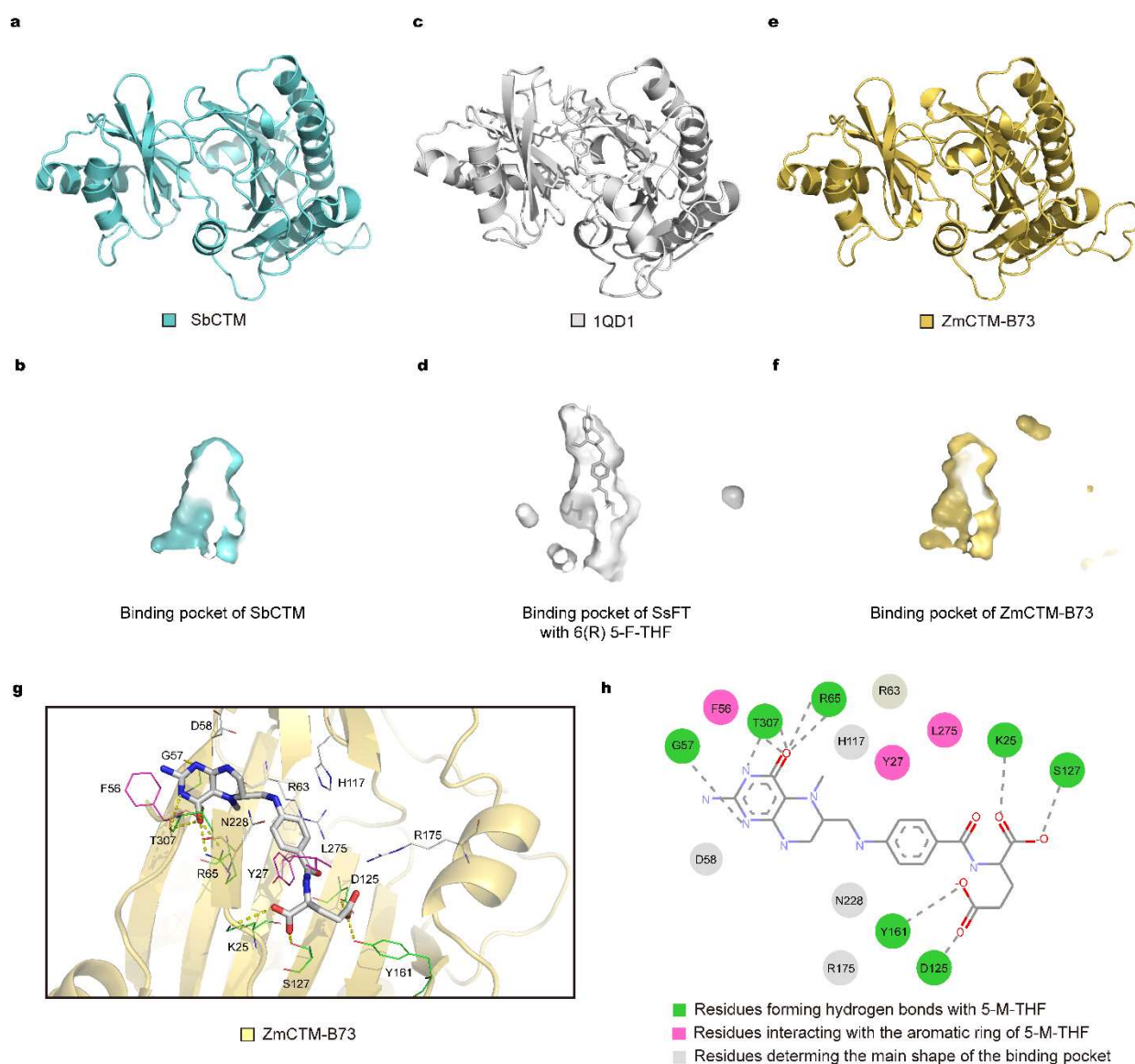


Supplemental material for “Catalysis of 5-methyltetrahydrofolate to MeFox facilitates folate biofortification in crops”

