An Interactome Map of Maize (*Zea mays* L.)

Linqian Han

National Key Laboratory of Crop Genetic Improvement

Wanshun Zhong

National Key Laboratory of Crop Genetic Improvement

Jia Qian

National Key Laboratory of Crop Genetic Improvement

Minliang Jin

National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University

Peng Tian

National Key Laboratory of Crop Genetic Improvement

Wanchao Zhu

National Key Laboratory of Crop Genetic Improvement

Hongwei Zhang

Chinese Academy of Agricultural Sciences

Yonghao Sun

National Key Laboratory of Crop Genetic Improvement

Jia-Wu Feng

National Key Laboratory of Crop Genetic Improvement

Guo Chen

National Key Laboratory of Crop Genetic Improvement

Ruonan Li

National Key Laboratory of Crop Genetic Improvement

Zimo Xiong

National Key Laboratory of Crop Genetic Improvement

Juan Li

National Key Laboratory of Crop Genetic Improvement

Zi Luo

National Key Laboratory of Crop Genetic Improvement

Sijia Chen

National Key Laboratory of Crop Genetic Improvement

Qixiao Jin

National Key Laboratory of Crop Genetic Improvement

Jiaxin Li

National Key Laboratory of Crop Genetic Improvement

Zhao Li

National Key Laboratory of Crop Genetic Improvement

Yan Liang

National Key Laboratory of Crop Genetic Improvement

Xiaomeng Jin

National Key Laboratory of Crop Genetic Improvement

Xiangguo Liu

Jilin Academy of Agricultural Sciences

Xinnan Ye

Jilin Academy of Agricultural Sciences

Yuejia Yin

Jilin Academy of Agricultural Sciences

Yong Peng

Huazhong Agricultural University <https://orcid.org/0000-0002-5379-4348>

Xiaoduo Lu

Qilu Normal University

Lingling Chen

Huazhong Agricultural University <https://orcid.org/0000-0002-3005-526X>

Qing Li

National Key Laboratory of Crop Genetic Improvement

Jianbing Yan (✉ yjianbing@mail.hzau.edu.cn)

Huazhong Agricultural University <https://orcid.org/0000-0001-8650-7811>

Fang Yang (✉ fyang@mail.hzau.edu.cn)

National Key Laboratory of Crop Genetic Improvement

Lin Li (✉ hzaulilin@mail.hzau.edu.cn)

National Key Laboratory of Crop Genetic Improvement

Biological Sciences - Article

Keywords: Maize, Interactome, Functional genomics, Protein–protein interaction, Machine learning

Posted Date: December 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-88694/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

An Interactome Map of Maize (*Zea mays* L.)

Linqian Han^{1, #}, Wanshun Zhong^{1, #}, Jia Qian^{1, #}, Minliang Jin^{1, #}, Peng Tian^{1, #}, Wanchao Zhu^{1, #}, Hongwei Zhang^{2, #}, Yonghao Sun^{1, #}, Jia-Wu Feng¹, Guo Chen^{1,3}, Ruonan Li¹, Zimo Xiong¹, Juan Li¹, Zi Luo¹, Sijia Chen¹, Qixiao Jin¹, Jiaxin Li¹, Zhao Li¹, Yan Liang¹, Xiaomeng Jin¹, Yong Peng¹, Xiaoduo Lu⁴, Xiangguo Liu⁵, Xinnan Ye⁵, Yuejia Yin⁵, Ling-Ling Chen¹, Qing Li¹, Jianbing Yan^{1,*}, Fang Yang^{1,*}, and Lin Li^{1,*}

¹National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

²Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

³Institute of Nuclear and Biological Technology, Xinjiang Academy of Agricultural Sciences, Urumqi 830091, China

⁴Institute of Molecular Breeding for Maize, Qilu Normal University, Jinan 250200, China

⁵Institute of Agricultural Biotechnology, Jilin Academy of Agricultural Sciences, Changchun, 130033, China

[#]These authors contributed equally to this work

*Corresponding authors: Lin Li - hzaulilin@mail.hzau.edu.cn; Fang Yang - fyang@mail.hzau.edu.cn; Jianbing Yan - yjianbing@mail.hzau.edu.cn

Author contributions: L.L., F.Y., and J.Y. designed and supervised this study. L.H., W.Z., J.Q., M.J., P.T., W.Z., H.Z., Y.S., J.F., G.C., R.L., Z.X., J.L., Z.L., S.C., Q.J., J.X.L., Y.P., X.L., L.C., Q.L., F.Y., and L.L. performed the data analysis. L.L., F.Y., J.Y., W.Z., and L.H. wrote the manuscript.

Summary

Interactomes are powerful tools for encoding and decoding complex life systems. Here, we generated a maize interactome map that integrates genomic interactions, transcriptomic co-expression networks, translatomic co-expression networks, and protein–protein interactions throughout the maize lifecycle. This map, containing over 9 million interactions in more than 5,000 functional modules, reveals extensive functional divergence for duplicate genes and a progressive increase in regulatory divergence between the two maize subgenomes during the flow of genetic information. This network enables dissecting and validating gene functions, re-constructing regulatory pathways, and deciphering molecular mechanisms underlying complex traits combining big data mining technique-machine learning. By applying this map to flowering-time, we identified 1,843 high-confidence genes enriched in eight molecular pathways that are related to flowering time. The function of 30 (out of 58 tested) genes, including 27 novel genes, was verified by loss-of-function mutagenesis. Furthermore, a new pathway involving histone modification was identified and confirmed to regulate flowering time. The interactome map illustrates how coherent sets of molecular interactions connect different types of functional elements and pathway modules to map a genome-wide functional wiring landscape of maize, which will be applicable in a wide range of species.

Key words: Maize, Interactome, Functional genomics, Protein–protein interaction, Machine learning

Organisms and cells are inherently complex¹. Although thousands of high-quality genome sequences (https://wikimili.com/en/List_of_sequenced_eukaryotic_genomes) and numerous genotype–phenotype associations have been uncovered using advanced biological and genetic techniques², our understanding of life systems remains limited³. Interactomes are excellent tools for describing such ultracomplex systems⁴. Many interactomes have been constructed for humans and model animals, but a few have been generated for plants (<http://www.pathguide.org/>). A genome-wide protein–protein binary interaction map was constructed for ~6,200 highly reliable interactions between ~2,700 proteins in *Arabidopsis thaliana*, revealing dramatic functional divergence of *Arabidopsis* genes⁵. In addition, a quantitative atlas of the transcriptomes, proteomes, and phosphoproteomes of 30 *Arabidopsis* tissues was recently constructed⁶. Despite this progress, interactions that cover the genetic flow from genome to transcriptome and proteome are lacking. Comprehensive interactome maps involving the genome, transcriptome, translome, and proteome simultaneously are needed to decode complex life systems.

An interactome map of maize

Maize, one of the most important crops worldwide and an ideal biological and genetic model, has been extensively explored for genotypic and phenotypic variations². Recently, large datasets of mRNA, protein, and phosphoprotein have been successfully integrated to construct a developmental atlas for maize⁷. An interactome covering all layers of genetic information flow would further promote the precise genetic improvement of this crop. Here, we leveraged advances in high-throughput sequencing techniques to measure

genomic interactions by ChIA-PET⁹; transcriptomic interactions for all detectable elements by strand-specific RNA-seq, circRNA-seq, and small RNA-seq; translatomic interactions by Ribo-seq; and protein–protein binary interactions (PPIs) by recombination-based library vs. library yeast-2-hybrid (RLL-Y2H)⁸ analysis throughout the lifecycle of the maize inbred line B73 (Fig. 1a).

