zm00001d041625 | 4.33 | 4.78 | 8.49* | 7.1* | 1.72 |
| *DRTS4* | Zm00001d049188 | 14.63 | 16.45 | 6.3** | 5.83** | 0.39 |
| *GGH* | Zm00001d010744 | 30.41 | 33.03 | 11.71 | 17.03 | 0.45 |
| *GCSP* | Zm00001d023437 | 25.1 | 23.3 | 6.37** | 6.63** | 0.27 |
| *GCST* | Zm00001d002258 | 16.98 | 17.92 | 23.92** | 23.85** | 1.37 |
| *GCSH2* | Zm00001d015378 | 51.54 | 49.76 | 95.06 | 123.71 | 2.16 |
| *GCSH3* | Zm00001d053800 | 54.67 | 51.8 | 238.49** | 261.89** | 4.70 |
| *GCSH4* | Zm00001d048974 | 0.34 | 0.47 | 4.09** | 5.43** | 12.06 |
| *SHMT-GLY1* | Zm00001d049234 | 126.9 | 122.74 | 191.25** | 198.73** | 1.56 |
| *SHMT7-1* | Zm00001d042661 | 10.84 | 12.34 | 2.48** | 1.75** | 0.18 |
| *SHMT7-2* | Zm00001d012247 | 12.16 | 14.26 | 3.49** | 2.89** | 0.24 |
| *MS1* | Zm00001d031128 | 287.42 | 291.19 | 326.88 | 363.5 | 1.19 |
| *MS3* | Zm00001d033480 | 18.91 | 19.93 | 24.75** | 25.91** | 1.31 |
| *DHC1* | Zm00001d005614 | 6.5 | 7.89 | 17.57** | 20.43** | 2.67 |
| *DHC2* | Zm00001d010867 | 33.38 | 33.2 | 83.95* | 61.97* | 2.19 |
| *DHC3* | Zm00001d053900 | 17.11 | 17.38 | 33.64** | 38.66** | 2.10 |
| *FTHS* | Zm00001d020669 | 43.31 | 45.36 | 51.63* | 51.24* | 0.35 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *10-FDF* | Zm00001d048550 | 2.09 | 2.37 | 0.75* | 0.19* | 0.57 |

CML530, wild type; QCL8011, mutant *o2o2o16o16*; 18DAP, 18 days after pollination; A, B, two replicates; FPKM (fragments per kilobase per million mapped reads) was used to estimate the expression level of individual genes. *, t-test value less than 0.05; **, t-test value less than 0.0

# 4

# Genetic Mapping of Folate QTLs Using a Segregated Population in Maize (*Zea mays L.*)

*In this chapter, we complete the genetic mapping of folate quantitative trait loci (QTLs) using a segregated population crossed by two maize lines. The QTLs are obtained to further identify the candidate gene. We make the hypothesis that gene ZmCTM could convert 5-M-THF to other folate derivatives in folate metabolism, and its function analysis by gene editing is performed to confirm it.*

From Wenzhu Guo*, **Tong Lian***, Baobao Wang et al. Genetic Mapping of Folate QTLs Using a Segregated Population in Maize (Zea mays L.), J Integr Plant Biol. 2019, 61(6): 675-690, doi: 10.1111/jipb.12811

From Ling Jiang, Wenzhu Guo, Yanjing Wang, Weiwei Wen, Weixuan Wang, **Tong Lian** et al. Conversion of 5-methyl-tetrahydrofolate to MeFox facilitates folate biofortification in maize. Doi:10.21203/rs.3.rs-308995/v1 (pre-print)

**Abstract：** As essential vitamins B for humans, folates accumulation in edible parts of crops, such as maize kernels, is of great importance for human health. Molecular breeding is an efficient approach for folate fortification, but is constrained by shortage of the knowledge on folate metabolism regulation. Here we report the genetic mapping of folate quantitative trait loci (QTLs) using a segregated population crossed by two maize lines, one high in folate (GEMS31) and the other low in folate (DAN3130). Two folate QTLs on chromosome 5 were obtained by the combination of F2 whole-exome sequencing and F3 kernel-folate profiling. These two QTLs had been confirmed by bulk segregant analysis using F6 pooled DNA and F7 kernel-folate profiling, and were overlapped with QTLs identified by another segregated population. Then the candidate gene, named as CTM, was identified as folate-binding molecular that may contribute to folate metabolism in maize. Loss of the *ZmCTM* function enhanced 5-methyl-tetrahydrofolate accumulation by three folds. We conclude that *ZmCTM* may participate in folate metabolism by converting 5-methyl-tetrahydrofolate to other folate derivatives in maize kernel.

**Keywords：** folates, QTL analysis, BSA, maize inbred lines

# 1 Introduction

Folates are water-soluble B vitamins that are essential for the growth and development of all species. Folates are coenzymes of many reactions in the synthesis of purine and pyrimidine, amino acid, pantothenic acid, and methionine. In microorganisms and plants, the basic pathways for folate synthesis and one-carbon metabolism have been well studied (Hanson and Gregory 2011; Strobbe and Van Der Straeten 2017). However, humans cannot biosynthesize folates de novo, and need a daily intake of more than 400 micrograms of folates per day to satisfy the needs of health. In pregnant women, folate deficiency can lead to the growth retardation and neural tube defects in the fetus (Rader and Schneeman 2006; Silver et al. 2015). Moreover, folate deficiency directly leads to megaloblastic anemia and high homocysteine essential hypertension, and increases the risk of cancer, cardiovascular disease, Alzheimer's disease, coronary atherosclerosis and other diseases (Blancquaert et al. 2010; Guo et al. 2017). Folate deficiency has become a global public health problem, especially in China (Bhutta and Salam 2012). In 2016, the five-year plan for national economic and social development proposed that the folate deficiency rate in pregnant women should be brought to below 5% in 2020 in China, highlighting the urgency of fighting folate deficiency.

Although humans can acquire folates from fruits and vegetables, the intake of folates from the staple crops is more important to human health, especially in developing countries. In Chinese diets, staple crops (e.g., cereals and tubers) provide the majority of calories, and serve as important sources of necessary macronutrients and key vitamins and minerals (e.g., folates, and calcium) (Chang et al. 2018). However, the edible parts of crops tend to be low in folate, and this is especially true in cereals, such as maize and rice (https://ndb.nal.usda.gov/ndb/nutrients/index). Targeted breeding could be applied to increase of folate contents in cereals. However, the prohibitive cost of folate measurement limits the phenotype-based selection, and it is difficult to screen for higher-folate-level varieties in breeding without any molecular genetic information (Dong et al. 2011).

Folates derivatives consist of a pteridine ring, a para-aminobenzoate ring, a tail of one or more L-glutamate residues, and different one-carbon units attached to the N5 and/or N10 positions of tetrahydrofolate (THF) molecules (Gorelova et al. 2017). The pathways for folates synthesis and one-carbon metabolism in different species are highly conserved. In plants, pteridine and para-aminobenzoate are biosynthesized in the cytoplasm and chloroplast, respectively, and then transported to the mitochondrion to form folpolyglutamates (Blancquaert et al. 2010). Based on the pathways of folate metabolism in plants, different strategies have been designed to increase the amounts of folates in crops, including enhancing folate and one-carbon metabolism, and increasing the stability of folates by genetic modification (Jiang et al. 2017). For example, to enhance folate biosynthesis, increasing dihydrofolate synthetase or folylpolyglutamate synthetase alone, reducing the activity of gamma glutamyl hydrolase alone, or increasing the transcriptional levels of genes encoding GTP cyclohydrolase I and aminodeoxychorismate synthase individually or together are three successful strategies in Arabidopsis (*Arabidopsis thaliana*), maize (*Zea mays*),

potato (Solanum tuberosum), rice (Oryza sativa), and tomato (Solanum lycopersicum) (Diaz de la Garza et al. 2004; Diaz de la Garza et al. 2007; Storozhenko et al. 2007; Naqvi et al. 2009; De Lepeleire et al. 2018; Liang et al. 2019). To increase the stability of folates, introducing mammalian folate-binding proteins and improving folate polyglutamylation have been shown to enable long-term storage of biofortified high-folate rice grains (Blancquaert et al. 2015). But a better understanding of folate genetic networks that control the folate accumulation in the edible parts of crops is still required, so are new candidate genes for breeding aimed at the enrichment of folates in maize kernels.

5-Formyltetrahydrofolate (5-F-THF) is the main derivative among folates, and 5-methyltetrahydrofolate and THF are also present in maize kernels (Lian et al. 2015). 5-F-THF is a relatively stable natural folate and can be used in dietary supplementation and the treatment of cancer (Suh et al. 2001). In one-carbon metabolism, the conversion between 5-F-THF and 5, 10-methenyltetrahydrofolate needs two different enzymes, and 5, 10-methenyltetrahydrofolate can form 5-methyltetrahydrofolate and other derivatives through a series of enzymatic reactions (Blancquaert et al. 2010). The genes related to folate metabolism in maize share high sequence similarity with those in other organisms, such as Arabidopsis, rice and Escherichia coli. However, there is no significant correlation between the transcription levels of these genes and the accumulation of folates in grains, indicating that there might be other key genes regulating folate metabolism in maize (Lian et al. 2015).

Quantitative trait loci (QTL) mapping is an efficient way to identify underlying genes and elements for phenotype variation, but it is rarely conducted in folates in maize. Due to the complexity of plant genomes and low density of markers, exploring genetic base of quantitative traits faces huge challenges (Yu et al. 2011). QTL mapping depends on the marker density and population size, and higher genetic marker density and larger population size can improve the precision of QTL mapping (Chen et al. 2014a). Relatively small population size of advanced populations such as recombination inbred lines (RIL) can reduce the cost of genotyping and phenotyping, but time consuming and expensive for advanced population construction. Conversely, early generation crosses such as F2s and backcross containing abundant recombinant, combined with high through-put genotyping method are powerful to detect QTLs (Vales et al, 2005; Chen et al, 2014a). Gene pyramiding by marker-assisted selection is an effective means for fortifying vitamins, such as vitamin A and vitamin E (Jiang et al. 2017), and the lack of functional-loci identification and prohibitive cost for folate profiling limits the molecular breeding of folates in maize to a large extent.

In this study, the detected major folate derivatives (5T) represented a mix of 5-F-THF and a speculated folate derivative which had not been identified as MeFox until a specific methodology for MeFox detection was developed in plants (Zhang, Jha et al. 2019, Shahid, Lian et al. 2020). We used a segregated population that was constructed from two maize inbred lines, one high in folate (GEMS31) and the other low in folate (DAN3130). Two QTLs on chromosome 5 were obtained by the combination of F2 whole-exome sequencing and F3 kernel-folate profiling. The phenotypic variation contribution rates were 26.7% (*q5T-a*) and 14.9% (*q5T-b*),

respectively. These two QTLs had been confirmed by bulk segregant analysis (BSA) using F6 pooled DNA and F7 kernel-folate profiling and QTL verification using another segregated population K22×DAN340 (Xu, Wang et al. 2017). BSA obtained several regions overlapped with *q5T-a* and *q5T-b*. And among QTLs obtained from the latter population, q5T-d and q5T-e could explain 10.6% and 7.1% of the phenotypic variation, and were overlapped with *q5T-a* and *q5T-b*, respectively. Thus, we successfully obtained two QTLs (*q5T-a* and *q5T-b*) that control 5T variation in maize grains. Subsequently, a folate-related gene (GRMZM2G124863) described as transferase with a folic acid binding domain was identified for the further functional analyses. Knock-out this candidate gene results in significantly increased 5-M-THF and decreased MeFox reveal its function of affecting 5-M-THF and MeFox accumulation in maize.

# 2 Results

## *2.1 Development of a Folate Segregated Population from phenotype-extreme-differing parents*

We investigated folate levels in maize kernels using a global germplasm collection (Yang, Gao et al. 2011). A mixture of folate derivates (5T) accounted for over 70.0 % of the total folates in 70.5 % of the inbred lines, which confirms that 5T is the main derivative among kernel folates (**Table 4-1**). Further phenotyping indicated that, irrespective of significant variations across the environments, GEMS31 contained more 5T than DAN3130, and the 5T level in GEMS31 was always around 20 times than that in DAN3130 (**Table 4-2**). Therefore, we used these two inbred lines to construct a segregated population by single seed descent beginning in 2014, with the seeds of F7 harvested in 2017.

During the development of this segregated population, genetic mapping using F2 whole-exome sequencing and F3 kernel folate profiling was carried out to obtain QTLs. Later, BSA using F6 bulked segregant sequencing and F7 kernel-folate profiling was performed to confirm the location of QTLs.