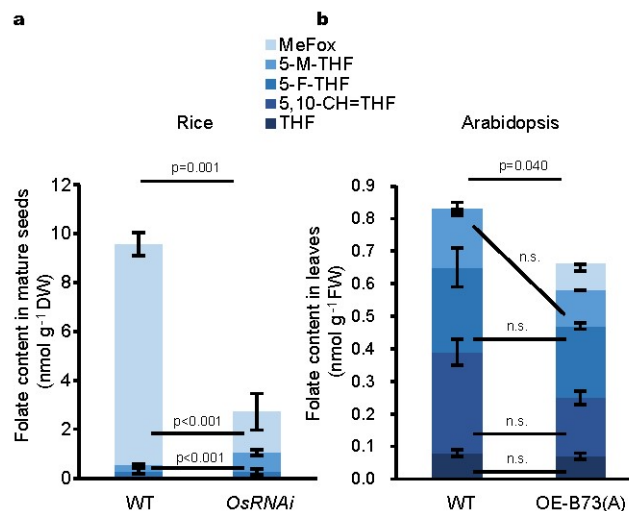
Extended data figures and tables



**Extended Data Fig. 1 Global structure and binding pocket of SbCTM (PDB: 7DYH), SsFT (PDB: 1QD1), and 3D model of ZmCTM-B73. Related to Fig. 2. (a) Global structure of SbCTM. (b) Binding pocket of SbCTM. (c) Global structure of SsFT. (d) Binding pocket of SsFT. (e) Global structure of ZmCTM-B73. (f) Binding pocket of ZmCTM-B73. (g) 3D binding model of 5M-THF with ZmCTM-B73. Residues in the binding pocket are shown as lines and 5-M-THF is shown as a**

12 grey stick. Hydrogen bonds are shown as dashed lines (h) Two-dimensional interaction map of 5M-  
 13 THF with binding pocket residues.

14



15

16 **Extended Data Fig. 2 Folate profiles of *OsCTM*-RNAi rice plants and overexpression of *ZmCTM*-**  
 17 ***B73* in *Arabidopsis*. Related to Fig. 4. (a) Folate profile (mean  $\pm$  SD of five biological replicates**  
 18 **with  $n = 3$  each) in mature seeds of wild-type (WT) and *OsCTM*-RNAi (*OsRNAi*) rice plants. *P*-value**  
 19 **(Student's *t*-test,  $P < 0.05$ ) represents *OsCTM*-RNAi compared with the wild-type rice. (b) Folate**  
 20 **profile (mean  $\pm$  SD of two biological replicates with  $n = 3$  each) in 30-day rosette leaves of *ZmCTM*-**  
 21 ***B73* transgenic (OE-B73) and wild-type *Arabidopsis* plants. A in the brackets after OE-B73**  
 22 **represents A-allele at S2069. *P*-value (Student's *t*-test,  $P < 0.05$ ) represents *ZmCTM*-*B73***  
 23 **overexpressor compared with the wild-type *Arabidopsis*. n.s., not significant.**

24

25 **Extended Data Table 1. Phenotypic variation and heritability analysis in the association panel**  
 26 **(n = 531). Related to Fig. 1.**

| Category               | Source of variation       | 5-F-THF-M (nmol g <sup>-1</sup> ) | 5-M-THF (nmol g <sup>-1</sup> ) |
|------------------------|---------------------------|-----------------------------------|---------------------------------|
| Descriptive Statistics | Range                     | 0.58-8.18                         | 0.19-0.79                       |
|                        | Mean ± SD.                | 1.66 ± 0.98                       | 0.33 ± 0.08                     |
| ANOVA <sup>a</sup>     | Environments              | 345.26 (2) **                     | 14.08 (2) **                    |
|                        | Lines                     | 4.63 (530) **                     | 0.07 (530) **                   |
|                        | Error                     | 1.17 (834)                        | 0.04 (833)                      |
|                        | Repeatability             | 0.76                              | 0.42                            |
|                        | 95% CI for H <sup>b</sup> | 0.68-0.84                         | 0.40-0.43                       |

27 <sup>a</sup>ANOVA, analysis of variance, showing mean square and degrees of freedom (in parentheses). The F-  
 28 test was applied to determine the significance level. Both environments and lines were regarded as  
 29 random effects.