Comprehensive transcriptomic and translatomic deep sequencing generated 1.78 trillion sequencing bases (Supplementary Fig. 1; Supplementary Table 1). Transcriptomic profiling uncovered the expression of 37,701 protein coding genes (9.5% of the genome), 9,858 long non-coding RNAs (lncRNAs) (1.4%), 61,137 circRNAs (2.3%), 343,507 small RNA (sRNA) clusters (67–87% using different gap thresholds), and 145 fusion expressed elements (0.001%) across 26 different stages and tissues (Fig. 1b; Supplementary Fig. 1). For translatomic profiling, we performed Ribo-seq¹⁰ on 20 out of the 26 tissues and identified 34,949 genes that were embedded in ribosomes. These expressed elements at both the transcriptome and translatomic levels span 90.5% of the reference genome, pointing to pervasive transcription in maize.

We conducted a series of multi-omic experiments to dissect interactions of any pair of functional elements and functional modules, where functional elements with similar roles were clustered coherently (see Methods). At the genome level, 13,142 loci formed 14,788 3-D physical interactions within 689 modules⁹ (Fig. 1c). At the transcriptome level, 178,025 expressed elements coherently functioned in 2,944,080 co-expression pairs for 5,998 modules (Fig. 1d; Supplementary Table 2). At the translatomic level, 23,646 genes formed 68,972 translational co-expression pairs in 1,485 modules (Fig. 1e; Supplementary Table 2).

To construct a genome-wide protein–protein interactome (PPI), we used eight bait and eight prey cDNA libraries from eight distinct tissues across the whole developmental stages and to screen for interactions via 130 library vs. library matings using the RLL-Y2H system. We performed Pacific Biosciences (PacBio) Sequel sequencing (<https://www.pacb.com/>) of plasmids isolated from the surviving positive yeast clones from each mating (Supplementary Fig. 3; Supplementary Table 3). We obtained nearly 9 million consensus PacBio full-length sequences and identified 87,775 high-confidence PPIs between 9,386 Binding Domain (BD) bait genes and 16,477 Activation Domain (AD) prey genes, filtered from over 360,000 detectable PPIs (Fig. 1f, Supplementary Table 4).

The multi-omic interactomes exhibited markedly different topological hierarchy (Fig. 1c–1f). Based on different confidence cut-offs (See Methods), the interactomes were grouped into low-, middle-, and high-confidence versions that showed similar trends (Supplementary Fig. 3). The high-confidence interactome was used for subsequent analyses. Although a subset of hub nodes (functional elements) from different -omics data overlap (Fig. 1g; Supplementary Fig. 4), the overall clustering coefficients was 0.254 in the 3D genome, 0.321 in the transcriptome, 0.12 in the translome, and 0.006 in the proteome (Supplementary Table 2), exhibiting a distinct topological structure for each regulatory layer (Fig. 1c–1f).

We integrated the interaction network from different layers. In all, we generated over 2 trillion bases and detected over 182,995 functional elements and ultimately constructed an interactome with over 3 million edges (Fig. 1h; Supplementary Fig. 5). Diverse types of regulatory elements interacted in this network, for example, *teosinte glume architecture1 (tga1)*¹¹ showed interactions with lncRNAs, circRNAs, and miRNAs

114 which were not found previously (Fig. 1i). We've developed a user-friendly website to store
115 all the interactome information (<http://cbi.hzau.edu.cn/interactome/index.php>), which can
116 be searched easily by gene name or using node information.

117 A comparison of our interactome with previously generated interactomes revealed
118 significant conservation of interactions (Supplementary Fig. 6a–6j), and over 50% of PPIs
119 could be validated using other biological techniques (Supplementary Table 5;
120 Supplementary Fig. 6k–6n; Supplementary Fig. 7). Although it is still not saturated
121 (Supplementary Fig. 6o), this is the largest protein interaction database to date (~360,000),
122 and the interactome map represents the largest, nearly complete, high-quality interactome
123 in plants^{5,7}.

Figure 1. The interactome uncovers the genome-wide regulatory landscape in maize. a, Data collection. Complete omics data covering all layers of genetic information flow were collected from maize inbred B73 throughout its entire lifecycle. The numbers in the cells of the matrix indicate the number of tissues collected for each of the four types of high-throughput data. b, Distribution and proportion of different types of expressed RNAs across the maize genome. c–f, Genome-wide landscape of interactions in the genome, transcriptome, translome, and proteome, respectively. Numbers in each panel (from top to bottom) represent the number of functional elements, interactions (edges), and modules. g, Total number of hub genes and number of overlapping hub genes (nodes above the 90th percentile for the number of edges) among the transcriptome, translome, and proteome. h, Genome-wide landscape of integrative interactome based on multi-omics. i, An example of an interaction module for the well-known gene *tg1* involving lncRNAs (orange long line), circRNAs (orange circle), small RNAs (orange short line), and protein-coding mRNAs (orange oval). Solid and dotted lines represent direct and indirect interactions, respectively.

Progressive functional divergence between the maize subgenomes from DNA to proteins

The interactome map reveals dramatic regulatory divergence for duplicated genes as expected (Supplementary text; Supplementary Fig. 8). Notably, maize is an allotetraploid with distinct ancient subgenomes, Maize1 and Maize2, that exhibit asymmetric divergence in both gene content and expression levels¹². Although Maize1, the dominant subgenome, contains more orthologs and highly expressed genes than Maize2, there is no evidence for regulatory bias between the two subgenomes^{12,13}. We detected no or only subtle differences between the subgenomes at the transcriptome level. However, significant differences were detected at both the translome and proteome levels (Fig. 2a; Supplementary Fig. 8g-8h). We aligned the maize genome with its ancestral sorghum genome and quantified the overall degree of differentiation of genes in a sliding window (100 genes). The regulatory divergence of the subgenomes increased from transcriptome to translome to proteome and exhibited dramatic variation across different windows (Fig. 2b; Supplementary Fig. 9-10).

To further explore the evolutionary relevance of subgenome divergence, we focused on hub genes at different omic layers. Intriguingly, Maize1 hub genes of both the transcriptome and translome levels were significantly over-represented ($P < 0.01$) at genomic regions targeted for domestication and improvement¹⁴; however, Maize1 hub genes of the proteome level were not over-represented (Fig. 2c). Accordingly, Maize2 hub genes of the transcriptome level were significantly enriched ($P < 0.001$) in loci targeted for domestication and improvement, and Maize2 hub genes of both the translome and

proteome levels were underrepresented in domestication loci (Supplementary text; Supplementary Fig.11). These results indicate that the variation of Maize1 hub genes at both the transcriptome and translome levels and of Maize2 hub genes at the transcriptome-level might be important for speciation and crop improvement. The non-enrichment of hub genes in neither subgenomes suggests that selection based on protein-coding sequence variation of maize hub genes is risky probably due to the detrimental nature of protein sequences¹⁵.