**Table 4-1.** The relationship between the percentages of inbred lines and percentages of 5T in total folates of dry maize seeds in 2013, N=359.

| The percentage of 5T in total folates (%) | The percentages of inbred lines (%) |
|---|---|
| 30.0-39.9 | 0.8 |
| 40.0-49.9 | 2.5 |
| 50.0-59.9 | 10.0 |
| 60.0-69.9 | 16.2 |
| 70.0-79.9 | 23.1 |
| 80.0-89.9 | 34.3 |
| 90.0-99.9 | 13.1 |

**Table 4-2.** Profiling of 5T in GEMS31 and DAN3130 inbred lines across there environments.

| Year | Location | 5T (nmol/g DW) | |
|---|---|---|---|
| | | GEMS31 | DAN3130 |
| 2009 | Hainan, China | 17.83±0.56 | 1.05±0.08 |
| 2010 | Yunnan, China | 8.01±0.43 | 0.37±0.02 |
| 2013 | Beijing, China | 5.27±0.32 | 0.26±0.01 |

## *2.2 Bin Map Construction using whole-exome sequencing data*

In all, 216 $F_2$ plants were sequenced using the exon-capture sequencing strategy. The sequences of each plant were aligned to the B73 reference genome Version 2 (only to the SNPs positions retained from the two parents), for SNP detection. The called SNP were filtered by the criteria of MAF>0.1 and missing rate<0.9. After the filtering, a total of 114,183 SNPs was selected for the following analysis. A sliding window method that calculates the ratio between the numbers of SNPs from the three genotypes (GEMS31, DAN3130 and Heterozygous) was used for genotype calling (Huang, Feng et al. 2009). The uninterrupted segments across the population were regarded as bins used for recombination map construction (see Methods). The resulting 3780 recombination bins captured the majority of the recombination events detected in the population (**Figure 4-1A**). The average physical length of the recombination bins was 548 kb, ranging from 479 Kb to 726 Kb across chromosomes (**Table 4-3**). A linkage map was constructed with these bins which served as markers. The map had a total genetic distance of 1482 cM with an average interval of 0.729 cM between the bins (**Table 4-3**). There were 12 bins that were >10 Mb, and 36 bins that were >5 Mb but <10 Mb. All of the >10 Mb bins and 22 of the 5-10 Mb bins were within 20 Mb of centromeres. Pairwise recombinational fraction analysis revealed that the linkage signals were diagonally displayed, validating the quality of the bin map (**Figure 4-1B**).

**Table 4-3.** The information of the linkage map.

| Chr. | Bins | | | Linkage map (cM) | | | Recombination (cM/Mb) | NO. of crossover |
|---|---|---|---|---|---|---|---|---|
| | Number | Mean (Mb) | Max (Mb) | Total | Mean | Max | | |
| 1 | 579 | 0.519 | 19.872 | 201.57 | 0.349 | 1.44 | 0.671 | 4.1 |
| 2 | 443 | 0.534 | 14.771 | 160.49 | 0.363 | 2.909 | 0.68 | 3.3 |
| 3 | 439 | 0.523 | 14.98 | 196.95 | 0.449 | 8.623 | 0.858 | 4 |
| 4 | 406 | 0.593 | 16.547 | 151.88 | 0.375 | 1.924 | 0.632 | 3.1 |
| 5 | 409 | 0.531 | 27.083 | 141.54 | 0.347 | 2.41 | 0.653 | 2.9 |
| 6 | 340 | 0.494 | 8.611 | 126.91 | 0.374 | 2.412 | 0.757 | 2.6 |
| 7 | 242 | 0.726 | 22.692 | 119.04 | 0.489 | 1.925 | 0.674 | 2.4 |
| 8 | 320 | 0.546 | 20.041 | 139.12 | 0.436 | 2.659 | 0.797 | 2.8 |
| 9 | 292 | 0.53 | 14.223 | 115.84 | 0.396 | 3.068 | 0.747 | 2.4 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 10 | 310 | 0.479 | 5.85 | 128.78 | 0.415 | 2.414 | 0.865 | 2.7 |
| Total | 3780 | 0.548 | 27.083 | 1482.12 | 0.399 | 8.623 | 0.729 | 3 |

To evaluate the power and accuracy of this genetic map for QTL mapping, QTL analysis of cob color was first performed in this population. There were 53 white cobs and 163 red cobs in the F2 population. The ratio of white: red fit the ratio of 1:3 ($p = 0.875$, Chi-square test), indicating that there might be only one Mendelian locus responsible for the trait variation. QTL mapping yielded only one locus, that was at Bin1_171-Bin1_175, with a physical position of 47.325–51.286 Mb on chromosome 1 (**Figure 4-2**). The peak bin, Bin1_172 (Chr1: 47.81–49.22 Mb) just encompassed the cloned gene pericarp color 1 (P1) at position 48.1Mb. P1 has been reported to regulate red pigmentation in cob, pericarp, tassel glumes, and husks (Robbins, Wang et al. 2009). Exact mapping of this locus for cob color demonstrated the quality of our bin map.

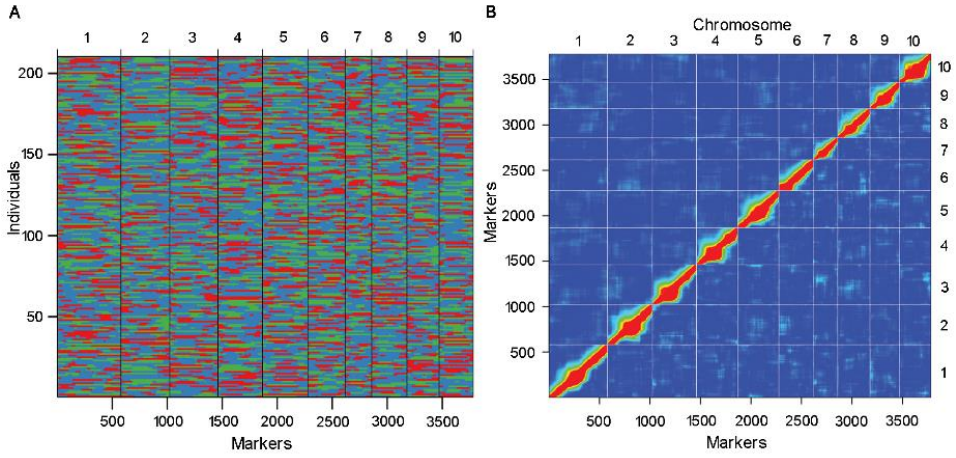



**Figure 4-1.** Recombination bin-map and heatmap for all markers of F2 whole-exome sequencing of GEMS31×DAN3130 population. (A) Bin-map consists of 3780 bin markers inferring 114,183 high qulity SNPs in F2 population. Physical position is based on B73 RefGen V2 sequence. Red: GEMS31 genotype; Green: DAN3130 genotype; Blue: heterozygote. (B) Pairwise recombination fractions are shown (upper left) as well as logarithm (base 10) of odds (LOD) which score for tests of linkage (bottom right) for all markers. Red corresponds to low recombination or a high LOD score, while blue represents unlinked markers.

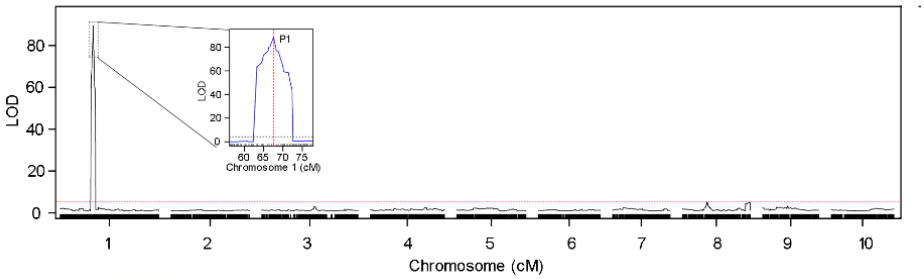



**Figure 4-2.** Mapping of a QTL controlling cob color in F2 population and the location of P1. Curve in plot indicate the genetic coordinate (X-axis) and LOD score (Y-score) of detected QTL, and the horizontal dashed line presents the LOD threshold. The peak encompasses the cloned gene pericarp color 1 (P1) at position 48.1Mb on chromosome 1. The box inside is the zoom-in image of the peak on chromosome 1. The vertical gray dashed line presents the peak position of P1 gene.

## 2.3 QTL analysis for 5T variation

GEMS31, DAN3130 and 216 F2 plants of the GEMS31×DAN3130 population were grown in Langfang, Hebei, China, in 2015. 5T contents of GEMS31 and DAN3130 were 4.61 ± 0.21, and 0.21 ± 0.02 nmol/g DW, respectively. 5T contents of the F2 population ranged from 0.08 ± 0.01, to 4.88 ± 0.32 nmol/g DW, and the average level was 1.51 nmol/g DW. The phenotypes in the F2 population showed a continuous distribution, indicating the quantitative characteristics of this trait (**Figure 4-3**).

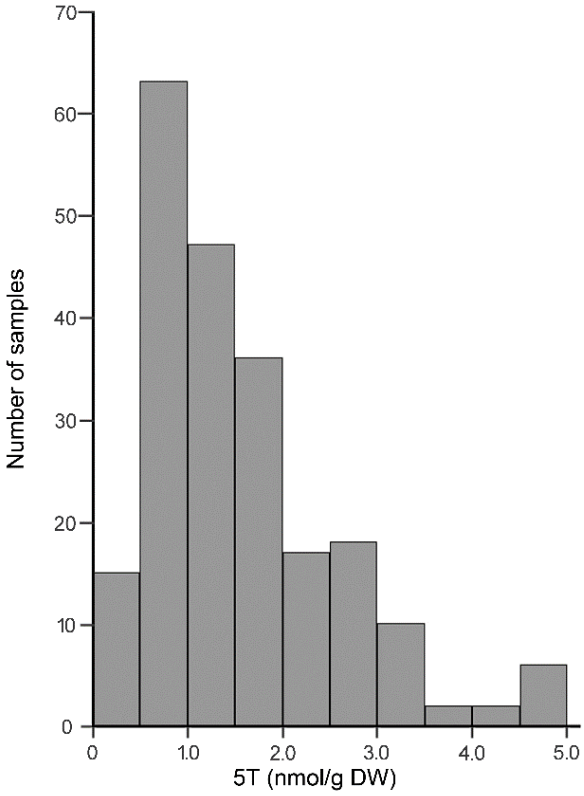
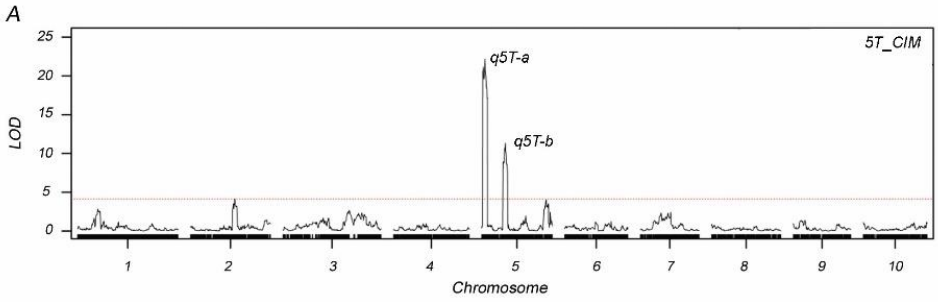



**Figure 4-3.** Variation of 5T levels in F3 kernels of the F2 population. The number of samples is shown as gray columns in the variation of 5T levels. 5T levels ranged from 0.08 ± 0.01, to 4.88 ± 0.32 nmol/g DW, and the average level was 1.51 nmol/g DW.

Next, we used our bin map to identify QTLs controlling kernel 5T variation. Our analysis of the 216 F2 plants detected two QTLs (*q5T-a* and *q5T-b*), with likelihood of odds (LOD) peaks overlapping with Bin5_17 and Bin5_128 on chromosome 5, which explained 26.7% and 14.9% of the phenotypic variation, respectively **(Figure 4-4A**). *q5T-a* occupied a physical position of 1.532–2.672 Mb, and *q5T-b* was at 16.638–22.402 Mb on chromosome 5, respectively (**Figure 4-4B**). The positive contributors to these two QTLs were both derived from the inbred line GEMS31. According to maize gene annotation database accessible at MaizeGDB (www.maizegdb.org), there are 88 protein-coding genes and 142 protein-coding genes, respectively. We have compared these protein-coding genes with genes expressed in seeds, and 68 and 102 genes have been obtained in *q5T-a* and *q5T-b*, respectively (Chen, Zeng et al. 2014). Then we filter these genes with tissue-specific expressions

by using high density NimbleGen Microarrays (Winter, Vinegar et al. 2007, Sekhon, Lin et al. 2011) and RNA Sequencing (Stelpflug, Sekhon et al. 2016). 22 and 32 genes have relatively high transcript levels during milking stage (DAP 18 to DAP 24) in either embryo, endosperm or whole seeds in *q5T-a* and *q5T-b*, respectively (**Table S4-1**). Among these genes after filtration, there are 22 genes coding functional unclassified proteins, 13 genes involved in protein synthesis, modification and degradation, 7 genes involved in RNA regulation. Based on the annotated function involved in metabolic pathway, two genes might be involved in folate-related metabolism: *GRMZM2G121840*, a S-adenosyl-L-methionine-dependent methyltransferases, and *GRMZM2G124863*, a transferase which contain a folic acid binding domain. More biochemical and genetic work is needed to dissect the relationship between these genes and folate metabolism.




B

| QTL | Chr | Peak_Bin | Peak_Lod | Peak_Pos_Link (cM) | Peak_Pos_Phy (Mb) | QTL_information | | | | | | Var% | Add | Dominance | Positive_Allele |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Int_Left | Int_Right | Left_Link | Right_Link | Left_Phy | Right_Phy | | | | |
| q5T-a | 5 | Bin5_17 | 22.187 | 6.727 | 2.297 | Bin5_5 | Bin5_21 | 1.715 | 7.920 | 1.532 | 2.672 | 26.723 | 0.616 | -0.527 | GEMS31 |
| q5T-b | 5 | Bin5_128 | 11.340 | 47.314 | 20.139 | Bin5_122 | Bin5_133 | 45.166 | 49.223 | 16.638 | 22.402 | 14.942 | 0.544 | 0.154 | GEMS31 |

**Figure 4-4.** Mapping of QTLs for 5T variation using GEMS31×DAN3130 population. (A) Location of 5T QTLs. Peaks of QTLs are shown as q5T-a and q5T-b on chromosome 5, respectively. Curve in plot indicate the genetic coordinate (X-axis) and LOD score (Y-score) of detected QTL, and the horizontal dashed line presents the LOD threshold. (B) Details of 5T QTLs. Chr, chromosome; Var, phenotypic variation contribution rate; Add, additive effect.