30 <sup>b</sup>95% confidence interval for broad-sense heritability. \*\*, *P*-value < 0.01; SD, standard deviation.

31

32 **Extended Data Table 2. Genes associated with the content of 5-F-THF-M identified in the**  
 33 **genome-wide association analysis. Related to Fig. 1.**

| Lead SNP                     | Chromosome | Position <sup>a</sup> | Allele <sup>b</sup> | Frequency | <i>P</i> -value        | Gene ID <sup>c</sup> | Annotation <sup>d</sup>                                                    | R <sup>2</sup> (%) |
|------------------------------|------------|-----------------------|---------------------|-----------|------------------------|----------------------|----------------------------------------------------------------------------|--------------------|
| chr1.S_18977644              | 1          | 18977644              | A/ <u>G</u>         | 22/346    | 8.58×10 <sup>-8</sup>  | GRMZM2G031177        | putative MATE efflux family protein                                        | 5.4                |
| chr2.S_188526643             | 2          | 188526643             | <u>A</u> /G         | 297/71    | 8.48×10 <sup>-7</sup>  | GRMZM2G009406        | DNA binding with one finger2 MNB1A                                         | 4.5                |
| chr3.S_10281356              | 3          | 10281356              | C/ <u>T</u>         | 36/332    | 9.50×10 <sup>-7</sup>  | GRMZM2G118243        | pleiotropic drug resistance protein 6-like                                 | 8.7                |
| chr3.S_208614813             | 3          | 208614813             | <u>A</u> /C         | 343/25    | 1.78×10 <sup>-7</sup>  | GRMZM2G077607        | PR domain zinc finger protein 13-like isoform X2protein 13-like isoform X2 | 5                  |
| chr3.S_220643522             | 3          | 220643522             | <u>T</u> /G         | 338/30    | 3.86×10 <sup>-7</sup>  | GRMZM2G130356        | esterase                                                                   | 6.2                |
| chr5.S_14926370              | 5          | 14926370              | A/ <u>G</u>         | 27/341    | 4.94×10 <sup>-8</sup>  | GRMZM2G035632        | hypothetical protein                                                       | 9.5                |
| chr5.S_15703594              | 5          | 15703594              | <u>C</u> /T         | 334/34    | 5.64×10 <sup>-11</sup> | AC197122.3_FGT003    | general negative regulator of transcription protein                        | 11.2               |
| chr5.S_19676906 <sup>e</sup> | 5          | 19676906              | <u>A</u> /G         | 245/123   | 1.46×10 <sup>-21</sup> | GRMZM2G124863        | glutamate formiminotransferase <sup>f</sup>                                | 27.6               |
| chr9.S_97200362              | 9          | 97200362              | T/ <u>G</u>         | 19/349    | 6.30×10 <sup>-8</sup>  | GRMZM2G104347        | hypothetical protein                                                       | 6.7                |

34 <sup>a</sup>Position in base pairs for the lead SNP based on the B73 reference genome sequence version 5b.60  
 35 (<http://www.maizesequence.org>).

36 <sup>b</sup>Major allele and minor allele. Underlined SNP is the favorable allele.

37 <sup>c</sup>Plausible biological candidate gene in the locus or annotated gene closest to the lead SNP.

38 <sup>d</sup>Each candidate gene is annotated according to Interproscan (<http://www.ebi.ac.uk/interpro>).

39 <sup>e</sup>Chr5.s\_19676907 has a genotype and *P*-value identical to those of Chr5.s\_19676906, only is  
 40 Chr5.s\_19676906 shown.