Figure 2. Asymmetric regulatory divergence indicates differences in evolutionary selection between the two maize subgenomes. a, Degree of variation of genes in subgenomes Maize1 and Maize2 across different regulatory levels. Maize1 and Maize2 represent homologous genes from subgenome Maize1 and Maize2, respectively. Maize1 without Maize2 indicates Maize1 genes whose corresponding Maize2 genes were lost. Maize2 without Maize1 indicates Maize2 genes whose corresponding Maize1 genes were lost. b, Biased regulatory fractionation at the transcriptome (top), translome (middle), and proteome (bottom) levels is observed for the reconstructed or “sorghumized” pair maize ancestral chromosome–chromosome 1. Dominant bin is defined as a genome region in which more genes in one subgenome have a significantly higher degree of bias than the corresponding homologous genes in the other subgenome. Lines are LOESS regression lines. The pie chart to the right represents the number of dominant bins from each subgenome. c, Enrichment of Maize1 and Maize2 hub genes at the transcriptome (top), translome (middle), and proteome (bottom) levels at loci targeted for domestication and improvement. *, **, and *** indicate P-values < 0.05, 0.01, and 0.001 by Chi-square test, respectively. Grey, yellow, and blue bars represent genome-wide, Maize1, and Maize2, respectively.

The interactome recapitulates the networks of important genes

The interactome map can reconstruct the networks of well-known genes and uncover new crosstalk genes with similar function. We explored an interaction subnetwork involving three key maize tillering genes, *Teosinte Branched 1 (TB1)*¹⁶, *Grassy Tillers1 (GT1)*¹⁷, and *Tassels Replace Upper Ears1 (TRUI)*¹⁸. Loss-of-function mutations in any of these three genes lead to more tillers, a characteristic of the maize ancestor teosinte. *TB1*, *GT1*, and *TRUI* demonstrated an interaction with *ZmALOG1* (Zm00001d003057) and *ZmALOG2* (Zm00001d032696), two functionally unknown genes belong to the ALOG (*Arabidopsis* LSH1 and *Oryza* G1) transcription factor family (Fig. 3a). Intriguingly, the loss-of-function mutations in *ZmALOG1* and 2 caused enhanced tillering, similar as that in *tb1*, *gt1*, and *trul* mutants (Fig. 3b; Supplementary Fig. 12a). The interactions between *ZmALOG1* and 2 and *TRUI* were confirmed by Y2H assays (Fig. 3c). Moreover, *ZmALOG1* and 2 also interact with *TB1* (Fig. 3c). These results demonstrate that the interactions between known and unknown genes in subnetworks have biological meaning. Therefore, our maize interactome can be used to reliably predict the functions of genes of interest, as well as their putative interaction pathways, shedding light on the regulatory mechanisms of known and unknown genes.

The interactome map contains functional modules with coherent sets of molecular interactions for genes of similar genetic functions. *CUP-SHAPED COTYLEDON (CUC)* genes establish boundaries between organs in many plant species¹⁹. However, the roles of *CUC* genes in maize have not yet been well recognized. To test the suitability of the maize interactome for discovering the functions of genes of interest, we searched the interactome

modules and observed that the maize *CUC* genes (*ZmCUC3* and its two paralogs *ZmNAM1*, and *ZmNAM2*) interacted with *TASSEL SHEATH1* (*TSH1*), *BARREN STALK1* (*BAI*), and *PIN-FORMED1*(*ZmPIN1*) (Fig. 3d); these genes (or their homologues) function in lateral organ formation and development in maize or other species^{20–26}. Thus, we reasoned that maize *CUC* homologues might play similar roles in lateral organ development. Indeed, loss-of-function *zmnaml* and *zmnam2* single and double mutants generated by CRISPR technology exhibited suppressed lateral bud formation (Fig. 3e; Supplementary Fig. 12b), confirming that these genes play conserved roles in lateral organ formation by establishing boundaries, which are required for proper lateral meristem initiation. The interactions between *ZmCUCs* and other genes, including *BAI*, *TSH1* and unknown genes *bHLH168* and *MADS838* in the subnetwork are candidates for further research.

The interactome map can illustrate the molecular interactions for genes, which were cloned separately for complex traits. Maize kernel size and weight are the most important components of maize yield and have been intensively investigated. A total of 63 well-known kernel mutant genes have been functionally cloned (Supplementary Table 6). We identified 55 out of 63 genes in the interactome and divided them into two groups (40 for constructing kernel-related subnetworks and 15 for validating the robustness of the kernel-related regulatory network). The subnetworks of 40 randomly selected kernel genes (1000 simulations) successfully predicted up to 40% of the validated kernel genes, a value significantly higher ($P = 1\text{E-}6$) than that of randomly selected functionally unrelated genes (Supplementary Fig. 13a). Notably, all 55 kernel genes could be assembled into a linked subnetwork in the maize interactome (Fig. 3f), which was further clustered into eight modules. The genes in six modules were significantly enriched in the functional groups of

embryo development, starch metabolic process, sugar mediated signalling pathway, and metabolic process. Among these, the mutation of an unreported gene *Zm00001d003047*, encoding a Pentatricopeptide Repeat-containing (PPR) protein, led to shrunken kernels (Fig. 3f–3h; Supplementary Fig. 13b). Using the maize interactome, we also successfully reconstructed regulatory networks involved in starch synthesis pathway^{27–30} and other established pathways (Supplementary Fig. 14; Supplementary Table 7–8). Thus, the maize interactome is a powerful tool for exploring and assembling gene regulatory networks underlying complex traits, which could help breeders to optimize trait improvement.

Figure 3. The interactome can predict gene function and re-construct regulatory networks underlying complex traits. a, A network showing that the well-known tillering-related genes *tb1*, *tru1*, and *gt1* are connected to two novel genes (*ZmALOG1* and *ZmALOG2*). b, Knockout mutants of *zmalog1* and *zmalog2* show altered tillering phenotypes. c, Protein interactions between *ZmALOG1* and 2 and *TRU1*, *TB1* were confirmed by Yeast-2-Hybrid assays. d, A network showing that *CUC3*, a gene known to function in lateral tissue development, is connected to two novel genes, *NAM1* and *NAM2*. e, CRISPR/Cas9 knockout mutants of *NAM1* and *NAM2* show suppressed development of lateral meristem. f, Construction of regulatory networks to uncover the molecular mechanism conferring kernel development. The nodes contain genes with previously reported functions in kernel development. Different colours represent different modules associated with kernel variation. g, A gene of unknown function (*Zm00001d003047*) was identified in the newly constructed regulatory network based on the interactome of well-known kernel genes. h, The shrunk kernel phenotype

of the knockout mutant (*Zm00001d003047*) co-segregated with the genotypes, whose phenotype and genotypes fit a 3:1 segregation ratio.

Systematically deciphering complex traits: flowering time as a proof of concept

Flowering time (FT) is an important agronomic trait involved in the adaptation of maize to a wide range of climates worldwide^{31–33}. Although many key FT genes in maize have been cloned (Supplementary Table 9), the genetic control and molecular mechanisms underlying this trait remain elusive. To dissect the potential regulatory networks, we constructed a neutral prediction model based on the interactome map and machine learning method using known FT genes (reported FT genes in maize and maize homologs to FT genes cloned in other species) as the training dataset. The model showed high prediction accuracy, and different layers of the interactome had variable prediction power, with AUC values ranging from 0.68 to 0.89 (Fig. 4a). Notably, the integrative interactome has the highest AUC value (up to 0.9). Totally, 3,553 genes were predicted to be associated with FT at different confidence levels (Fig 4b; Supplementary Table 10), suggesting that an ultracomplex molecular mechanism underlies FT in maize. We predicted the shortest distance (SD) to known validated FT genes between all predicted FT genes and randomly selected genes, and found that the predicted FT genes had significantly lower SD values with validated FT genes compared to randomly selected genes (Fig. 4b). Using more restrictive filtering via empirical cut-off of SD distribution, we obtained 1,843 high-confidence FT genes (Fig. 4b; Supplementary Table 10). Compared with random genes, the predicted FT genes are significantly enriched in both the *Arabidopsis* homology dataset

(<http://www.phytosystems.ulg.ac.be/florid/>) and SNP-GWAS loci³⁴ that are associated with flowering time (Supplementary Fig. 15).