## *2.4 Validation of folate QTLs*

We had used two methods to validate above two folate QTLs: bulk segregant analysis using GEMS31×DAN3130 RIL population and QTL verification using another segregated population K22×DAN340. First, GEMS31, Dan3130 and 216 F6 plants of the GEMS31×DAN3130 population were additionally grown in Langfang, Hebei, China, in 2017. The 5T contents of GEMS31 and DAN3130 were 5.29 ± 0.45, and 0.70 ±0.01 nmol/g DW, respectively. The 5T contents of the GEMS31×DAN3130

F7 seeds ranged from 0.57±0.05 to 8.01±0.51 nmol/g DW, and the average level was 1.57 nmol/g DW. For bulk sergeant analysis, 30 DNA samples from 15 lines (2 siblings for each line) of F6 plants that exhibited either extreme-high or extreme-low 5T in the F7 seeds were pooled to form the extreme-high and extreme-low groups, respectively (**Table S4-2**). Both pools were sequenced and analyzed via BSA (Takagi, Abe et al. 2013), identifying several regions at the physical position from 1.40 to 21.20 Mb on chromosome 5 (**Figure 4-5**). *q5T-a* was found to be overlapped with one region (1.40-2.65 Mb); *q5T-b* was found to be overlapped with some regions (16.55-17.10 Mb, 17.75-17.85 Mb, 18.55-19.50 Mb, 19.80-20.55 Mb, 20.60-21.20 Mb, **Figure 4-5B**). Second, we used another segregated population K22×DAN340 originally constructed for tocopherols QTL analysis to do folate QTL verification (Xu et al. 2012). The 5T levels have been measurement in K22, DAN340 and 188 F3 kernels of K22×DAN340 population. 5T level in DAN340 kernels was about 2.1 times of that in K22 kernels (2.85 ± 0.26 nmol/g DW vs 1.35 ± 0.11 nmol/g DW). 5T content displays abundant phenotypic variation ranging from 0.62 ± 0.05 to 2.89 ± 0.19 nmol/g DW in the population. QTL mapping results showed that total four QTLs were detected in this population (*q5T-c* to *q5T-f*) (**Figure 4-6A**). Among them, there are two main genetic loci (*q5T-d* and *q5T-e*) that have overlapped with the localization of *q5T-a* and *q5T-b* obtained from GEMS31×DAN3130 population (**Figure 4-4B**, **Figure 4-6B**). *q5T-d* and *q5T-e* could explain 10.6% and 7.1% of the phenotypic variation in K22×DAN340 population, respectively. These results thus confirmed that there are two folate QTLs (*q5T-a* and *q5T-b*) related to 5T variations in maize kernels.

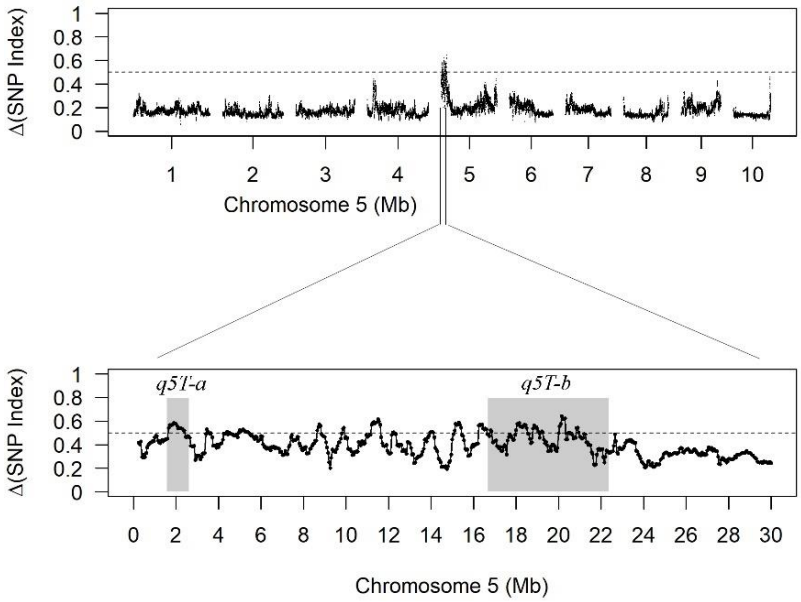
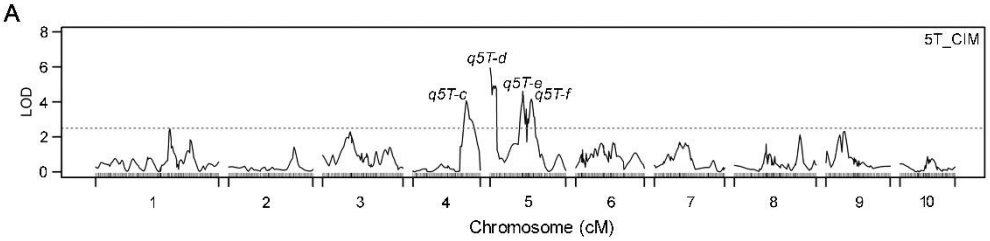



**Figure 4-5.** BSA results for GEMS31×DAN3130 F6 populations. (A) BSA regions at the physical position from 1.40 to 21.20 Mb on chromosome 5. Curve in plot indicate the

physical coordinate (X-axis) and ΔSNP index (Y-score) of detected regions, and the horizontal dashed line presents the ΔSNP index threshold. (B) The zoom-in image of the above regions on chromosome 5. Gray columns indicate q5T-a and q5T-b from QTL analysis.

A




B

| QTL | Chr | Peak_LOD | Peak_Pos_Phy (Mb) | QTL_int (Mb) | | Var% | Add | Dominance |
|---|---|---|---|---|---|---|---|---|
| | | | | Left_Phy | Right_Phy | | | |
| q5T-c | 4 | 4.07 | 235.44 | 233.628 | 239.510 | 7.027 | -0.155 | -0.138 |
| q5T-d | 5 | 6.01 | 3.539 | 0.000 | 6.860 | 10.624 | -0.228 | -0.089 |
| q5T-e | 5 | 4.61 | 35.277 | 18.349 | 58.604 | 7.099 | -0.062 | -0.270 |
| q5T-f | 5 | 4.16 | 91.814 | 66.432 | 188.609 | 6.335 | -0.042 | -0.279 |

**Figure 4-6.** Mapping of QTLs for 5T variation using K22×DAN340 population. (A) Location of 5T QTLs. Peaks of QTLs are shown as q5T-c to q5T-f on chromosome 5, respectively. Curve in plot indicate the genetic coordinate (X-axis) and LOD score (Y-score) of detected QTL, and the horizontal dashed line presents the LOD threshold. (B) Details of 5T QTLs. Chr, chromosome; Var, phenotypic variation contribution rate; Add, additive effect.

## *2.5 CTM affects 5-M-THF accumulation in maize*

A folate-related gene (GRMZM2G124863) described as transferase with a folic acid binding domain was identified from QTL results, and we named it as *ZmCTM*. To investigate whether *ZmCTM* is responsible for folate accumulation in corn seeds, *ZmCTM*-edited maize plants carrying a stop-gain mutation were generated using the CRISPR/Cas9 system (**Figure 4-7A**). The mutant plants showed no obvious difference in morphology compared with the wild type (**Figure 4-7B**). With the development of folate profiling, 5T was separated into 5-F-THF and MeFox (Wan, Han et al. 2019, Shahid, Lian et al. 2020). Folates in young seeds (DAP 25) of the mutant plants were profiled, showed 1.55-1.87-fold increase of 5-M-THF-Glu1, 2.72-3.62-fold increase of 5-M-Glu2, 4.77-5.81-fold increase of 5-M-Glu3, and 2.24-2.71-fold increase of total 5-M-THF. Besides, 91.8% to 93.8% decrease of MeFox were found, while not significant variation was observed in 5-F-THF, 5,10-$CH_2$-THF and THF. Our data shown 1.97 to 2.30-fold increase of total folate, including total 5-M-

THF, total 5-F-THF, 5,10-$CH_2$-THF and THF. Thus, *ZmCTM* was considered responsible for folate derivative accumulation, especially 5-M-THF, in maize kernel.

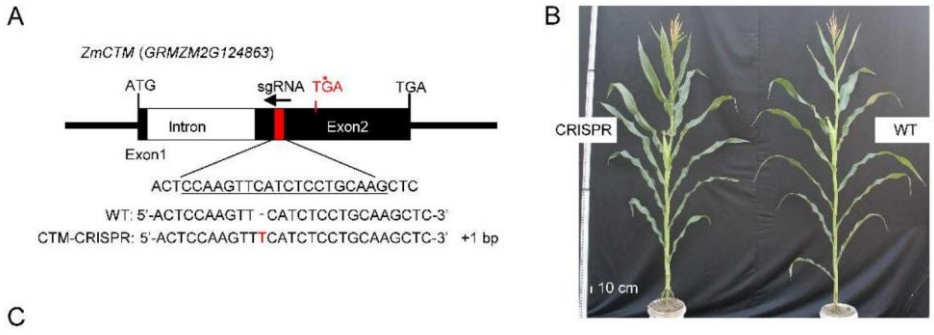



**Figure 4-7.** Phenotypes of ZmCTM editing in maize. (A) Gene structure of the *ZmCTM* and sgRNA target site in exon 2. The sgRNA target sequence is underlined. TGA, indicated by an asterisk, represents the resulting stop-gain mutation due to the insertion of T, highlighted in red. (B) The photos of wild-type (WT) maize and *ZmCTM*-edited (CRISPR) plants at day after sowing 70.

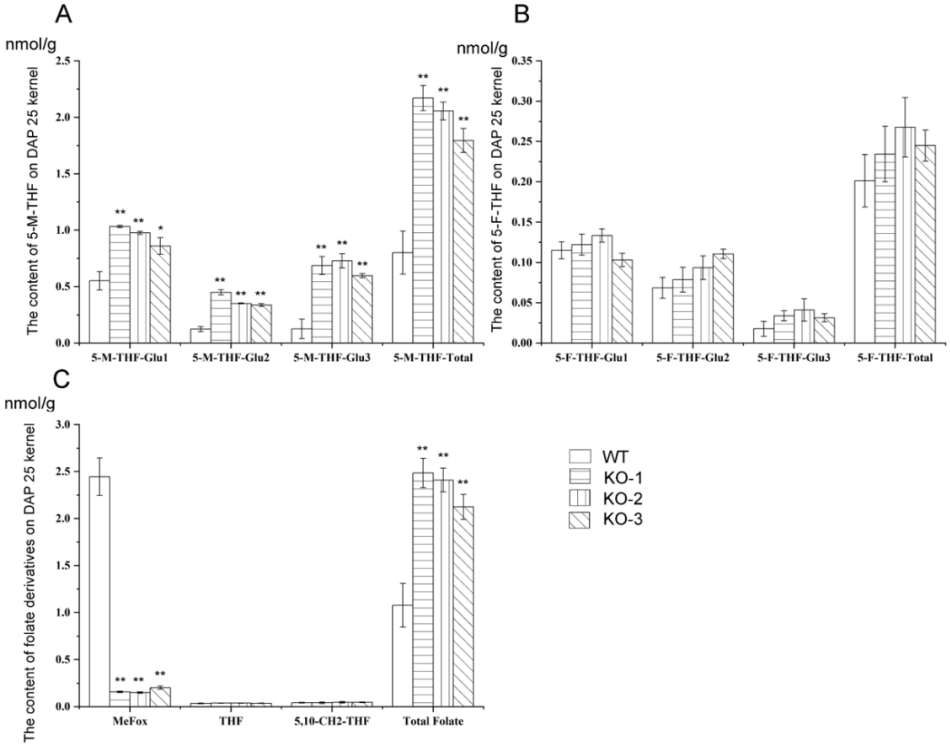


**Figure 4-8.** Effects of ZmCTM editing on folate accumulation in maize plants. (A) The content of 5-M-THF on DAP 25. (B) The content of 5-F-THF on DAP 25. (C) The contents of other folate derivatives on DAP 25 kernel. WT, wild type; KO, knock-out; *, p-value < 0.05; **, p-value < 0.01.

## *2.6 Conserved analysis of ZmCTM*

*ZmCTM* was predicted to encode a glutamate formiminotransferase (GFT) of 326 amino acids (aa) because the presumed protein showed 20.7% amino acid sequence identity to the formiminotransferase (FT) domain (326 aa) of the porcine (Sus scrofa) formiminotransferase cyclodeaminase (SsFTCD, 541 aa; **Figure 4-9A**). Algal and land plant CTMs showed a well conserved motif pattern2. For example, ZmCTM had 81.6% amino acid sequence identity to sorghum glutamate formiminotransferase (SbGFT, SbCTM hereafter; 317 aa; **Figure 4-9B**; http://blast.ncbi.nlm.nih.gov).