41 <sup>f</sup>Protein characterised in this study for catalysis of 5-M-THF to MeFox.

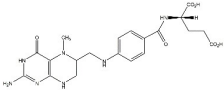
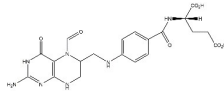
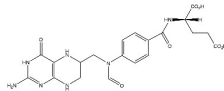
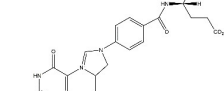
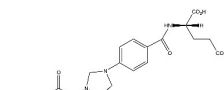
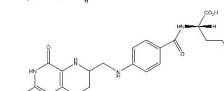
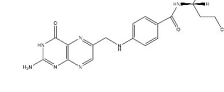
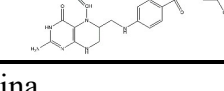
42

**Extended Data Table 3. Data collection and refinement statistics. Related to Fig. 2.**

|                                                     | SbCTM (7DYH)                                  |
|-----------------------------------------------------|-----------------------------------------------|
| Number of crystals                                  | 1                                             |
| Space group                                         | P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> |
| Cell dimensions                                     |                                               |
| <i>a</i> , <i>b</i> , <i>c</i> (Å)                  | 46.58, 77.89, 83.33                           |
| <i>a</i> , <i>b</i> , <i>g</i> (°)                  | 90, 90, 90                                    |
| Wavelength (Å)                                      | 0.9793                                        |
| Resolution (Å)                                      | 45-1.80 (1.84-1.80)                           |
| <i>R</i> <sub>merge</sub> (%)                       | 5.7 (71.7)                                    |
| <i>I</i> / <i>σ</i> <i>I</i>                        | 32.5 (3.8)                                    |
| Completeness (%)                                    | 100.0 (100.0)                                 |
| Number of measured reflections                      | 372,440 (22,519)                              |
| Number of unique reflections                        | 28,849 (1,672)                                |
| Redundancy                                          | 12.9 (13.5)                                   |
| Wilson B factor (Å <sup>2</sup> )                   | 21.5                                          |
| <i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> | 17.61 / 19.45                                 |
| No. atoms                                           |                                               |
| Protein                                             | 2,312                                         |
| Water                                               | 281                                           |
| B-factors                                           |                                               |
| Protein                                             | 29.3                                          |
| Water                                               | 39.2                                          |
| R.m.s. deviations                                   |                                               |
| Bond lengths (Å)                                    | 0.004                                         |
| Bond angle (°)                                      | 0.703                                         |
| Ramachandran plot statistics (%)                    |                                               |
| Most favorable                                      | 98.66                                         |
| Disallowed                                          | 0                                             |

44 The highest-resolution shell is shown in parentheses.  $R_{merge} = \frac{\sum h \sum i |I_{h,i} - \bar{I}_h|}{\sum h \sum i I_{h,i}}$ , where  $\bar{I}_h$  is the  
45 mean intensity of *i* observations of symmetry related reflections of *h*.  $R = \frac{\sum |F_{obs} - F_{calc}|}{\sum F_{obs}}$ , where  
46 *F*<sub>calc</sub> is the calculated protein structure factor from the atomic model (*R*<sub>free</sub> was calculated with 5% of  
47 the reflections selected).

48 **Extended Data Table 4. Virtual screen against ZmCTM-B73 protein. Related to Fig. 2.**

| Compound name                               | Molecular structure                                                                 | Affinity (kal mol <sup>-1</sup> ) <sup>a</sup> |
|---------------------------------------------|-------------------------------------------------------------------------------------|------------------------------------------------|
| 5-Methyltetrahydrofolate (5-M-THF)          |    | -9.0                                           |
| 5-Formyltetrahydrofolate (5-F-THF)          |    | -8.2                                           |
| 10-Formyltetrahydrofolate                   |    | -8.2                                           |
| 5,10-Methenyltetrahydrofolate (5,10-CH=THF) |    | -8.1                                           |
| 5,10-Methylenetetrahydrofolate (5,10-M-THF) |    | -8.0                                           |
| Tetrahydrofolate (THF)                      |   | -8.0                                           |
| Folic acid                                  |  | -7.7                                           |
| 5-formiminotetrahydrofolate                 |  | -7.4                                           |

49 <sup>a</sup>Affinity is calculated using Autodock Vina.