To assess the precision of our predictions, we chose 48 predicted genes and 10 training FT genes and successfully knocked out their functions by generating either truncated protein or shifted reading frames for functional validation by CRISPR knockout or EMS mutagenesis. In total, 30 genes (27 newly validated in maize and 3 reported FT maize genes) were confirmed to be associated with FT variation (Fig. 4c-4h; Supplementary Table 11; Supplementary Fig. 16–45). Mutations in 9 of the 10 known FT genes led to significant deviations in FT compared to the wild type in at least one growing location, while knockout mutations in 43.75% (21/48) of the predicted FT genes led to significant FT variation (Fig. 4c; Supplementary Table 11; Supplementary Fig. 16-45). For example, loss-of-function mutation in the predicted gene *Zm00001d06236*, which encodes an MYB transcription factor, could result in postponed flowering time in three environments (Fig. 4d-4f; Supplementary Fig. 16). Field tests on the other loss-of-function mutants showed significant FT variations as expected (Fig. 4g; Supplementary Fig. 17-45).

Furthermore, based on a previous gene regulatory model³⁵, our interaction network, and the 30 validated FT genes, eight molecular pathways were found to be associated with FT variation in maize: the light transduction pathway, circadian clock pathway, photoperiod pathway, autonomous pathway, gibberellin (GA) pathway, a pathway for the maintenance of inflorescence meristem identity (Integrator), the floral transition pathway, and a newly uncovered pathway associated with histone modification (Fig. 4h; Supplementary Table 12). These results indicate that machine learning mining on our

interactome map is suitable for the large-scale identification of functional genes underlying complex traits.

The addition of the 30 validated genes to molecular pathways involving FT in maize further our understanding of FT variation (Fig. 4h). For example, Six genes were placed into the light transduction pathway: *Zm00001d024126* (Supplementary Fig. 17), *Zm00001d013402* (Supplementary Fig. 18), *ZmPHYB1*³⁶ (Supplementary Fig. 19), *Zm00001d033799* (Supplementary Fig. 20), *ZmPHYC1*³⁷ (Supplementary Fig. 21), *Zm00001d045944* (*ZmCRY2*, *CRY2* homologue³⁸) (Supplementary Fig. 22). These genes encode phytochrome A/B/C and photoreceptor cryptochrome proteins. Seven genes were added into the circadian clock-related pathway: *Zm00001d006236* (*ZmMYB31*) (Supplementary Fig. 16), *Zm00001d021291* (*ZmPRR95a*, circadian clock component *PRR9* homologue³⁹) (Supplementary Fig. 23), *Zm00001d006212* (*ZmPRR95b*, circadian clock component *PRR9* homologue³⁹) (Supplementary Fig. 24), *Zm00001d017241* (*ZmTOC1b*, *TOC1* homologue⁴⁰) (Supplementary Fig. 25), *Zm00001d049543* (*ZmCCA1*, *LHY* and *CCA1* homologue⁴¹) (Supplementary Fig. 26), *Zm00001d034313* (Supplementary Fig. 27), and *Zm00001d006918* (Supplementary Fig. 28).

Notably, a newly identified FT pathway was uncovered in our model that includes five newly validated genes (Fig. 4h): *Zm00001d010625* (Supplementary Fig. 29), *Zm00001d011748* (Supplementary Fig. 30), *Zm00001d017193* (Supplementary Fig. 31), *Zm00001d032894* (Supplementary Fig. 32), and *Zm00001d006918* (Supplementary Fig. 33). Homologues of these newly discovered genes in *Arabidopsis* regulate the vernalization response gene *FLC*⁴⁴ via epigenetic modification (H4K5ace⁴⁵; antisense RNA synthesis⁴⁶) and nuclear complex formation⁴⁷. Although vernalization does not occur in maize, maize

genes whose homologues in *Arabidopsis* function in vernalization remain functional in maize FT pathways. Genes in this pathway were enriched in the GO terms gene silencing and histone modification (Supplementary Table 13). The unexpected discovery of a core gene of this newly identified FT pathway in maize lays the foundation for studying the relationship between FT and vernalization in domestication in the future^{13,32}.

Since we systematically uncovered and validated a series of molecular pathways underlying FT in maize, we asked whether genes from these pathways contribute in consistent direction to FT variation. Loss-of-function mutants of FT genes grown in the same environment (Shandong, China in the summer of 2020) exhibited a wide range of significant FT variations (Fig. 4i; Supplementary Table 11). Interestingly, loss-of-function mutations of genes from the same pathway had inconsistent genetic effects on FT variation (Fig. 4i).

Flowering time is a key adaptive trait that has helped maize spread to a wide range of environments worldwide. Many FT genes were selected during maize domestication and improvement¹⁴. We asked whether specific FT pathways were preferentially subjected to selection during maize domestication and improvement. Enrichment analysis of FT pathway genes in genomic regions targeted for domestication and improvement¹⁴ demonstrated that the floral transition pathway was significantly more likely to be subject to selection during maize domestication than the other pathways (Supplementary Fig. 46). Moreover, the GA pathway was significantly more likely to be selected during maize improvement (Supplementary Fig. 46). Most importantly, genes with relatively high genetic effects in the circadian clock and light transduction pathways have not been subject

to selection during modern maize breeding, which represent potential selection targets for modern breeding.

Figure 4. The interactome facilitates systematic dissection of the molecular mechanisms underlying flowering time in maize. a, Distribution of the AUC (Area Under Curve) values of different layers of the interactome for machine learning to predict flowering time (FT)-related genes. b, Distribution of SD values of predicted FT genes and randomly selected genes. SD represents the shortest distance to well-known FT genes. c, The percentage of genes with validated

functions in loss-of-function mutants. d–f, Validation of the function of *Zm00001d006236*. d, An early stop codon mutation occurred in the last exon of *Zm00001d006236*. e, *Zm00001d006236* mutants show delayed flowering time compared to the wild type. f, Statistical comparison of FT variation between the wild type and *Zm00001d006236* mutants show that the mutant has significantly delayed FT compared to wild-type plants. **, and *** indicate P-values < 0.01, and 0.001 by Student's T-test, respectively. g, The 30 FT genes were shown to be significantly associated with FT variation in multiple years or locations. The label in y-axis shows year.location.season.trait. LF: Langfang, Hebei, China; SD: Zibo, Shandong, China; JL: Gongzhuling, Jilin, China; YN: Xishuangbanna, Yunnan, China; HN: Sanya, Hainan, China; C: spring; X: summer; D: winter; T: DTT (days to tasselling); A: DTA (days to anthesis), S: DTS (days to silking). h, All validated FT genes (including well-known genes and newly validated genes in this study) and their interaction genes could be classified into eight pathways. The 30 validated genes were highlighted in red. i, Genetic effects of 21 FT genes in the same growing location. Different colours of gene ID on the x-axis represent different pathways in Fig. 4h.

Beyond the first interactome of maize

The first interactome of maize uncovered more than 9 million interactions spanning all levels of genetic information flow across the entire maize lifecycle. It is the largest molecular functional network generated to date, however, it is still not saturated, especially at the proteome level. To examine binary protein–protein interactions, we only screened 10,049 BD bait genes against 24,586 AD prey genes, representing one-quarter of all possible PPIs. Furthermore, the current interactome only describes intra-omic interactions and not inter-omic interactions such as protein–DNA interactions, and protein–RNA interactions. These are interesting questions for future research.