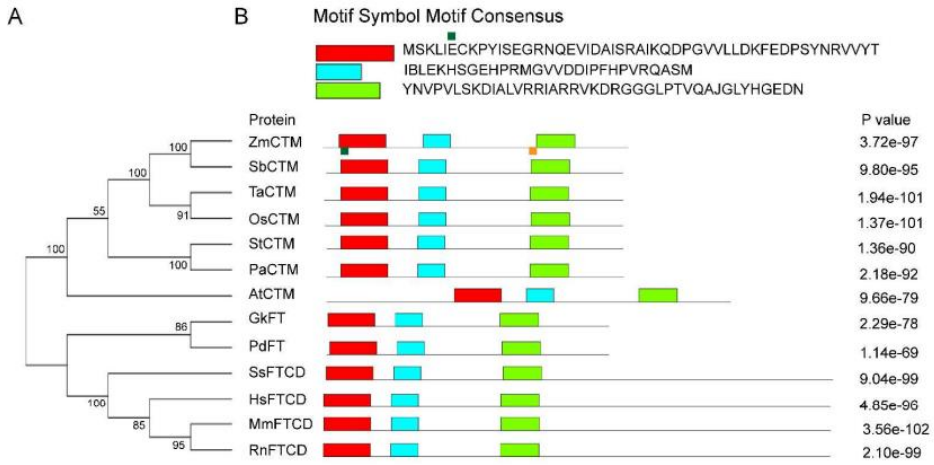



**Figure 4-9.** Phylogenetic-tree (A) and motif analysis (B) of CTM orthologues from plants and mammals. ZmCTM-B73, sequence from maize (Zea mays) inbred line B73, NP_001130076.1; ZmCTM-Qi319, sequence from maize inbred line Qi319, AMK92167.1; SbCTM, sequence from Sorghum bicolor, XP_002466878.1; TaCTM, sequence from Triticum aestivum, KAF6990789.1; OsCTM, sequence from Oryza sativa, XP_015633257.1; PaCTM, sequence from Prunus avium, XP_021832372.1; StCTM, sequence from Solanum tuberosum, XP_006357514.1; AtCTM, sequence from Arabidopsis thaliana, NP_973497.1; GkFT, sequence from Gloeobacter kilaueensis, WP_023175819.1; PdFT, sequence from Parabacteroides distasonis, WP_022193267.1; SsFTCD, sequence from Sus scrofa, NP_999440.1; HsFTCD, sequence from Homo sapiens, NP_006648.1; MnFTCD, sequence from Mus musculus, NP_543121.1; RnFTCD, sequence from Rattus norvegicus, NP_446019.1.

# 3 Discussion

## *3.1 The insistent demands for genetic mapping of new folate-candidate regions in maize*

Increasing folate intake is an efficient way to battle folate deficiency, and the current international practice is to take folic acid tablets and add folic acid to flour in developed countries (Bhutta and Salam 2012). However, the latest research shows that excessive intake of synthetic folic acid can lead to changes in the levels and patterns of DNA methylation in human bodies, accelerate the occurrence and development of some kinds of cancers, and also mask the deficiency of vitamin B12 (Cole, Baron et al. 2007, Bae, Ulrich et al. 2014, Moore, Ames et al. 2014). More importantly, some populations carry a homozygous mutation in the gene for methylenehydrofolate reductase (the 677TT genotype); in those cases, even supplying folic acid tablets does not effectively control the homocysteine level (Waskiewicz, Piotrowski et al. 2011). Thus, intake of bioactive folates from edible parts of crops is more health and important.

In maize, even though co-expressed *Gm8gGCHI* and *GmADCS* driven by endosperm-specific promoters results a 4.2-fold increase in folate levels in transgenic grains, the understanding of folate genetic networks remains unclear (Liang, Wang et al. 2019). And there is rarely a correlation between folate content and expression of conserved genes of folate biosynthesis and metabolism in kernels (Lian, Guo et al. 2015). Furthermore, the measurement of folates is unacceptably expensive for breeders, and different from most other agronomic traits, the breeding-based phenotype selection is untenable for folate breeding. Thus, genetic mapping of new folate-candidate regions is imperative. These regions could be new genetic regulators, and serve a key role in regulating the accumulation of folates in grains, and the development of corresponding molecular makers will be an efficient way to promote efforts for breeding higher-folate-level varieties of maize.

## *3.2 Rapidly mapping QTLs using a segregated population derived from phenotype-extreme-differing parents*

5T is the most dominant derivative in mature maize grains (Table 4-1). In this study, a folate segregated population developed from a cross of maize inbred lines with extreme-high and extreme-low levels of 5T was used to dissect the genetic architecture of folate accumulation in maize grains. This population consisted of 216 F2 plants and was genotyped by whole-exome sequencing (Lu, Liu et al. 2018). Two folate QTLs were obtained and validated by BSA using GEMS31×DAN3130 RIL population and QTL verification using another segregated population K22×DAN340 (**Figure 4-6**).

The mapping resolution for folate QTLs is much higher than in previous findings (Price 2006, Chen, Wang et al. 2014). Previously, the mapping resolution was 2 cM or less when using QTL mapping to position genes with a population size of 100–200 lines and hundreds of markers (Price 2006); 0.8–56.6 Mb for tassel and ear architecture QTLs were also obtained from a large early generation population (708 F2 progeny) combined with sequencing-based genotyping (Chen, Wang et al. 2014). In our case, a similar strategy, but with a much smaller population size (216 F2

progeny) was used here and this analysis achieved higher resolution than previous methods, with 1.1 and 5.8 Mb for each folate QTLs interval, respectively (**Figure 4-4**).

Normally the aim of QTL analysis with replication in different environments was to check the repeatability of the phenotype and to acquire more reliable results. Validations are usually needed when QTL analysis is using an F2 population without replication (Vales, Schon et al. 2005, Chen, Wang et al. 2014). In our case, the reliability of mapping loci had been validated by BSA using GEMS31×DAN3130 RIL population and QTL verification using another segregated population K22×DAN340. Both the regions from BSA and *q5T-d* and *q5T-e* from another QTL analysis were overlapped with *q5T-a* and *q5T-b*, respectively (**Figure 4-5B** and **Figure 4-6B**). Thus, we inferred that combining a small population derived from parents having extremely different phenotypes with genotyping by sequencing strategy would be another efficient method for rapid QTL identification, and reduce costs by reducing the number of individuals subjected to genotyping and phenotyping.

## *3.3 Two QTLs for folates accumulation in maize kernels*

Even though folates have the characteristics of quantitative traits (**Figure 4-3**), the genetic regulation of folate accumulation in kernels is much different from that of traditional flowering time, morphological and stress-related traits. Those traits have a highly polygenic genetic architecture that is mainly controlled by many small effect QTLs (Mickelson, Stuber et al. 2002, Peiffer, Romay et al. 2014, Raihan, Liu et al. 2016, Xu, Wang et al. 2017). Parental phenotypic and genetic distances are positively associated with the effects and numbers of QTLs yielded (Pan, Xu et al. 2017), and major-effect QTLs can be more easily mapped to a higher resolution (Wang, Zhu et al. 2018). However, in our case the two QTLs can explain 41.6% (with each of them >10%) of total phenotypic variation in 5T, reflecting that both are major-effect QTLs. This feature makes them well suited for marker-assisted selection, which is mainly used in maize breeding for qualitative-trait genes or major-effect QTLs (Pixley and Warburton. 2010). Furthermore, we have filtered protein-coding genes in QTLs using tissue-specific expressions data, and total 54 genes have relatively high transcript levels during milking stage (DAP 18 to DAP 24) in either embryo, endosperm or whole seeds (**Table 4-4**). Two candidate genes were inferred to be involved in folate-related metabolism: one encodes a protein whose coenzyme is S-adenosyl-L-methionine, which obtains a methyl group from 5-methyltetrahydrofolate during methyl cycle (Abbasi, Abbasi et al. 2018); the other one encodes a transferase protein containing folic acid binding domain. More biochemical and genetic work is needed to dissect the relationship between these genes and folate metabolism. We can expect that nutrition rich maize can be obtained combining these new elements involved in folate biosynthesis or one-carbon metabolism

## *3.4 CTM participates in folate metabolism in plants*

The identification of CTM function in 5-M-THF metabolism benefitted from development of MeFox chromatographic detection method. When we originally

detected the folate derivates for genome-wide association study (GWAS) and QTL analysis using the LC/MS method developed for Arabidopsis (Jiang, Liu et al. 2013), MeFox coeluted with 5-F-THF and thus these two folate derivatives couldn't be distinguished. Recently, the LC/MS detection for folate derivatives was improved in plants, allowing separation of MeFox from 5-F-THF in lentil and maize (Zhang, Jha et al. 2019, Shahid, Lian et al. 2020). The data from our lab and other labs showed that MeFox accumulation varied significantly amongst different plant species (Hilton, Christensen et al. 2003, Zhang, Jha et al. 2019, Shahid, Lian et al. 2020). For example, MeFox was undetectable in Arabidopsis leaves, but abundant in mature seeds of cereal crops (Ringling and Rychlik 2017). Using this newly developed method (Shahid, Lian et al. 2020), MeFox was eventually quantitatively detected and, consequently the *ZmCTM* gene was mapped by GWAS analysis. In human serum, 5-M-THF was found to be converted to MeFox (Fazili and Pfeiffer 2013). However, the protein that is involved in this conversion has not yet been identified (Fazili and Pfeiffer 2013, Fazili, Sternberg et al. 2014, Ringling and Rychlik 2017), and this is also the case for plants. In this study, *ZmCTM* was found associated with of the accumulation of 5-M-THF, and the functional domain was highly likely to be conserved in plants. However, the molecular mechanisms underlying the CTM protein converting 5-M-THF to MeFox remain largely unknown, and thus await further investigations.

# 4 Materials and Methods

## *4.1Plant material and the development of a folate segregated population*

The set of dry seeds powder (n=359) for the folate profiling from the global germplasm collection were kindly provided by Dr. Jianbing Yan in 2009, 2010, and 2013, respectively. The set of dry seeds powder (n=188) for the folate profiling from K22, DAN340 and 188 F3 kernels of segregated population K22×DAN340 were kindly provided by Dr. Jianbing Yan in 2009. DAN3130 is originated from China, belonging to the NSS subpopulation with pedigree being American hybrid P78599; GEMS31 is from the United States, belonging to the TST subpopulation with pedigree being 2282-01_XL380_S11_F2S4_9226-Blk26/00 (Yang et al. 2011). Both inbred lines were grown at Langfang, Hebei, China in the summer of 2014. The experimental field was loamy soil with pH 6.8, organic matter 0.7 %, phosphorus 13.8 mg/L, and potassium 48 mg/kg. During field preparation, 440 kg/acre of urea (46-0-0) were applied. The herbicides were applied 5 d after planting. Plants were hand planted in 5-m-long rows with row and plant spacing of 60 and 25 cm, respectively. The segregated population was developed by single seed descent from the cross (Murigneux, Barloy et al. 1993). F1 seeds were obtained, and later grew at Ledong, Hainan, China in the winter of 2014. F2 plants were grown at Langfang, Hebei, China in the summer of 2015 under the same conditions. All the DNA of 21-d-old leaves have been collected and frozen in -80 ℃ for the whole-exome sequencing. And F3 kernels have been collected and used for folate analysis. F3 plants have been grown at Ledong, Hainan, China in the winter of 2015; F4 plants have been grown at

Langfang, Hebei, China in the summer of 2016. F5 plants have been grown at Ledong, Hainan, China in the winter of 2016. F6 plants have been grown at Langfang, Hebei, China in the summer of 2017. All the DNA of 21-d-old leaves have been collected and frozen in -80 ℃ for the BSA analysis. And F7 kernels have been collected and used for folate analysis.

## *4.2 Whole-exome sequencing*

Genomic DNA was isolated from young leaves of 216 F2 plants using the Plant Genomic DNA Extraction Kit (Karroten). DNA was quantified using Nanodrop 2000 (Thermo Scientific). Then 4.5 μg of genomic DNA (35 ng/μl) was fragmented into 200–300 bp using Bioruptor UCD-200 (Diagenode). Libraries were constructed using NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs), the exomes were captured using the Capture Beads from the SeqCap EZ Pure Capture Bead Kit, washed, and amplified by Ligation-Mediated PCR, and were sequenced by Hiseq Xten (Illumina) to generate 150-bp paired-end reads and 8-bp indexed reads (Lu, Liu et al. 2018).

## *4.3 Folate analysis*

The samples of dry seeds collected from field were used for identification of folate profiles. The methods for sample preparation and metabolite measurement were described as previous study (Lian, Guo et al. 2015). Detection of folate derivatives, including MeFox in young seeds and mature seeds from maize, was carried out in accordance with the method described by Shahid et al (Shahid, Lian et al. 2020). Folates were measured in four biological replicates, and each sample consisted of 50 mg of plant material. The average content of folate was used as the phenotypic data.

## *4.4 SNP calling*

After being trimmed，the sequencing data was mapped to the reference genome using BWA software (Li and Durbin 2009)，we used the default parameter and pair-end mem model. The software Samtools was used to genotype every single nucleotide polymorphism of each sample within the background of polymorphism (191,223) between the parental samples (Li, Peng et al. 2013). Then we obtain the genotype of each F2 sample in the reference of parental polymorphism. Totally, 172,271 SNPs were obtained in the F2 population. The raw SNPs were further filtered by the criterions of MAF>0.1 and missing rate<0.9, thus 114,183 high-quality SNPs were used for subsequently analysis finally.

## *4.5 Bin map construction and QTL analysis*

The bin map was constructed using a sliding window approach as described by Huang et al (2009) with a little modification. Raw SNPs were scanned in 15-SNP-windows with a sliding step of one SNP. In each window, the genotype was defined by the ratio of three kinds of SNPs (GEMS31, DAN3130 and Heterozygous): the widows was called homozygous GEMS31, or DAN3130 genotype when more than 12/15 of SNPs were from either GEMS31, or DAN3130, otherwise it would be

defined as heterozygous genotype Regions with no more than 15 continuous heterozygous windows were set as breakpoint, since there will be several heterozygous windows for transition at the breakpoints, but these should not span more than 15 uninterrupted windows. Adjacent windows with same genotypes were merged as a block. Blocks with length less than 500kb, or windows less than 15 were set as missing to avoid false double crossover. Adjacent windows and successive small blocks (less than 100 kb or 5 mapped SNPs) with frequently transient genotypes were merged and defined as a larger heterozygous block. 5 kb incremental segments with the same genotype across the 216 F2 were merged together as a bin marker. Bins with extreme distortion (Chi- square, p-value < 1.0 E-10) were discarded. The bins with missing genotype were imputed by R/qtl software using the "argmax" method. The linkage map was constructed with the est.map function of R/qtl (Broman, Wu et al. 2003).