50 **Extended Data 5. Docking data of ZmCTM-Qi319 and ZmCTM-B73 with 5-M-THF. Related to**  
51 **Fig. 3.**

| Protein name | Docking method | Score                        |
|--------------|----------------|------------------------------|
| ZmCTM-Qi319  | Autodock Vina  | -9.1 kcal mol <sup>-1</sup>  |
| ZmCTM-B73    |                | -9.0 kcal mol <sup>-1</sup>  |
| ZmCTM-Qi319  | Autodock 4     | -13.9                        |
| ZmCTM-B73    |                | -11.8                        |
| ZmCTM-Qi319  | LigandFit      | 80.9                         |
| ZmCTM-B73    |                | 66.9                         |
| ZmCTM-Qi319  | Cdocker        | -63.8 kcal mol <sup>-1</sup> |
| ZmCTM-B73    |                | -62.2 kcal mol <sup>-1</sup> |

52

**Extended Data Table 6 Primers used in this study. Related to Methods**

| Primer names            | Sense primers (5'-3')                                                      | Antisense primers (5'-3')             | Usage                                       |
|-------------------------|----------------------------------------------------------------------------|---------------------------------------|---------------------------------------------|
| ZM-FAB-4863-1/4         | CCGTGTGCTTCTGGTGGATAAT                                                     | CGTCCTTGTTATGTCAGTGTAGT               | Resequencing primers                        |
| ZM-FAB-4863-7/10        | GCTACTACCGCCCAAACCTACAG                                                    | TTCTGGACACGACGGTAACTC                 | Resequencing primers                        |
| ZM-FAB4863-CDS-F/R      | AAAAAGCAGGCTATGGAGCCTCATCACGCAA<br>AC                                      | AGAAAGCTGGGTTTCAGTCGTCAGCGCAAGC       | Cloning ZmCTM                               |
| 540-F/R                 | GACTAGTCCAGTATTCTTGTATGGAGCAGC                                             | TCGGCACTGACATGATCTGGAT                | Cloning OsCTM                               |
| 4863CDS-BglII-F/        |                                                                            |                                       | RNAi fragment to                            |
| 4863CDS-BstEII-R        | AGATCTATGGAGCCTCATCACGCAAAC                                                | GGTTACCTCAGTCGTCAGCGCAAGC             | vector pCambia1301                          |
| ZM-FT-RT-F/ZM-FT-RT-R   | GCGTTCTCAAGCATAGACCTG                                                      | ACCGTCGTAGCGCCTCTCTT                  | Cloning CTM CDS                             |
| BAR-F/BAR-R             | GTCTGCACCATCGTCAACCACTACAT                                                 | TGAAGTCCAGCTGCCAGAAACCCA              | to overexpression                           |
| ZN-actin-F/ ZN-actin-R  | GGCCCAACTGCCGAAGCCAT                                                       | GAGAGGGGCCTCGGTGAGCA                  | vector pCambia3301                          |
| MU6 1-1 F/R             | TGCACTGCACAAGCTGTGGTGCCAAAGAAT<br>TTCCAGG                                  | AATTCGGTGCTTGCGGCTC                   | qRT-PCR of <i>ZmCTM</i>                     |
| MUsgR-1F/R              | GAGCCGCAAGCACCGAATTGAGCTTGACAGG<br>AGATGAAGCTGTTTTAGAGCTAGAAATAGCAA<br>GTT | GGCCAGTGCCAAGCTTAAAAAAGCACCGA<br>CTCG | Identification of overexpressors            |
| FTCD_M1_NdeI            | AAGAAGGAGATATACATATGATGAGCCAGCT<br>GGTTGAATG                               |                                       | qRT-PCR of <i>ACT1N</i>                     |
| FTCD_A541_XhoI          | TGGTGATGGTGATGCTCGAGTGCCTGACGT<br>GCTTCCAGACG                              |                                       | Cloning maize promoter to CRISP/cas9 vector |
| FT_V326_XhoI            | TGGTGATGGTGATGCTCGAGAACCAGGTATT<br>CAATAATACG                              |                                       | Cloning ZmCTM                               |
| Qi319-M1-NdeI           | TCGGGCGCGGATCCCATATGATGGAGCCTC<br>ATCACGCAAAC                              |                                       | sgRNA target to Crispr/cas9 vector          |
| Qi319-D330-XhoI(nostop) | TGGTGATGGTGATGCTCGAGGTCGTCAGCG<br>CAAGCAATC                                |                                       | Cloning                                     |
| SB-M1-NdeI              | AAGAAGGAGATATACATATGATGGAACCCCA<br>CCACGCTAAC                              |                                       | Cloning                                     |
| SB-D317-XhoI(nostop)    | TGGTGATGGTGATGCTCGAGGTCGTCGGTG<br>CAGGCGATCTTC                             |                                       | Cloning                                     |
| SB-A6-NdeI              | AAGAAGGAGATATACATATGGCTAACAAGCC<br>CAAGCCCAAG                              |                                       | Cloning                                     |
| SB-S-T315-XhoI(nostop)  | TGGTGATGGTGATGCTCGAGGGTGGAGGCG<br>ATCTTCAGGTAC                             |                                       | Cloning                                     |
| Ta_M1_NdeI              | AAGAAGGAGATATACATATGATGGAGGGCAA<br>GCGCACCGACCAG                           |                                       | Cloning                                     |