Conclusions

Genetic information is derived from the genome and transferred to the transcriptome, translome, and finally the proteome. This process is regulated at multiple

366 levels simultaneously¹. Dissecting these regulatory networks at multi-omics levels is a
367 highly efficient way to explore the functional genomics and life systems of an organism,
368 which are highly dynamic and complex¹. We assembled a comprehensive interactome,
369 which contains genomic, transcriptomic, translatomic, and proteomic interactions. Our
370 data demonstrate that functional divergence exists at different layers and leads to gradually
371 increasing functional differentiation. Compared to the existing regulatory networks in
372 plants^{5,7}, the current version of the maize interactome involves all layers of genetic
373 information flow, representing the largest interactome in plants generated to date. We
374 showed that the interactome facilitates the identification of the functions of key genes and
375 the systematic elucidation of the molecular mechanisms underlying phenotypic variations.
376 Thus, the maize interactome introduced here paves the way for systematically dissecting
377 the molecular mechanisms underlying agronomic traits, thereby revolutionizing genetic
378 research. It allows the simultaneous identification of hundreds of thousands genes involved
379 in a trait and the construction of their interaction network. Coupled with genome editing
380 technologies, the interactome can guide systematic engineering of the agronomic ally
381 important trait in the near future.

382 **Acknowledgements:** This research was supported by the National Natural Science
383 Foundation of China (91735305; 31771798; 31922068), the National Key Research and
384 Development Program of China (2016YFD0100800), Open funds of the National Key
385 Laboratory of Crop Genetic Improvement, the Fundamental Research Funds for the Central
386 Universities (ZK201908), and Huazhong Agricultural University Scientific &
387 Technological Self-innovation Foundation (2015R016).

388 **Competing interests**

389 The authors declare no competing interests.

References

1. Trewavas, A. A Brief History of Systems Biology: “Every object that biology studies is a system of systems.” Francois Jacob (1974). *Plant Cell* **18**, 2420–2430 (2006).
2. Wallace, J. G., Larsson, S. J. & Buckler, E. S. Entering the second century of maize quantitative genetics. *Heredity* **112**, 30–38 (2014).
3. Ma’ayan, A. Complex systems biology. *J. R. Soc. Interface* **14**, 20170391 (2017).
4. Dixon, S. J., Costanzo, M., Baryshnikova, A., Andrews, B. & Boone, C. Systematic Mapping of Genetic Interaction Networks. *Annu. Rev. Genet.* **43**, 601–625 (2009).
5. Consortium, A. I. M. Evidence for Network Evolution in an *Arabidopsis* Interactome Map. *Science* **333**, 601–607 (2011).
6. Mergner, J. *et al.* Mass-spectrometry-based draft of the *Arabidopsis* proteome. *Nature* **579**, 409–414 (2020).
7. Walley, J. W. *et al.* Integration of omic networks in a developmental atlas of maize. *Science* **353**, 814–818 (2016).
8. Yang, F. *et al.* Development and application of a recombination-based library versus library high-throughput yeast two-hybrid (RLL-Y2H) screening system. *Nucleic Acids Res.* **46**, e17–e17 (2018).
9. Peng, Y. *et al.* Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. *Nat. Commun.* **10**, 1–11 (2019).
10. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **324**, 218–223 (2009).
11. Wang, H., Studer, A. J., Zhao, Q., Meeley, R. & Doebley, J. F. Evidence That the Origin of Naked Kernels During Maize Domestication Was Caused by a Single Amino Acid Substitution in *tg1*. *Genetics* **200**, 965–974 (2015).
12. Schnable, J. C., Springer, N. M. & Freeling, M. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. USA.* **108**, 4069–4074 (2011).
13. Li, L. *et al.* Co-expression network analysis of duplicate genes in maize (*Zea mays* L.) reveals no subgenome bias. *BMC Genomics* **17**, 875 (2016).
14. Hufford, M. B. *et al.* Comparative population genomics of maize domestication and improvement. *Nature Genet.* **44**, 808–811 (2012).
15. Ecker, J. R. *et al.* ENCODE explained. *Nature* **489**, 52–54 (2012).

- 423 16. Doebley, J., Stec, A. & Hubbard, L. The evolution of apical dominance in maize. *Nature* **386**,
424 485–488 (1997).
- 425 17. Whipple, C. J. *et al.* *grassy tillers1* promotes apical dominance in maize and responds to shade
426 signals in the grasses. *Proc. Natl. Acad. Sci. USA*. **108**, 13375–13376 (2011).
- 427 18. Dong, Z. *et al.* Ideal crop plant architecture is mediated by *tassels replace upper ears1*, a
428 BTB/POZ ankyrin repeat gene directly targeted by TEOSINTE BRANCHED1. *Proc. Natl.*
429 *Acad. Sci. USA*. **114**, E8656–E8664 (2017).
- 430 19. Hake, S. Identification of cup-shaped cotyledon: New Ways to Think about Organ Initiation.
431 *Plant Cell* **31**, 1202–1203 (2019).
- 432 20. Gallavotti, A., Yang, Y., Schmidt, R. J. & Jackson, D. The Relationship between Auxin
433 Transport and Maize Branching. *Plant Physiol.* **147**, 1913–1923 (2008).
- 434 21. Yao, H. *et al.* The *barren stalk2* Gene Is Required for Axillary Meristem Development in Maize.
435 *Mol. Plant*. **12**, 374–389 (2019).
- 436 22. Skirpan, A., Wu, X. & McSteen, P. Genetic and physical interaction suggest that BARREN
437 STALK1 is a target of BARREN INFLORESCENCE2 in maize inflorescence development.
438 *Plant J.* **55**, 787–797 (2008).
- 439 23. Whipple, C. J. *et al.* A Conserved Mechanism of Bract Suppression in the Grass Family. *Plant*
440 *Cell* **22**, 565–578 (2010).
- 441 24. Gälweiler, L. *et al.* Regulation of Polar Auxin Transport by AtPIN1 in *Arabidopsis* Vascular
442 Tissue. *Science* **282**, 2226–2230 (1998).
- 443 25. Wang, Q., Kohlen, W., Rossmann, S., Vernoux, T. & Theres, K. Auxin Depletion from the Leaf
444 Axil Conditions Competence for Axillary Meristem Formation in *Arabidopsis* and Tomato.
445 *Plant Cell* **26**, 2068–2079 (2014).
- 446 26. Gallavotti, A. *et al.* The role of *barren stalk1* in the architecture of maize. *Nature* **432**, 630–635
447 (2004).
- 448 27. Zhang, Z., Zheng, X., Yang, J., Messing, J. & Wu, Y. Maize endosperm-specific transcription
449 factors O2 and PBF network the regulation of protein and starch synthesis. *Proc. Natl. Acad.*
450 *Sci. USA*. **113**, 10842–10847 (2016).
- 451 28. Wang, W. *et al.* The *Zea mays* mutants *opaque2* and *opaque16* disclose lysine change in waxy
452 maize as revealed by RNA-Seq. *Sci Rep* **9**, 1–12 (2019).
- 453 29. Feng, F. *et al.* OPAQUE11 Is a Central Hub of the Regulatory Network for Maize Endosperm
454 Development and Nutrient Metabolism. *Plant Cell* **30**, 375–396 (2018).
- 455 30. Mach, J. Clarifying the Opaque: Identification of Direct Targets of Maize Opaque2. *Plant Cell*
456 **27**, 484–484 (2015).