Composite interval mapping (CIM) was performed for QTL analysis using the R/qtl package (Broman et al. 2003) with a scanning window size of 10-cM. The LOD threshold for QTL defining was calculated by 1000 times permutations ($p < 0.05$) for each trait. The confidence intervals were estimated using the 1.5 LOD-drop method (Wang et al. 2018).

In the F2 K22×DAN340 population, composite Interval Mapping (CIM) implemented in Windows QTL Cartographer V2.5 was used for QTL mapping in the population of Dan340/K22 F2:3. Zmap (model 6) with a 10 cM window and an interval-mapping increment of 2 cM were selected. The confidence interval was determined 1.5 LOD value down flanking the peak (Zeng et al. 1994; Silva et al. 2012).

## *4.6 BSA analysis*

Two libraries were sequenced by Hiseq Xten, and we obtained 110.27 and 128.18 million cleaned paired-end reads in each library (GEMS31 and DAN3130), respectively. The Q20 percentages were from 92.19 to 98.38, respectively. Next, the reads ere aligned to the maize genome (ftp://ftp.ensemblgenomes.org/pub/plants/release-37/fasta/zea_mays/dna/), and the SNPs and Indels were called using HaplotypeCaller in Genome Analysis Toolkit (GATK) after bam sorting, index using SAMtools v0.1.18 (Li, Handsaker et al. 2009, McKenna, Hanna et al. 2010). The raw variants were then filtered using GATK VariantFiltration (for SNPs: QD <2, MQ<40, FS>60, MQRankSum < -12.5, ReadPosRankSum < −8.0 and SOR >3; for Indels: QD < 2, FS > 200, SOR > 10, MQRankSum < -12.5 and ReadPosRankSum < -8). Besides that, the SNPs and Indels with only two alleles, a coverage of more than ten reads and detected in both pools were retained for further analysis. Finally, we obtained 3.88 M SNPs and 0.473 M Indels for GEMS31 and DAN3130 group, respectively. A SNP-index was calculated for each detected SNPs. An average SNP-index was computed for at least five consecutive SNPs in a 500 Kb sliding window and 50 Kb step. The Δ (SNP-index) was calculated combining the information of SNP-index in both pools and plotted against the chromosome position (Takagi et al. 2013).

## *4.7 Re-sequencing of candidate gene ZmCTM*

The sequence of the candidate gene ZmCTM was obtained from the B73 reference at the MaizeSequence database (http://ensembl.gramene.org/Zea_mays/Info/Index?db=core). Primers were designed using Primer Premier 5 software to cover the entire gene, including regions of the promoter, exons, introns, and 5' and 3' UTRs.

## *4.8 Transgenic analysis*

ZmCTM-edited maize plants carrying stop-gain mutation were generated using the CRISPR-Cas9 system (Li, Liu et al. 2017), with an SgRNA target sequence of CCAAGTTCATCTCCTGCAAG. After T1 seeds had been obtained, all transgenic plants were self-pollinated for two generations. Three individual lines of homozygous ZmCTM-edited maize were characterized for 2 years, and representative data from 1 year were shown in this study because the lines exhibited similar patterns. Young seeds (DAP25) and mature seeds of third-generation transgenic plants were collected for folate profiling. Plant heights and ear lengths of the mature wild-type and transgenic maize measured by a ruler.

## *4.9 Sequence alignment*

Phylogenetic evolutionary analyses were conducted using MEGA version 5 (Tamura, Peterson et al. 2011). The amino acid sequences of CTM in maize and other species were aligned using ClustalW (https://www.genome.jp/tools-bin/clustalw), and output by Espript3.0 (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi). Motif were predicted by MEME Version 5.3.3 （Multiple Em for Motif Elicitation, https://meme-suite.org/meme/tools/meme) using motif discovery mode.

**Acknowledgements**

This work was financially supported by the Ministry of Science and Technology of China (2016YFD0100503 to L.J.), the National Natural Science Foundation of China (31870283 to L.J.), Shanghai Agriculture Applied Technology Development Program (Z20180103 to L.J.), Beijing Natural Science Foundation (6172032 to B.W.).

**Authors' contributions**

W.G., L.J., T.L., and B.W. analyzed the data and wrote the manuscript. L.J., B.W., and C.Y.Z. designed the study. L.J., T.L and D.Y. developed the folate RIL population. B.W. and W.G. performed QTL analysis using different segregated populations, respectively. T.L. and W.W. constructed the libraries for sequencing. J.G. and H.W. assembled the genome and performed BSA analysis. L.J., T.L., F.M.S.A., X.W. and Q.L. performed the folate analysis and cob color analysis. L. J., T. L. performed the analysis on transgenic plants. H.Y.W., J.T. and C.Y.Z. supervised the project and revised the manuscript. All authors read and approved the final manuscript.

**Conflict of interest**

The authors declare that they have no competing financial interests.

# References

Abbasi, I. H. R., F. Abbasi, L. Wang, M. E. Abd El Hack, A. A. Swelum, R. Hao, J. Yao and Y. Cao (2018). "Folate promotes S-adenosyl methionine reactions and the microbial methylation cycle and boosts ruminants production and reproduction." AMB Express 8(1): 65.

Bae, S., C. M. Ulrich, L. B. Bailey, O. Malysheva, E. C. Brown, D. R. Maneval, M. L. Neuhouser, T. Y. D. Cheng, J. W. Miller, Y. Y. Zheng, L. R. Xiao, L. F. Hou, X. L. Song, K. Buck, S. A. A. Beresford and M. A. Caudill (2014). "Impact of folic acid fortification on global DNA methylation and one-carbon biomarkers in the Women's Health Initiative Observational Study cohort." Epigenetics 9(3): 396-403.

Bhutta, Z. A. and R. A. Salam (2012). "Global nutrition epidemiology and trends." Ann Nutr Metab 61 Suppl 1: 19-27.

Blancquaert, D., S. Storozhenko, K. Loizeau, H. De Steur, V. De Brouwer, J. Viaene, S. Ravanel, F. Rebeille, W. Lambert and D. Van Der Straeten (2010). "Folates and Folic Acid: From Fundamental Research Toward Sustainable Health." Critical Reviews In Plant Sciences 29(1): 14-35.

Blancquaert, D., J. Van Daele, S. Strobbe, F. Kiekens, S. Storozhenko, H. De Steur, X. Gellynck, W. Lambert, C. Stove and D. Van Der Straeten (2015). "Improving folate (vitamin B9) stability in biofortified rice through metabolic engineering." Nat Biotechnol 33(10): 1076-1078.

Broman, K. W., H. Wu, S. Sen and G. A. Churchill (2003). "R/qtl: QTL mapping in experimental crosses." Bioinformatics 19(7): 889-890.

Chang, X., R. S. DeFries, L. Liu and K. Davis (2018). "Understanding dietary and staple food transitions in China from multiple scales." PLoS One 13(4): e0195775.

Chen, J., B. Zeng, M. Zhang, S. Xie, G. Wang, A. Hauck and J. Lai (2014). "Dynamic transcriptome landscape of maize embryo and endosperm development." Plant Physiol 166(1): 252-264.

Chen, Z., B. Wang, X. Dong, H. Liu, L. Ren, J. Chen, A. Hauck, W. Song and J. Lai (2014). "An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F(2) maize population." BMC Genomics 15: 433.

Chen, Z. L., B. B. Wang, X. M. Dong, H. Liu, L. H. Ren, J. Chen, A. Hauck, W. B. Song and J. S. Lai (2014). "An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F-2 maize population." Bmc Genomics 15.

Cole, B. F., J. A. Baron, R. S. Sandler, R. W. Haile, D. J. Ahnen, R. S. Bresalier, G. McKeown-Eyssen, R. W. Summers, R. I. Rothstein, C. A. Burke, D. C. Snover, T. R. Church, J. I. Allen, D. J. Robertson, G. J. Beck, J. H. Bond, T. Byers, J. S. Mandel, L. A. Mott, L. H. Pearson, E. L. Barry, J. R. Rees, N. Marcon, F. Saibil, P. M. Ueland and E. R. Greenberg (2007). "Folic acid for the prevention of

colorectal adenomas - A randomized clinical trial." Jama-Journal Of the American Medical Association 297(21): 2351-2359.

De Lepeleire, J., S. Strobbe, J. Verstraete, D. Blancquaert, L. Ambach, R. G. F. Visser, C. Stove and D. Van Der Straeten (2018). "Folate Biofortification of Potato by Tuber-Specific Expression of Four Folate Biosynthesis Genes." Mol Plant 11(1): 175-188.

Diaz de la Garza, R., E. P. Quinlivan, S. M. Klaus, G. J. Basset, J. F. Gregory, 3rd and A. D. Hanson (2004). "Folate biofortification in tomatoes by engineering the pteridine branch of folate synthesis." Proc Natl Acad Sci U S A 101(38): 13720-13725.

Diaz de la Garza, R. I., J. F. Gregory, 3rd and A. D. Hanson (2007). "Folate biofortification of tomato fruit." Proc Natl Acad Sci U S A 104(10): 4218-4222.

Dong, W., Z. Cheng, X. Wang, B. Wang, H. Zhang, N. Su, C. Yamamaro, C. Lei, J. Wang, J. Wang, X. Zhang, X. Guo, F. Wu, H. Zhai and J. Wan (2011). "Determination of folate content in rice germplasm (Oryza sativa L.) using tri-enzyme extraction and microbiological assays." Int J Food Sci Nutr 62(5): 537-543.

Fazili, Z. and C. M. Pfeiffer (2013). "Accounting for an isobaric interference allows correct determination of folate vitamers in serum by isotope dilution-liquid chromatography-tandem MS." J Nutr 143(1): 108-113.

Fazili, Z., M. R. Sternberg, N. Paladugula, R. D. Whitehead, Jr., H. Chen and C. M. Pfeiffer (2014). "The loss of 5-methyltetrahydrofolate in human serum under suboptimal preanalytical conditions can only partially be recovered by an oxidation product." J Nutr 144(11): 1873-1879.

Gorelova, V., J. De Lepeleire, J. Van Daele, D. Pluim, C. Mei, A. Cuypers, O. Leroux, F. Rebeille, J. H. M. Schellens, D. Blancquaert, C. P. Stove and D. Van Der Straeten (2017). "Dihydrofolate Reductase/Thymidylate Synthase Fine-Tunes the Folate Status and Controls Redox Homeostasis in Plants." Plant Cell 29(11): 2831-2853.

Guo, Q. N., H. D. Wang, L. Z. Tie, T. Li, H. Xiao, J. G. Long and S. X. Liao (2017). "Parental Genetic Variants, MTHFR 677C>T and MTRR 66A>G, Associated Differently with Fetal Congenital Heart Defect." Biomed Research International.

Hanson, A. D. and J. F. Gregory, 3rd (2011). "Folate biosynthesis, turnover, and transport in plants." Annu Rev Plant Biol 62: 105-125.

Hilton, J. F., K. E. Christensen, D. Watkins, B. A. Raby, Y. Renaud, S. de la Luna, X. Estivill, R. E. MacKenzie, T. J. Hudson and D. S. Rosenblatt (2003). "The molecular basis of glutamate formiminotransferase deficiency." Hum Mutat 22(1): 67-73.

Huang, X., Q. Feng, Q. Qian, Q. Zhao, L. Wang, A. Wang, J. Guan, D. Fan, Q. Weng, T. Huang, G. Dong, T. Sang and B. Han (2009). "High-throughput genotyping by whole-genome resequencing." Genome Res 19(6): 1068-1076.

Jiang, L., Y. Liu, H. Sun, Y. Han, J. Li, C. Li, W. Guo, H. Meng, S. Li, Y. Fan and C. Zhang (2013). "The mitochondrial folylpolyglutamate synthetase gene is required for nitrogen utilization during early seedling development in arabidopsis." Plant Physiol 161(2): 971-989.

Jiang, L., W. Wang, T. Lian and C. Zhang (2017). "Manipulation of Metabolic Pathways to Develop Vitamin-Enriched Crops for Human Health." Front Plant Sci 8: 937.

Li, C., C. Liu, X. Qi, Y. Wu, X. Fei, L. Mao, B. Cheng, X. Li and C. Xie (2017). "RNA-guided Cas9 as an in vivo desired-target mutator in maize." Plant Biotechnol J 15(12): 1566-1576.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics 25(14): 1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and G. P. D. Proc (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics 25(16): 2078-2079.

Li, H., Z. Peng, X. Yang, W. Wang, J. Fu, J. Wang, Y. Han, Y. Chai, T. Guo, N. Yang, J. Liu, M. L. Warburton, Y. Cheng, X. Hao, P. Zhang, J. Zhao, Y. Liu, G. Wang, J. Li and J. Yan (2013). "Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels." Nat Genet 45(1): 43-50.

Lian, T., W. Guo, M. Chen, J. Li, Q. Liang, F. Liu, H. Meng, B. Xu, J. Chen, C. Zhang and L. Jiang (2015). "Genome-wide identification and transcriptional analysis of folate metabolism-related genes in maize kernels." BMC Plant Biol 15: 204.

Liang, Q., K. Wang, X. Liu, B. Riaz, L. Jiang, X. Wan, X. Ye and C. Zhang (2019). "Improved folate accumulation in genetically modified maize and wheat." J Exp Bot 70(5): 1539-1551.