|                      |                                                |                                       |          |
|----------------------|------------------------------------------------|---------------------------------------|----------|
| Ta-D316-XhoI(nostop) | TGGTGATGGTGATGCTCGAGGTCAGCGGCG<br>GACACCATCTTG |                                       | Cloning  |
| St-M1-NdeI           | ATGGACCCCAACAGCGAGTG                           |                                       | Cloning  |
| St-S316-XhoI(nostop) | GGAGCTGCTGGAGATCAGCTTC                         |                                       | Cloning  |
| PA-M1-NdeI           | ATGGATCATAGTCCGGCATG                           |                                       | Cloning  |
| PA-N318-XhoI(nostop) | ATTACGATCTTCGCTGGTCAGC                         |                                       | Cloning  |
| Qi319-G232N          | CTTGGATCGAGAACTACAATGTC                        | GACATTGTAGTTCTCGATCCAAG               | Mutation |
| B73-N228G            | TGGATCGAGGGCTACAATGTC                          | GACATTGTAGCCCTCGATCCA                 | Mutation |
| SB_D57A              | CAGCTGGGCGCCCACCACTAC                          | GTAGTGGTGGGCGCCCAGCTG                 | Mutation |
| SB_R64A              | CAACCGTTTCGCCTACACCTG                          | CAGGGTGTAGCGAAACGGTTG                 | Mutation |
| SB-H109A             | TCCGGTACTGCCCTCGTATC                           | GATACGAGGGGCAGTACCGGA                 | Mutation |
| SB_G219N             | CATGGATCGAAAATTACAACATC                        | GATGTTGTAATTTTCGATCCATG               | Mutation |
| SB_H256A             | GCTCTGGCCCTGTTTCGCCGGTGACGACTGC<br>ACC         | GGTGCAGTCGTCACCGGCGAACAGGGCCAG<br>AGC | Mutation |
| SB_Y296A             | GAACAGGGTGCCTACACTG                            | CAGTGTAGGCACCCTGTTC                   | Mutation |
| SB_T298A             | GTTACTACGCTGACATCAC                            | GTGATGTCAGCGTAGTAAC                   | Mutation |