- 457 31. Kuleshov, N. N. World's diversity of phenotypes of maize. *Journal of the American Society of*
458 *Agronomy* (1933).
- 459 32. Buckler, E. S. *et al.* The Genetic Architecture of Maize Flowering Time. *Science* **325**, 714–718
460 (2009).
- 461 33. Heerwaarden, J. van, Hufford, M. B. & Ross-Ibarra, J. Historical genomics of North American
462 maize. *Proc. Natl. Acad. Sci. USA*. **109**, 12420–12425 (2012).
- 463 34. Liu, H.-J. *et al.* CUBIC: an atlas of genetic architecture promises directed maize improvement.
464 *Genome Biol.* **21**, 20 (2020).
- 465 35. Dong, Z. *et al.* A Gene Regulatory Network Model for Floral Transition of the Shoot Apex in
466 Maize and Its Dynamic Modeling. *PLoS One* **7**, e43450 (2012).
- 467 36. Sheehan, M. J., Kennedy, L. M., Costich, D. E. & Brutnell, T. P. Subfunctionalization of *PhyB1*
468 and *PhyB2* in the control of seedling and mature plant traits in maize. *Plant J.* **49**, 338–353
469 (2007).
- 470 37. Li, Q. *et al.* CRISPR/Cas9-mediated knockout and overexpression studies reveal a role of maize
471 phytochrome C in regulating flowering time and plant height. *Plant Biotechnol. J.* (2020).
- 472 38. Guo, H., Yang, H., Mockler, T. C. & Lin, C. Regulation of Flowering Time by *Arabidopsis*
473 Photoreceptors. *Science* **279**, 1360–1363 (1998).
- 474 39. Matsushika, A., Imamura, A., Yamashino, T. & Mizuno, T. Aberrant Expression of the Light-
475 Inducible and Circadian-Regulated *APRR9* Gene Belonging to the Circadian-Associated
476 *APRR1/TOC1* Quintet Results in the Phenotype of Early Flowering in *Arabidopsis thaliana*.
477 *Plant Cell Physiol.* **43**, 833–843 (2002).
- 478 40. Somers, D. E., Webb, A. A., Pearson, M. & Kay, S. A. The short-period mutant, *toc1-1*, alters
479 circadian clock regulation of multiple outputs throughout development in *Arabidopsis thaliana*.
480 *Development* **125**, 485–494 (1998).
- 481 41. Mizoguchi, T. *et al.* LHY and CCA1 Are Partially Redundant Genes Required to Maintain
482 Circadian Rhythms in *Arabidopsis*. *Dev. Cell* **2**, 629–641 (2002).
- 483 42. Zhang, Y., Cao, G., Qu, L.-J. & Gu, H. Characterization of *Arabidopsis* MYB transcription
484 factor gene *AtMYB17* and its possible regulation by LEAFY and AGL15. *J. Genet. Genomics*
485 **36**, 99–107 (2009).
- 486 43. Liu, H. *et al.* High-Throughput CRISPR/Cas9 Mutagenesis Streamlines Trait Gene
487 Identification in Maize. *Plant Cell* (2020) doi:10.1105/tpc.19.00934.
- 488 44. Michaels, S. D. & Amasino, R. M. FLOWERING LOCUS C Encodes a Novel MADS Domain
489 Protein That Acts as a Repressor of Flowering. *Plant Cell* **11**, 949–956 (1999).

- 490 45. Xiao, J. *et al.* Requirement of histone acetyltransferases HAM1 and HAM2 for epigenetic
491 modification of *FLC* in regulating flowering in *Arabidopsis*. *J. Plant Physiol.* **170**, 444–451
492 (2013).
- 493 46. Shin, J.-H. & Chekanova, J. A. *Arabidopsis* RRP6L1 and RRP6L2 Function in *FLOWERING*
494 *LOCUS C* Silencing *via* Regulation of Antisense RNA Synthesis. *PLoS Genet.* **10**, e1004612
495 (2014).
- 496 47. Gu, X. *et al.* *Arabidopsis* FLC clade members form flowering-repressor complexes
497 coordinating responses to endogenous and environmental cues. *Nat. Commun.* **4**, 1947 (2013).

Methods

Plant material

Twenty-six tissues or stages (Supplementary Table 1) were sampled from B73 maize plants in a greenhouse. The plants were cultivated under a 12-h light/12-h dark cycle at 25 °C. Three replicates were prepared per tissue. RNA was extracted from 5–10 individual plants per replicate. Equal amounts of the three replicates were combined and used for sequencing.

RNA sequencing platform

Total RNA was isolated from the 26 tissues using a Direct-zol RNA Microprep kit (Zymo Research) following the manufacturer's guidelines. Total RNA from each tissue was reverse-transcribed to cDNA, which was used as a template to prepare the mRNA-seq, sRNA-seq, and circRNA-seq libraries. The 26 RNA-seq libraries were prepared using Ribo-Zero rRNA Removal kits and TruSeq Stranded Total RNA Library Prep Plant (Illumina). The libraries were sequenced on the Illumina X-ten platform at the Beijing Genomics Institute (BGI). The 26 sRNA-seq libraries were prepared using an MGIEasy Small RNA Library Prep Kit (BGI) and sequenced on a BGISEQ-500 sequencer at BGI. The 26 circRNA-seq libraries were prepared using Ribo-Zero rRNA Removal kits and a TruSeq RNA Library Prep kit (Illumina). The circRNA libraries were sequenced on an X-ten platform at BGI.

RNA-seq analysis

Quality control of raw data

529 BGI used SOAPnuke1.6.5⁴⁸ to process the raw RNA-seq and circRNA-Seq data
530 with the parameters “-n 0.01 -l 20 -q 0.3 -A 0.25” and used SOAPnuke1.6.5⁴⁸ filters RNA
531 to process the raw sRNA-Seq data with the parameters “-p 1 -z 18 -s -Q 2 --sanger”.

532 *Identification and quantification of total RNA*

533 Total RNA-seq data from the 26 tissues or stages were mapped to the B73 AGPv4
534 reference genome using the default parameters of HISAT and StringTie^{49–51}. For each
535 tissue, a gtf file was generated and merged with an integrative annotation gtf file using
536 StringTie merge⁵⁰. TPM (Transcripts Per Million reads), a standardized measure of gene
537 expression, was calculated with a 0.03 mismatch rate using RSEM 1.3.1⁵².

538 *Identification and quantification of LncRNAs*

539 The improved version of LncRNA_Finder⁵³ was used to identify lncRNAs from
540 total RNAs. Transcripts with more than 300 nucleotides were removed, as >95% of the
541 proteins contained more than 100 amino acids⁵⁴. The remaining transcripts were assessed
542 by CPC2⁵⁵, which rapidly and accurately assesses the coding ability of RNAs to
543 discriminate and remove potential protein-coding transcripts. To remove housekeeping
544 RNAs, including tRNAs (transfer RNAs) (<http://gtrnadb.ucsc.edu/>), rRNAs (ribosomal
545 RNAs) (<http://maize.jcvi.org/>), snRNAs (small nuclear RNAs), and snoRNAs (small
546 nucleolar RNAs) (<http://noncode.org/>), the remaining transcripts were compared to these
547 housekeeping RNA databases. The remaining transcripts were lncRNAs. RSEM⁵² software
548 was used to quantify lncRNA expression with a 0.03 mismatch rate.