Lu, X., J. Liu, W. Ren, Q. Yang, Z. Chai, R. Chen, L. Wang, J. Zhao, Z. Lang, H. Wang, Y. Fan, J. Zhao and C. Zhang (2018). "Gene-Indexed Mutations in Maize." Mol Plant 11(3): 496-504.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly and M. A. DePristo (2010). "The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data." Genome Research 20(9): 1297-1303.

Mickelson, S. M., C. S. Stuber, L. Senior and S. M. Kaeppler (2002). "Quantitative trait loci controlling leaf and tassel traits in a B73 x MO17 population of maize." Crop Science 42(6): 1902-1909.

Moore, E. M., D. Ames, A. G. Mander, R. P. Carne, H. Brodaty, M. C. Woodward, K. Boundy, K. A. Ellis, A. I. Bush, N. G. Faux, R. N. Martins, C. L. Masters, C. C. Rowe, C. Szoeke and D. A. Watters (2014). "Among Vitamin B12 Deficient Older People, High Folate Levels are Associated with Worse Cognitive Function: Combined Data from Three Cohorts." Journal Of Alzheimers Disease 39(3): 661-668.

Murigneux, A., D. Barloy, P. Leroy and M. Beckert (1993). "Molecular and morphological evaluation of doubled haploid lines in maize. 1. Homogeneity within DH lines." Theor Appl Genet 86(7): 837-842.

Naqvi, S., C. Zhu, G. Farre, K. Ramessar, L. Bassie, J. Breitenbach, D. Perez Conesa, G. Ros, G. Sandmann, T. Capell and P. Christou (2009). "Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways." Proc Natl Acad Sci U S A 106(19): 7762-7767.

Olmstead, D. L., B. A. Nault and A. M. Shelton (2016). "Biology, Ecology, and Evolving Management of Helicoverpa zea (Lepidoptera: Noctuidae) in Sweet Corn in the United States." J Econ Entomol 109(4): 1667-1676.

Pan, Q. C., Y. C. Xu, K. Li, Y. Peng, W. Zhan, W. Q. Li, L. Li and J. B. Yan (2017). "The Genetic Basis of Plant Architecture in 10 Maize Recombinant Inbred Line Populations." Plant Physiology 175(2): 858-873.

Peiffer, J. A., M. C. Romay, M. A. Gore, S. A. Flint-Garcia, Z. W. Zhang, M. J. Millard, C. A. C. Gardner, M. D. McMullen, J. B. Holland, P. J. Bradbury and E. S. Buckler (2014). "The Genetic Architecture Of Maize Height." Genetics 196(4): 1337-+.

Price, A. H. (2006). "Believe it or not, QTLs are accurate!" Trends Plant Sci 11(5): 213-216.

Rader, J. I. and B. O. Schneeman (2006). "Prevalence of neural tube defects, folate status, and folate fortification of enriched cereal-grain products in the United States." Pediatrics 117(4): 1394-1399.

Raihan, M. S., J. Liu, J. Huang, H. Guo, Q. C. Pan and J. B. Yan (2016). "Multi-environment QTL analysis of grain morphology traits and fine mapping of a kernel-width QTL in Zheng58 x SK maize population." Theoretical And Applied Genetics 129(8): 1465-1477.

Ringling, C. and M. Rychlik (2017). "Simulation of Food Folate Digestion and Bioavailability of an Oxidation Product of 5-Methyltetrahydrofolate." Nutrients 9(9).

Robbins, M. L., P. H. Wang, R. S. Sekhon and S. Chopra (2009). "Gene Structure Induced Epigenetic Modifications of pericarp color1 Alleles of Maize Result in Tissue-Specific Mosaicism." Plos One 4(12).

Sekhon, R. S., H. Lin, K. L. Childs, C. N. Hansey, C. R. Buell, N. de Leon and S. M. Kaeppler (2011). "Genome-wide atlas of transcription during maize development." Plant J 66(4): 553-563.

Shahid, M., T. Lian, X. Wan, L. Jiang, L. Han, C. Zhang and Q. Liang (2020). "Folate monoglutamate in cereal grains: Evaluation of extraction techniques and determination by LC-MS/MS." Journal of Food Composition and Analysis 91: 103510.

Silver, M. J., N. J. Kessler, B. J. Hennig, P. Dominguez-Salas, E. Laritsky, M. S. Baker, C. Coarfa, H. Hernandez-Vargas, J. M. Castelino, M. N. Routledge, Y. Y.

Gong, Z. Herceg, Y. S. Lee, K. Lee, S. E. Moore, A. J. Fulford, A. M. Prentice and R. A. Waterland (2015). "Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment." Genome Biol 16: 118.

Stelpflug, S. C., R. S. Sekhon, B. Vaillancourt, C. N. Hirsch, C. R. Buell, N. de Leon and S. M. Kaeppler (2016). "An Expanded Maize Gene Expression Atlas based on RNA Sequencing and its Use to Explore Root Development." Plant Genome 9(1).

Storozhenko, S., V. De Brouwer, M. Volckaert, O. Navarrete, D. Blancquaert, G. F. Zhang, W. Lambert and D. Van Der Straeten (2007). "Folate fortification of rice by metabolic engineering." Nat Biotechnol 25(11): 1277-1279.

Strobbe, S. and D. Van Der Straeten (2017). "Folate biofortification in food crops." Curr Opin Biotechnol 44: 202-211.

Suh, J. R., A. K. Herbig and P. J. Stover (2001). "New perspectives on folate catabolism." Annual Review Of Nutrition 21: 255-282.

Takagi, H., A. Abe, K. Yoshida, S. Kosugi, S. Natsume, C. Mitsuoka, A. Uemura, H. Utsushi, M. Tamiru, S. Takuno, H. Innan, L. M. Cano, S. Kamoun and R. Terauchi (2013). "QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations." Plant Journal 74(1): 174-183.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar (2011). "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods." Mol Biol Evol 28(10): 2731-2739.

Vales, M. I., C. C. Schon, F. Capettini, X. M. Chen, A. E. Corey, D. E. Mather, C. C. Mundt, K. L. Richardson, J. S. Sandoval-Islas, H. F. Utz and P. M. Hayes (2005). "Effect of population size on the estimation of QTL: a test using resistance to barley stripe rust." Theor Appl Genet 111(7): 1260-1270.

Wan, X., L. D. Han, M. Yang, H. Y. Zhang, C. Y. Zhang and P. Hu (2019). "Simultaneous extraction and determination of mono-/polyglutamyl folates using high-performance liquid chromatography-tandem mass spectrometry and its applications in starchy crops." Anal Bioanal Chem 411(13): 2891-2904.

Wang, B. B., Y. B. Zhu, J. J. Zhu, Z. P. Liu, H. Liu, X. M. Dong, J. J. Guo, W. Li, J. Chen, C. Gao, X. M. Zheng, L. Z. E, J. S. Lai, H. M. Zhao and W. B. Song (2018). "Identification and Fine-Mapping of a Major Maize Leaf Width QTL in a Re-sequenced Large Recombinant Inbred Lines Population." Frontiers In Plant Science 9.

Waskiewicz, A., W. Piotrowski, G. Broda, A. Sobczyk-Kopciol and R. Ploski (2011). "Impact of MTHFR C677T gene polymorphism, folate, B6 and B12 vitamin intake on homocysteine concentration in Polish adult population." European Heart Journal 32: 720-720.

Winter, D., B. Vinegar, H. Nahal, R. Ammar, G. V. Wilson and N. J. Provart (2007). "An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets." PLoS One 2(8): e718.

Xu, G., X. Wang, C. Huang, D. Xu, D. Li, J. Tian, Q. Chen, C. Wang, Y. Liang, Y. Wu, X. Yang and F. Tian (2017). "Complex genetic architecture underlies maize tassel domestication." New Phytol 214(2): 852-864.

Yang, X. H., S. B. Gao, S. T. Xu, Z. X. Zhang, B. M. Prasanna, L. Li, J. S. Li and J. B. Yan (2011). "Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize." Molecular Breeding 28(4): 511-526.

Yu, H., W. Xie, J. Wang, Y. Xing, C. Xu, X. Li, J. Xiao and Q. Zhang (2011). "Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers." PLoS One 6(3): e17595.

Zhang, H., A. B. Jha, D. De Silva, R. W. Purves, T. D. Warkentin and A. Vandenberg (2019). "Improved folate monoglutamate extraction and application to folate quantification from wild lentil seeds by ultra-performance liquid chromatography-selective reaction monitoring mass spectrometry." J Chromatogr B Analyt Technol Biomed Life Sci 1121: 39-47.

# *Supplementary Materials*

**Table S4-1.** Filtered candidate genes in *q5T-a* and *q5T-b*

| Position (Chr 5) | Gene ID | Annotation |
|---|---|---|
| 1565340..1568691 | GRMZM2G125300 | Ribosomal protein S21e |
| 1603428..1605321 | GRMZM2G010861 | Acyl-CoA binding protein 4 |
| 1608896..1612025 | GRMZM2G326195 | Expressed protein |
| 1621338..1624924 | GRMZM2G326272 | Translocon at the outer envelope membrane of chloroplasts 159 |
| 1682058..1686853 | GRMZM2G004140 | Expressed protein |
| 1822186..1829125 | GRMZM2G138523 | Expressed protein |
| 1829948..1838934 | GRMZM2G138511 | Chaperone DnaJ-domain superfamily protein |
| 1846096..1853085 | GRMZM2G138421 | Squamosa promoter-binding protein-like 12 |
| 1898216..1902113 | GRMZM2G100419 | Expressed protein |
| 2116106..2121486 | GRMZM2G121826 | Protein kinase protein with tetratricopeptide repeat domain |
| 2126865..2129228 | GRMZM2G121840 | S-adenosyl-L-methionine-dependent methyltransferase |
| 2181763..2184186 | GRMZM2G085301 | Major facilitator superfamily protein |
| 2259767..2263658 | GRMZM2G144008 | GTPase activity |
| 2265501..2268450 | GRMZM2G144020 | Expressed protein |
| 2275248..2277812 | GRMZM2G445169 | Barwin-like endoglucanases superfamily protein |
| 2284054..2290332 | GRMZM2G144042 | Protein kinase superfamily protein |
| 2350546..2357084 | GRMZM2G003108 | SEC14 cytosolic factor / phosphoglyceride transfer family protein |
| 2358646..2361757 | GRMZM2G003068 | Zinc knuckle (CCHC-type) family protein |
| 2362606..2365060 | GRMZM2G002825 | Actin depolymerizing factor 4 |
| 2387257..2400960 | GRMZM2G002765 | NAD(P)-binding Rossmann-fold superfamily protein |
| 2524531..2535342 | GRMZM2G458266 | Transducin/WD40 repeat-like superfamily protein |
| 2536464..2538475 | GRMZM2G160046 | Expressed protein |
| 16701221..16703382 | GRMZM5G895064 | DNA-directed RNA polymerase family protein |
| 17077279..17083849 | GRMZM2G018943 | NagB/RpiA/CoA transferase-like superfamily protein |
| 17123432..17138841 | GRMZM2G095239 | Telomere repeat binding factor 1 |
| 17290718..17296942 | GRMZM2G417223 | RNA-binding KH domain-containing protein |
| 17340297..17343212 | GRMZM2G023347 | ABI3-interacting protein 3 |
| 17881123..17884825 | GRMZM2G046676 | SNF7 family protein |

| | | |
|---|---|---|
| 18192142..18193721 | GRMZM2G060940 | 2-Oxoglutarate and Fe (II)-dependent oxygenase superfamily protein |
| 18358282..18370223 | GRMZM2G067306 | Ribosomal protein L25/Gln-tRNA synthetase anti-codon-binding domain |
| 18469442..18472522 | GRMZM2G066867 | Sucrose nonfermenting 1(SNF1)-related protein kinase |
| 18473592..18476671 | GRMZM2G066996 | Serine protease inhibitor (SERPIN) family protein |
| 18650274..18656914 | GRMZM2G044900 | Seed maturation protein PM23 |
| 18789191..18791839 | GRMZM2G087326 | Reversibly glycosylated polypeptide 2 |
| 19014015..19016625 | GRMZM2G045192 | No annotation |
| 19069664..19077156 | GRMZM2G375984 | DEK domain-containing chromatin associated protein |
| 19145071..19148243 | GRMZM2G074898 | Ribosomal protein L24e family protein |
| 19207810..19212507 | GRMZM2G081221 | SMAD/FHA domain-containing protein |
| 19263506..19269985 | GRMZM2G403149 | GTP-binding family protein |
| 19269959..19277452 | GRMZM2G403218 | FRIGIDA-like protein |
| 19460235..19471332 | GRMZM2G386714 | Lysyl-tRNA synthetase 1 |
| 19478193..19481880 | GRMZM2G086904 | RNA polymerase Rpb6 |
| 19696664..19698973 | GRMZM2G124863 | Transferases, folic acid binding |
| 19939913..19945652 | GRMZM2G001631 | Dicer-like 2 |
| 20784279..20795347 | GRMZM2G002874 | RNA-binding protein 45A |
| 21005810..21011591 | GRMZM2G024054 | Ubiquitin-protein ligase 2 |
| 21337753..21339972 | GRMZM5G871425 | Embryo defective 2016 |
| 21366640..21370672 | GRMZM5G823318 | Aspartate kinase 1 |
| 21459531..21465540 | GRMZM2G413544 | Protein kinase superfamily protein |
| 21592571..21599690 | GRMZM2G073934 | Tetratricopeptide repeat (TPR)-like superfamily protein |
| 21601089..21603875 | GRMZM2G074169 | Cytochrome B5 isoform E |
| 21719063..21724696 | GRMZM2G153815 | Mitochondrial HSO70 2, DnaK family protein |
| 21903403..21908619 | AC209835.4_FG004 | ChaC-like family protein |
| 21954101..21955171 | GRMZM2G096372 | Expressed protein |

**Table S4-2.** 5T profiles of the extreme-high and extreme-low groups from F7 seeds of the GEMS31×DAN3130 population.

| High 5T | | Low 5T | |
|---|---|---|---|
| Sample name | 5T (nmol/g DW) | Sample name | 5T (nmol/g DW) |
| A180-1 | 4.40±0.35 | E116-2 | 0.57±0.05 |
| A180-2 | 4.50±0.40 | E116-1 | 0.58±0.04 |
| A111-1 | 4.66±0.38 | E139-2 | 0.59±0.05 |
| A111-2 | 4.67±0.23 | E139-20 | 0.60±0.02 |
| B101-1 | 4.67±0.34 | C138-2 | 0.69±0.03 |
| A129-1 | 4.68±0.28 | B136-2 | 0.72±0.06 |
| D171-3 | 4.72±0.44 | B134-3 | 0.79±0.04 |
| B101-3 | 4.81±0.42 | B136-3 | 0.83±0.02 |
| A136-2 | 4.82±0.37 | B140-2 | 0.83±0.07 |
| B130-4 | 4.90±0.32 | C136-3 | 0.88±0.05 |
| C163-1 | 5.03±0.47 | B127-5 | 0.88±0.08 |
| B130-3 | 5.14±0.23 | B141-4 | 0.90±0.05 |
| A179-1 | 5.15±0.35 | C133-2 | 0.91±0.06 |
| D180-2 | 5.16±0.46 | C133-3 | 1.01±0.05 |
| B103-2 | 5.16±0.16 | C139-1 | 1.02±0.07 |
| B103-4 | 5.23±0.34 | C138-1 | 1.06±0.03 |
| C163-2 | 5.24±0.27 | B141-3 | 1.10±0.06 |
| A136-3 | 5.45±0.51 | E170-3 | 1.10±0.08 |
| A171-4 | 5.53±0.26 | C137-3 | 1.11±0.03 |
| A171-2 | 5.62±0.29 | B127-1 | 1.13±0.06 |
| F121-1 | 5.74±0.41 | B140-3 | 1.16±0.08 |
| D171-4 | 5.84±0.36 | E170-1 | 1.19±0.07 |
| F121-2 | 5.87±0.31 | C135-2 | 1.22±0.08 |
| E103-1 | 6.04±0.57 | C135-1 | 1.26±0.12 |
| A179-2 | 6.12±0.15 | B134-5 | 1.26±0.05 |
| D180-3 | 6.22±0.56 | B179-3 | 1.30±0.07 |
| E103-3 | 6.25±0.21 | C136-1 | 1.58±0.06 |
| A125-3 | 6.49±0.33 | C139-3 | 1.61±0.04 |
| A129-2 | 7.13±0.52 | B179-5 | 1.61±0.09 |
| A125-1 | 8.01±0.51 | C137-1 | 1.68±0.11 |

# 5

# Discussion and prospects

# 1 Conclusion

Maize is a major global staple crop that provides essential nutrients for human. To date, many studies have unravelled the regulatory mechanism of different metabolites such as sucrose, fructose, starch, fatty acids and proteins during maize kernel development. For example, the transcriptomic behaviour of maize plants during kernel formation and the regulatory roadmap for the biosynthesis of starch, fatty acids, and storage proteins in seeds have been summarized (Wang, Wang et al. 2012). Folate is a key soluble B vitamin that is indispensable in our daily diet. However, the regulatory mechanism of folate accumulation in maize kernel is complicated and poorly understood. Several research findings suggest that folate metabolism is associated with varieties of metabolic pathways. Here, we screen out two different maize inbred lines with folate variation as experimental material and perform the transcriptome analysis to find out those key pathways and genes related to folate accumulation. Such a low-cost strategy will enable us to preliminarily explore the closely related pathways of folate and provide the ideas for futural research direction. Meanwhile, we performed the genetic mapping of folate quantitative trait loci using a segregated population to identify the candidate genes involved in folate metabolism in maize seeds. Finally, transgenic experiments are carried out to analyze the function of the candidate gene to support our findings further.

## *1.1 The pathways associated with folate accumulation during kernel development*

Different maize inbred lines accumulate folate derivatives in different patterns during kernel development. In the development stage from DAP 24 to DAP 35, folate accumulation is mainly influenced by circumjacent pathways of folate biosynthesis like pyruvate metabolism, glutamate metabolism and serine metabolism. The pyruvate metabolism can provide PEP and formate for chorismate biosynthesis and source of OCM, respectively (Herrmann and Weaver 1999, Bykova, Stensballe et al. 2003, Wulfert, Schilasky et al. 2020). Glutamate has long been recognized as an essential substrate for folate biosynthesis and stability (Shane 1989). Serine is a crucial amino acid that has been well known to participate in one-carbon metabolism reactions. From DAP 35 to mature kernel, the stage of late seed maturation, not only pyruvate metabolism and glutamate metabolism but also the genes belonging to Acyl-CoA N-acetyltransferase superfamily protein and histone-lysine N-methyltransferase family are found. Several research results connect folate contents to these two specific protein families in the mammal (Upton, Smelt et al. 2000, Cao, Strnatka et al. 2010), while more experimental evidence is necessary to confirm their roles in plants. Our findings give a new insight into the Acyl-CoA N-acetyltransferase superfamily protein and histone-lysine N-methyltransferase family on folate accumulation in maize.

## *1.2 The genes associated with folate accumulation in different inbred lines*

In addition, the variation of folate contents in the mature kernel between different inbred lines might be associated with genes in 1) two branches of folate biosynthesis,

2) serine/THF/5-M-THF cycle and 3) conversion of THF monoglutamate to THF polyglutamate. The importance of ρ-ABA branch and pteridine branch for folate biosynthesis have been reported and summarized in our general introduction. Here, the genes *ADCS* and *DHNA2*, which have been utilized in folate biofortification of plants by engineering approach, show a positive correlation with folate contents in our analysis. Gene expression of *SHMT7-1*, *SHMT7-2, GCST*, *MTHFR1* and *MS2*, which are involved in the serine/THF/5-M-THF cycle, shows a positive correlation with folate contents. *FPGS2* and *GGH* which are involved in the folate polyglutamylation and deglutamylation have been reported to be tightly associated with folate content in our study. All the above results demonstrate that these three processes may perform as critical regulators for folate biosynthesis in maize kernels.

Furthermore, two folate QTLs on chromosome 5 were obtained from a segregated population using whole-exome sequencing technology and bulk segregant analysis. Then the new candidate gene, named *ZmCTM*, was identified with annotation of the folate-associated gene. Knock-out of the *ZmCTM* significantly enhanced 5-M-THF accumulation around three-fold in maize kernel. We put forward the hypothesis that the ZmCTM enzyme could convert 5-M-THF to other folate derivatives in folate metabolism of maize kernel, and more evidence of molecular mechanisms is needed to be investigated. Our findings suggest that those metabolic pathways and essential genes are related to folate accumulation during kernel formation. Our results contribute to further understanding the mechanisms of folate accumulation, and to identify candidate genes and pathways displaying some potential for folate biofortification.

# 2. Discussion

Crop vitamins are essential for human health, and insufficient vitamins intake threatens billions of people. Biofortification aims to enhance the micronutrient in food crops to alleviate the micronutrient deficiencies in poor and developing countries. Folate is one of the crucial water-soluble vitamin B, and its deficiency has been known associated with neural tube defects in newborns and some cancers in adults. Until recently, the research works on folate biofortification have been investigated to address the prevalence of folate deficiency. However, some challenges are arising that we must solve.

First, the cost and technology of folate measurement by LC/MS is a challenge that restricts the folate research progress. In our works, screening out the different inbred lines with folate-variation and the construction of the RIL population takes the expensive cost of detection and is hugely time-consuming. Due to technology limitations, the initially detected folate derivatives for QTL analysis can not distinguish the MeFox and 5-F-THF. After improving the LC/MS detection method, MeFox can be separated from 5-F-THF in lentils and other plants (Wan, Han et al. 2019, Zhang, De Silva et al. 2021). To solve it, lateral flow dipstick immunoassay is developed later to evaluate the folate levels in maize that will help roughly and fastly screen out high-/low-folate lines (Liang, Yi et al. 2017). Also, the measurement

method is updated for more accurate detection (Wan, Han et al. 2019, Shahid, Lian et al. 2020). More efficient and accurate phenotyping can save time-consuming and experiment costs and make gene mapping results more convenient and convincing. In our study, we used the whole-exome sequencing (WES) strategy instead of the whole-genome sequencing (WGS) for QTL analysis. The WGS can assess all regions of genomes, which is regarded as the gold standard for sequencing. WES is a probe-based enrichment approach that targets to specific region. Compared to the WGS, the WES strategy represents a sequencing option that is cheaper than WGS. Meanwhile, the WES strategy focus on the exon region that can reduce the complexity of data analysis and improve the sequencing depth of target regions (Singh, Singh et al. 2012, Kaur and Gaikwad 2017, Lu, Liu et al. 2018). On the other hand, the WES strategy is not limited by transcript abundance, and does not depend on the tissue/stage collected compared to the RNA sequencing strategy. In our study, the WES strategy was particularly suitable to identify the QTL. However, the limitation of WES in our research is also obvious. First, even both WGS and WES can obtain the functional molecular markers, while the WES can not obtain the SNPs which are important for controlling the transcriptional regulation. Second, those regions, which are not considered to be gene exons currently, will be missed in the WES strategy.

Second, our previous results indicate that 5-M-THF is the major form of folate derivative during maize kernel development (Lian, Guo et al. 2015, Song, Yu et al. 2021). The concentration of 5-M-THF reaches its peak at DAP 12, then sharply decreases in the mature seed. The folate variation between inbred lines DAN3130 and JI63 mainly comes from 5-M-THF in our study. Indeed, the 5-M-THF account for up to 89% of total folate in transgenic plants. In maize, the folate biofortification results in 2 to 4-fold increase in folate content (Naqvi, Zhu et al. 2009, Liang, Wang et al. 2019). In our results, a new candidate gene *ZmCTM* is in QTL *q5T-b*, and it's function is connected to folate contents by trangenic function analysis. The loss function of *ZmCTM* results in around three-fold increased 5-M-THF in maize kernel providing a potential prospect of folate biofortification. Transferring the *ADCS* (from Arabidopsis) and *GCHI* (from mice) into tomatoes is triggering an overproduction of monoglutamate folate. High levels of monoglutamate folate may feedback inhibit the activity of FPGS (Diaz de la Garza, Gregory et al. 2007). In our study, not only monoglutamated 5-M-THF but also polyglutamated 5-M-THF were enhanced in *ZmCTM* mutant. The enriched polyglutamated 5-M-THF may also contribute to the folate stability in maize kernel. Additional conversion activity is still required to reveal the molecular mechanism. Since 5-M-THF provides the methyl groups for the reaction of homocysteine to methionine. A gene encoding S-adenosyl-L-methionine (SAM)-dependent methyltransferase, which is considered to be associated with folate metabolism, is located in the QTL region *q5T-a*. Coincidently, GWAS of seed folate content in common bean has previously identified a candidate gene belonging to the SAM-dependent methyltransferase super family (Martin, Torkamaneh et al. 2021).

Third, transcriptome analyses have been applied in the identification of significant pathways and regulatory factors. For example, transcriptome analysis can be used to identify the hub genes involved in nitrogen metabolism for improving nitrogen use

efficiency in *Gossypium* (cotton) (Iqbal, Dong et al. 2020), to identify the candidate genes of terpenoid and phenylpropanoid metabolism in *Ferula assafoetida* (Amini, Naghavi et al. 2019), and to investigate the unique molecular mechanism and transcriptional regulation of sucrose metabolism occurred in oil tubers of *Cyperus esculentus* (Yang, Liu et al. 2018). In our results, folate biosynthesis and breakdown may be associated with glutamate metabolism, pyruvate metabolism and serine/glycine metabolism, which leads to a complex regulatory mechanism of folate accumulation. Similarly, the waxy maize endosperms from DAP 15 to DAP 27 are collected for folate profiling and transcriptome analysis. The glutamate metabolism pathway is enriched at DAP 18. Besides, total 518 transcription factors (TFs) are found from WGCNA results (Song, Yu et al. 2021). In comparison to our research, the transcriptome analysis of waxy maize endosperm focuses on the early stages of endosperm development and attempts to identify TFs associated with folate contents. Our study investigates some folate genes expression levels positively corelated to folate accumulation. Among those genes, *ADCS*, *DHNA2*, *GGH* and *FPGS2*, which are involved in folate biosynthesis, have proven to help boosting folate enhancement with transgenic approaches. The other genes related to OCM like *GCST*, *SHMT7-2*, *MTHFR1* and *MS2* also provided new insight into folate biofortification. Meanwhile, these genes may can be used as potential indicators for screening out high/low folate lines in maize through their variational expression levels. Even though the bioinformatic analysis suggests essential metabolic pathways and genes that might affect the folate accumulation in maize kernel. Experimental evidence is required to demonstrate their roles in folate metabolism. Nowadays, the purpose of folate biofortification is to strengthen the folate biosynthesis pathway. On the other hand, some new approaches to achieving folate biofortification without genetic modification are developed. For instance, foliar salicylic acid treatment and the addition of phenylalanine can boost folate accumulation in Arabidopsis and spinach, respectively (Watanabe, Ohtani et al. 2017, Puthusseri, Divya et al. 2018). Our study investigates the folate accumulation in maize kernel using transcriptome analysis that provides new insight into folate biofortification.