549 *Identification and quantification of fusion RNAs*

Mapssplice-v2.1.8⁵⁶ was used to identify fusion RNAs from total RNA-seq data from each tissue. The following filtering steps were performed to obtain high-confidence fusion RNAs. Firstly, PolyA sequences were removed using the following the criteria: the proportion of As and Ts was > 40%, and the proportion of As and Ts was > 80%. Secondly, to remove the fusion RNAs formed by homologous genes, the two parental genes (overlapping sequences > 50 bp) that formed the fusion RNA were identified, and the similarity between the parental genes was monitored. If the corresponding parental genes could not be found (overlapping sequences < 50 bp), the sequences were extended 500 bp up- and downstream of the splicing site, and the extended sequences were aligned the maize AGPv4 genome⁵⁷ to identify the parental genes. The remaining fusion RNAs were compared to the genome by Blast-2.2.18⁵⁷, and the number of overlapping bases of the splicing site was manually checked. Only fusion RNAs with < 7 overlapping bases were retained. Thirdly, since some fusion RNAs with high sequence similarity might be the same fusion RNAs, Blast⁵⁷ analysis was performed to divide the highly similar fusion RNAs into groups, and the fusion RNAs with the highest number of reads in each group were retained. Fourthly, to verify the identified fusion RNAs, their sequences were compared with their original sequences using TopHat 2.1.1⁵⁸, and RNAs that match at least 6 bases and had a read coverage of at least three at the junction site were retained. The normalised expression levels of fusion RNAs were calculated as

$$TPM_f = \frac{Ni \times 10^6}{Li \times \sum (Ni/Li + \dots + Nm/Lm)}$$

where Ni is the number of reads mapping to gene i , and Li is the sum of exon length of gene i .

572 *CircRNA identification and quantification*

573 To obtain clean reads from circRNA-seq, FastUniq⁵⁹ was used to remove duplicates
574 introduced by PCR amplification during high-throughput sequencing. CIRI v2.0.5⁶⁰ was
575 used to identify circRNAs by aligning them to the maize AGPv4 reference genome⁶¹, and
576 CIRCexplorer2 v2.0.1⁶² was used quantify the expression levels of the circRNAs. The
577 mean counts of junction reads per mapped billion reads (CPM) values were calculated as
578 the normalised expression level of each transcript. The circRNA CPM formula is as follows:

$$579 \quad CPM_c = \frac{Reads_j \times 10^6}{Reads_m}$$

580 , where Reads_j is the number of junction reads and Reads_m is the number of mapped reads.

581 *sRNA identification and quantification*

582 Small RNAs were identified and quantified as described by Wang et al.⁶³ with slight
583 modifications. After filtering out low-quality reads and eliminating those matching tRNAs,
584 rRNAs, snRNAs, and snoRNAs, the retained clean reads from 26 tissues or stages that
585 were 18–26 nt in size were mapped to the maize AGPv4 reference genome using Bowtie-
586 1.1.2 with the parameters unique location and no mismatch^{51,61,64,65}. All samples were
587 merged and the read number calculated using Samtools-1.5⁶⁶, and only sRNAs with more
588 than 6× coverage were retained. Given that sRNAs are distributed in clusters throughout
589 the maize genome and that gap distance is defined as the distance between two
590 neighbouring clusters, we compared the coverage of chromosomes between two
591 neighbouring sRNA clusters using different gap distances and selected a coverage of 6×

and a 300 bp gap distance as the thresholds for identifying sRNAs (Supplementary Fig. 1h,i). For each sample, the number of sRNA reads in each sRNA cluster was calculated and reconstructed as a matrix, with each row representing a sRNA cluster and each column representing a sample. Read abundances were normalised based on library size by scaling to TPM to allow for direct comparisons across libraries.

Ribo-seq: preparation, sequencing, and analysis

To extract polysomes, tissue (5 g) samples were pulverized and dissolved in polysome extraction buffer (44 mM Tris-HCl, pH 7.5, 175 mM KCl, 13 mM MgCl₂, 1% TritonX-100, 15 mM 2-mercaptoethanol, 100 µg/mL cycloheximide, 10 units/mL DNase I). The polysome pellets were treated with RNase I (10 units/µg RNA) at room temperature for 1 h. After terminating the reaction using RNase Inhibitor (20 units/µL), the solution was immediately transferred into a MicroSpinS-400 column to enrich the RNA-ribosome complex (monosomes). Ribosome-protected fragments were extracted from the samples using a miRNeasy RNA isolation kit (Qiagen) following the manufacturer's instructions. After removing rRNAs, the remaining RNAs were used to construct libraries, which were sequenced on the Illumina HiSeq 2500 platform.

After removing adaptors and quality control using FASTX_Toolkit-0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/index.html) fastx_clipper with the parameters “-l 5 -c -n -v -Q33” and fastx_trimmer with the parameters “-f 1 -Q 33”, the Ribo-Seq reads were aligned to the rRNA reference sequence from NCBI using Bowtie-1.1.2⁶⁷ to filter out reads derived from rRNAs. The Ribo-Seq reads were mapped to the maize AGPv4 reference genome⁶¹. The expression levels were measured by TPM using RSEM (version

1.3.0)⁵². Only reads that mapped to the coding sequence were used to calculate translational abundance.

Protein–protein interactions

PCR amplification

Total RNA was extracted from 20 tissues separately using a Direct-zol RNA Miniprep kit (Zymo Research). The total RNA was divided into eight groups (Supplementary Table 3) based on tissue type. Equal amounts of RNA from each group were pooled. Approximately 2 µg pooled RNA was used to generate first-strand cDNA using Clontech's SMART™ technology²¹ according to Clontech's Make Your Own "Mate & Plate™" Library System User Manual (Protocol No. PT4085-1) with modified primers. These modified primers were homologous to modified pGADT7 (mAD) or modified pGBKT7 (mBD) vectors from a previously reported recombination-based library versus library high-throughput yeast two-hybrid (RLL-Y2H) screening system⁸ and included modified SMART III oligo primers (CCATACGACGTACCAGATTACGCTCC/rG//rG//rG/ [rG represents a ribonucleotide] for the mAD vector and AAGCTGATCTCAGAGGAGGACCTGCC/rG//rG//rG/ for the mBD vector) and modified oligo-dT primers (GGGCGTTGACACTAGTCTCGAGGATC(T)₃₀VN, N=A, G, C, or T; V=A, G, or C) for the mAD vector and ACCCGCACGCTAACTAGTGTCGAC(T)₃₀VN for the mBD vector). The first-strand cDNA was amplified to obtain full-length double-stranded cDNA using a KAPA HiFi HotStart ReadyMix PCR kit with specific primers (Forward primers: AGCGCCGCCATGGAGTACCCATACGACGTACCAGATTACGCT and

636 GCCATCATGGAGGAGCAGAAGCTGATCTCAGAGGAGGACCTG for the mAD and
637 mBD vectors, respectively; Reverse primers:
638 CCTTTGAGTTCTCTCAGTTGGGGGCGTTGACACTAGTCTCGA and
639 CCAAGGGCACGCCCTGGCACCCGCACGCTAACTAGTGTCGAC for the mAD and
640 mBD vectors, respectively). Next, the full-length double-stranded cDNA was normalised
641 as described previously⁶⁸, and 10 µg of normalised cDNA was used for each yeast
642 transformation.

643 *Yeast library construction*

644 Modified mAD and mBD vectors were used to construct a maize cDNA yeast
645 library according to the Clontech Y2H transformation method⁸. The mAD and mBD
646 vectors were digested with the restriction enzymes BamHI and NdeI. Each 10 µg PCR-
647 amplified cDNA fragment was mixed with 5 µg linearized mAD or mBD vector and used
648 to co-transform 6 ml Y187 and Y2H yeast-competent cells, respectively, with the
649 Yeastmaker Yeast Transformation System (Clontech) following the manufacturer's
650 protocol. The cells were spotted onto SD -Leu (mAD clones) or SD -Trp + X-α-Gal (mBD
651 clones) medium and grown for 3–4 days at 30 °C. All blue clones including self-activated
652 functional fusion genes in the BD library were removed.