The two-genes strategy is insufficient for folate improvement in some species, like Arabidopsis, potatoes, maize and cassava. More candidate genes are proposed for three-genes or multi-genes strategy to boost folate accumulation in the staple crops. Both known folate-related genes which we are concerned with and the new candidate gene *ZmCTM* can be utilized for improving folate accumulation. On the other side, folate biofortification should not be limited to transgenic methods. Those candidate metabolism pathways and metabolites shall be further investigated to enhance folate levels via non-gene edited approaches. At the same time, the folate precursors are also required to be concerned in the biofortification field. Normally, huge amounts of folate precursors like pteridine are overproduced (more than 1,000-fold increase) in folate biofortified tomatoes (Diaz de la Garza, Quinlivan et al. 2004, Hossain, Rosenberg et al. 2004). More research is also needed to assess the safety of pteridines accumulating in transgenic crops. Generally speaking, even though we find out that some new pathways and genes might play a function in regulating folate accumulation during

kernel formation. However, their detailed mechanism and biochemistry process is unclear. Continued research is required to excavate their potential value in conventional breeding and engineering breeding. These newest results provide valuable references for folate research and the new direction and insights to study folate metabolism in plants.

# 3. Prospect

## *3.1 Future research on identified pathways and genes*

We have generated a gene-indexed maize mutant collection through ethyl methanesulfonate (EMS) mutagenesis and detected the mutations by combining exome capture and next-generation sequencing. Such gene-indexed EMS mutant population has several advantages. First, it is easy to generate mutations that cover as much of the genome as possible due to multiple mutations are simultaneously present in a single genome. Second, both strong and weak alleles are very likely to be isolated from the mutant population. Third, it enables functional analysis of a specific domain/motif within a gene (Lu, Liu et al. 2018). Such mutation library can help us to quickly and efficiently screen out the target genes. The EMS mutants of those candidate genes, which are involved in the pyruvate metabolism, glutamate metabolism, serine metabolism, GNAT-transcription factor family and histone-lysine N-methyltransferase family, will be utilized for analyzing folate profiling and gene transcriptional levels. To reach a conclusion, it is also necessary to clean the background by performing multiple backcrosses. Finally, key genes will be screened out for further functional analysis. (a) Gene encodes functional enzymes. The CRISPR/Cas9 mutants and overexpression lines will be used to investigate the candidate gene function in maize. Bioinformatic analysis such as molecular docking can be utilized to predict the substrate of enzyme reaction. Transcriptional levels of folate-related genes will be analyzed in transgenic lines to demonstrate the upregulated/downregulated genes. The SNPs information from the natural inbred population will be connected with folate profiling to evaluate their correlation. The potential effect of alleles will be investigated using overexpression lines with different allele sites. Moreover, enzyme activity assay analysis of protein expressed *in vitro* with different alleles will also contribute to the analysis of gene function. The key amino acid sites will be identified for breeding purposes in order to cultivate maize lines with high folate content. (b) Gene encodes TFs. The natural inbred population representing the global diversity of maize germplasm will be used to identify SNPs. Conserved DNA binding domain, transcription activity assay, ChIP-sequencing and yeast one-hybrid assays will be analyzed to identify the target genes regulated by candidate TFs; Meanwhile, the yeast two-hybrid assays will be performed to identify the proteins that interact with candidate TFs; The subcellular localization of candidate TFs and interacted proteins will be identified to further understand the mechanism of controlling folate phenotype. (c) The Arabidopsis mutants with disruption in candidate metabolism pathways will be used to unravel their impact on folate. Next, isotopic tracing of the metabolite flux will be combined with quadrupole time-of-flight mass

spectrometry (QTOF-MS) to determine the metabolite flux from candidate pathways to folate metabolism. Not only transcriptome analysis but also proteome and metabolome analysis are required for our futural research.

## *3.2 The potential value of ZmCTM*

The loss function of *ZmCTM* causes an increased 5-M-THF. The conserved domain analysis of ZmCTM in many different plants suggests that *ZmCTM* is highly conserved in crop species. This indicates that blocking the metabolic flow towards MeFox by the introduction of a loss-of-function mutation into CTM can significantly improve 5-M-THF accumulation, thus representing a new strategy for folate biofortification to benefit human health, especially for the people with only partial MTHFR activity (Rommer, Zschocke et al. 2017). Many cereal crops contain huge amounts of MeFox, while the function of MeFox in plants is unknown (Ringling and Rychlik 2017). The loss of ZmCTM function in maize contains a significantly reduced MeFox, which can be used to figure out the role of MeFox in plants.

## *3.3 The fine-mapping of the other QTL*

The re-sequencing of population parents has been done to mine their genomic sequence variation for marker primer design. The phenotyping and genotyping of the heterogeneous inbred family will be performed to narrow down the candidate QTL region *q5T-a*. Then we use the maize EMS mutants of candidate genes to verify the variation of folate concentration. Finally, the gene function analysis will be completed to understand the regulatory mechanism of folate accumulation in maize.

# References

Amini, H., M. R. Naghavi, T. Shen, Y. Wang, J. Nasiri, I. A. Khan, O. Fiehn, P. Zerbe and J. N. Maloof (2019). "Tissue-Specific Transcriptome Analysis Reveals Candidate Genes for Terpenoid and Phenylpropanoid Metabolism in the Medicinal Plant Ferula assafoetida." G3 (Bethesda) 9(3): 807-816.

Bykova, N. V., A. Stensballe, H. Egsgaard, O. N. Jensen and I. M. Moller (2003). "Phosphorylation of formate dehydrogenase in potato tuber mitochondria." J Biol Chem 278(28): 26021-26030.

Cao, W., D. Strnatka, C. A. McQueen, R. J. Hunter and R. P. Erickson (2010). "N-acetyltransferase 2 activity and folate levels." Life Sci 86(3-4): 103-106.

Diaz de la Garza, R. I., J. F. Gregory, 3rd and A. D. Hanson (2007). "Folate biofortification of tomato fruit." Proc Natl Acad Sci U S A 104(10): 4218-4222.

Fazili, Z., M. R. Sternberg, N. Potischman, C. Y. Wang, R. J. Storandt, L. Yeung, S. Yamini, J. J. Gahche, W. Juan, Y. P. Qi, N. Paladugula, G. Gabey and C. M. Pfeiffer (2020). "Demographic, Physiologic, and Lifestyle Characteristics Observed with Serum Total Folate Differ Among Folate Forms: Cross-Sectional Data from Fasting Samples in the NHANES 2011-2016." J Nutr 150(4): 851-860.

Herrmann, K. M. and L. M. Weaver (1999). "The Shikimate Pathway." Annu Rev Plant Physiol Plant Mol Biol 50: 473-503.

Hossain, T., I. Rosenberg, J. Selhub, G. Kishore, R. Beachy and K. Schubert (2004). "Enhancement of folates in plants through metabolic engineering." Proc Natl Acad Sci U S A 101(14): 5158-5163.

Iqbal, A., Q. Dong, X. Wang, H. Gui, H. Zhang, X. Zhang and M. Song (2020). "Transcriptome Analysis Reveals Differences in Key Genes and Pathways Regulating Carbon and Nitrogen Metabolism in Cotton Genotypes under N Starvation and Resupply." Int J Mol Sci 21(4).

Kaur, P. and K. Gaikwad (2017). "From Genomes to GENE-omes: Exome Sequencing Concept and Applications in Crop Improvement." Front Plant Sci 8: 2164.

Liang, Q., C. Yi, L. Jiang, G. Tan, C. Zhang and B. Wang (2017). "Development of a lateral flow dipstick immunoassay for evaluation of folate levels in maize." Anal Bioanal Chem 409(24): 5655-5660.

Liang, Q., K. Wang, X. Liu, B. Riaz, L. Jiang, X. Wan, X. Ye and C. Zhang (2019). "Improved folate accumulation in genetically modified maize and wheat." J Exp Bot 70(5): 1539-1551.

Lu, X., J. Liu, W. Ren, Q. Yang, Z. Chai, R. Chen, L. Wang, J. Zhao, Z. Lang, H. Wang, Y. Fan, J. Zhao and C. Zhang (2018). "Gene-Indexed Mutations in Maize." Mol Plant 11(3): 496-504.

Martin, C. J., D. Torkamaneh, M. Arif and K. P. Pauls (2021). "Genome-Wide Association Study of Seed Folate Content in Common Bean." Front Plant Sci 12: 696423.

Naqvi, S., C. Zhu, G. Farre, K. Ramessar, L. Bassie, J. Breitenbach, D. Perez Conesa, G. Ros, G. Sandmann, T. Capell and P. Christou (2009). "Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways." Proc Natl Acad Sci U S A 106(19): 7762-7767.

Puthusseri, B., P. Divya, V. Lokesh, G. Kumar, M. A. Savanur and B. Neelwarne (2018). "Novel Folate Binding Protein in Arabidopsis Expressed during Salicylic Acid-Induced Folate Accumulation." J Agric Food Chem 66(2): 505-511.

Ringling, C. and M. Rychlik (2017). "Origins of the difference between food folate analysis results obtained by LC-MS/MS and microbiological assays." Anal Bioanal Chem 409(7): 1815-1825.

Rommer, P. S., J. Zschocke, B. Fowler, M. Fodinger, V. Konstantopoulou, D. Moslinger, E. Stogmann, E. Suess, M. Baumgartner, E. Auff and G. Sunder-Plassmann (2017). "Manifestations of neurological symptoms and thromboembolism in adults with MTHFR-deficiency." J Neurol Sci 383: 123-127.

Shahid, M., T. Lian, X. Wan, L. Jiang, L. Han, C. Zhang and Q. Liang (2020). "Folate monoglutamate in cereal grains: Evaluation of extraction techniques and determination by LC-MS/MS." Journal of Food Composition and Analysis 91: 103510.

Shane, B. (1989). "Folylpolyglutamate synthesis and role in the regulation of one-carbon metabolism." Vitam Horm 45: 263-335.

Singh, D., P. K. Singh, S. Chaudhary, K. Mehla and S. Kumar (2012). "Exome sequencing and advances in crop improvement." Adv Genet 79: 87-121.

Song, L., D. Yu, H. Zheng, G. Wu, Y. Sun, P. Li, J. Wang, C. Wang, B. Lv and X. Tang (2021). "Weighted gene co-expression network analysis unveils gene networks regulating folate biosynthesis in maize endosperm." 3 Biotech 11(10): 441.

Upton, A., V. Smelt, A. Mushtaq, R. Aplin, N. Johnson, H. Mardon and E. Sim (2000). "Placental arylamine N-acetyltransferase type 1: potential contributory source of urinary folate catabolite p-acetamidobenzoylglutamate during pregnancy." Biochim Biophys Acta 1524(2-3): 143-148.

Wan, X., L. D. Han, M. Yang, H. Y. Zhang, C. Y. Zhang and P. Hu (2019). "Simultaneous extraction and determination of mono-/polyglutamyl folates using high-performance liquid chromatography-tandem mass spectrometry and its applications in starchy crops." Anal Bioanal Chem 411(13): 2891-2904.

Wang, G., G. Wang, F. Wang and R. Song (2012). A Transcriptional Roadmap for Seed Development in Maize. Seed Development: OMICS Technologies toward Improvement of Seed Quality and Crop Yield: 81-97.

Wulfert, S., S. Schilasky and S. Krueger (2020). "Transcriptional and Biochemical Characterization of Cytosolic Pyruvate Kinases in Arabidopsis thaliana." Plants (Basel) 9(3).

Watanabe, S., Y. Ohtani, Y. Tatsukami, W. Aoki, T. Amemiya, Y. Sukekiyo, S. Kubokawa and M. Ueda (2017). "Folate Biofortification in Hydroponically Cultivated Spinach by the Addition of Phenylalanine." J Agric Food Chem 65(23): 4605-4610.

Yang, Z., D. Liu and H. Ji (2018). "Sucrose metabolism in developing oil-rich tubers of Cyperus esculentus: comparative transcriptome analysis." BMC Plant Biol 18(1): 151.

Zhang, H., D. De Silva, D. Dissanayaka, T. D. Warkentin and A. Vandenberg (2021). "Validated B vitamin quantification from lentils by selected reaction monitoring mass spectrometry." Food Chem 359: 129810.

# Appendix – Publications

1. Guo W*, **Lian T***, et al. Genetic mapping of folate QTLs using a segregated population in maize. J Integr Plant Biol. 2019, 61(6): 675-690. (*: equal contribution)
2. **Lian T***, Wang X*, et al. Comparative transcriptome analysis reveals mechanisms of folate accumulation in maize grains. Int J of Mol Sci. 2022, 23(3): 1708 (*: equal contribution)
3. Mehmood S, **Lian T**, et al. Folate monoglutamate in cereal grains: Evaluation of extraction techniques and determination by LC-MS/MS. J Food Compost Anal. 2020, 91:103510
4. Jiang L, Wang W, **Lian T**, Zhang C. Manipulation of metabolic pathways to develop vitamin-enriched crops for human health. Front Plant Sci. 2017, 8:937.
5. Wang L, Kong D, …, **Lian T**, et al. Tetrahydrofolate modulates floral transition through epigenetic silencing. Plant Physiol. 2017, 174(2): 1274-1284.
6. Jiang L*, Guo W*, …, **Lian T**, et al. Conversion of 5-methyl-tetrahydrofolate to MeFox facilitates folate biofortification in maize. (preprint)
7. Patent: Jiang L, Zhang C, **Lian T**, et al. Enhancing the contents of 5-M-THF by gene-edit method. (CN111041045B) 2020.