653 *Yeast mating*

654 To identify positive yeast mating, 2 ml of Y2H and Y187 yeast libraries (at least 2
655 ×10⁷ cell each library) were combined in 50 ml 2× YPDA liquid medium and incubated at
656 30 °C for 24 h at a shaking speed of 50 rpm. The yeast cells were collected and resuspended
657 in 6 ml 0.9% NaCl, and 300 µL aliquots were plated onto a 15 cm-diameter cell-culture

dish containing SD/-Trp, -Leu, -His, -Ade selective solid medium and grown for 5–7 days at 30°C. Each library mating was repeated at least 10 times. A barcode sequence was added to each repeat, and ten repeats were pooled to construct a sequencing library.

Yeast plasmid isolation and PPI sequencing

Following yeast mating, the clones on plates for each repeat were collected and the fusion plasmids extracted from the yeast cells using a yeast plasmid kit (OMEGA). Barcoded (Third-generation sequencing barcodes) vector primers (Supplementary Table 14) were used to amplify the plasmids for each repeat. This PCR products were further purified using DNA-free beads to remove fragments shorter than 0.75 kb.

BiFC assays

To validate the positive rate of the PPI data, 114 protein–protein interactions were randomly chosen from three libraries and the corresponding cDNA fragments cloned in the Pxy104 and Pxy106 vectors using a ClonExpress II One Step Cloning Kit (Vazyme, Nanjing, China). These BiFC constructs were transformed into *Agrobacterium tumefaciens* strain GV3101 and co-transfected into *Nicotiana benthamiana* leaf epidermal cells. The YFP signals in infiltrated leaves were observed and photographed 48 h later under an SP8 confocal microscope (Leica) with 488 nm excitation and 515–555 nm emission filters.

PPI analysis

The libraries were sequenced on the PacBio platform, and CCS (Circular Consensus Sequencing) reads were called using SMRTLink-5.1.0.26412 (<https://www.pacb.com/support/software-downloads/>). The following analysis was

679 performed to obtain PPIs: Firstly, reads in FASTA format were extracted from BAM files
 680 using SAMtools-1.9⁶⁶. Next, because ten barcoded samples were combined to generate
 681 each sequencing library, reads from each sample were divided based on their barcode
 682 sequences using Cutadapt-1.9.1⁶⁹. Parameters -e and -m for running reads separation were
 683 0.3 and 200, respectively. The reads were aligned to the coding sequence data from the
 684 maize AGPv4 genome to identified candidate PPI genes using BLAST+/2.7.1⁵⁷. Only PPI
 685 genes with BLAST e-value $< 1.0e^{-4}$ that matched to two genes were retained. Lastly, the
 686 mAD-primer (CATACGACGTACCAGATTACG), mBD-primer
 687 (CATCATGGAGGAGCAGAAGCT), and recombination sequences
 688 (TAGCGTGCGGGTGCCAGGGCGTGCCCTTGAGTTCTCTCAGTTGGGGGCGTTG
 689 AC) were aligned to the retained CCS reads using needle
 690 (<http://biopython.org/DIST/docs/tutorial/Tutorial.html#htoc86>). If a CCS contained
 691 sequences from two different genes, and the primer sequence and the recombination
 692 sequence were aligned consistently, the proteins encoded by the two genes were considered
 693 to be interacting proteins.

694 The PPIs were divided into three confidence groups based on the probability of
 695 self-activation. Low-confidence PPIs indicates interaction edges detected using offline data.
 696 Because the mating of yeasts with empty vectors generated interaction signals (self-
 697 activation), to remove edges caused by self-activation, all PPIs involving eight empty BD
 698 libraries were removed to obtain the middle-confidence PPIs. To further remove false PPIs
 699 introduced by self-activation, PPIs identified using empty vectors were used as the positive
 700 training data, true PPIs confirmed by BiFC were used as the negative training data, and the
 701 frequencies of each gene in the AD and BD libraries were used as training features to

estimate the self-activating genes using the M5P machine learning classification algorithm (<https://cran.r-project.org/web/packages/RWeka/index.html>). Edges for self-activating genes in the BD library were removed, and the remaining PPIs were considered to be high-confidence PPIs.

Construction of co-expression networks

An expression matrix of 179,309 components (25,319 genes, 5,469 lncRNAs, 45 fusion RNAs, 4,318 circRNAs, and 144,158 small RNAs) detected in more than 10 of 26 distinct tissues or stages were used to construct a transcriptome co-expression network. To reduce the bias of Gini Correlation Coefficient (GCCs) caused by overexpressed transcripts, TPM values were transformed by a logarithmic function. GCCs between all pairs of all elements ($179,309 \times 179,309$) were calculated using the adjacency matrix `rsgcc_1.0.6`⁷⁰ with “saveType=“bigmatrix””, which is suitable for big data. Any edge with GCCs < 0.3 was excluded. Next, Mutual Rank_{a,b}, i.e., the mutual rank between element a and element b, was employed to assess the correlation of co-expression according to the formula^{71,72} :

$$\text{Mutual Rank}_{a,b} = \sqrt{\text{rank}(GCC_a \rightarrow GCC_b) \times \text{rank}(GCC_b \rightarrow GCC_a)}$$

Mutual correlation was divided into three confidence levels based on the relationship between gene number and MR (Supplementary Fig. 3). MR values at rapidly increasing intervals, slowly increasing intervals, and intervals reaching a steady level were defined by dividing by the high-, middle-, and low confidence values, respectively (Supplementary Fig. 3).

Finally, the Edge Weight (EW) between node a and node b was converted by $MR_{a,b}$ values. Larger EW values indicate greater correlation according to the formula:

$$\text{Edge Weight}_{a,b} = e^{-\frac{MR_{a,b} - \min(MR_{a,b})}{\max(MR_{a,b}) - \min(MR_{a,b})}}$$

, where EW ranges from $\frac{1}{e}$ to 1.

To obtain a network containing high-confidence small RNAs with potential functional implications, co-expression networks were constructed using 25,300 genes, 5,469 lncRNAs, 14 fusion RNAs, 4,318 circRNAs, and 159 micro RNAs. High, middle, and low-confidence edges were identified as described above. In addition, a co-expression network only including 24,394 annotated genes was constructed named the Slim-transcriptome network; high, middle, and low-confidence edges were identified in this network as described above.

Construction of an integrative omics database

Four omic data sets, including ChIA-PET data from Peng et al.⁹, a co-expression network constructed using RNA-Seq data from 26 tissues or stages, a co-expression network constructed using Ribo-seq data from 20 tissues or stages, and PPIs from yeast-2-hybrid screening were used to construct an integrative omic database including all possible temporal and spatial interactions. An integrative EW (EW_i) was calculated as detailed below. If an edge was present in one omic data set, the EW was raised to the fourth power to obtain the EW_i . If an edge was present in two omic data sets, EW_i was calculated by multiplying the square of the average EW by each EW. If three edges were present, EW_i was calculated by multiplying the average EW by each EW. If four edges were present,

EW_i was calculated by multiplying each EW¹². Each integrative omic edge included four (genome, transcriptome, translome, and proteome) corresponding omic edges. Because each omic data point had three corresponding confidence levels, three EW_i values representing three confidence levels were ultimately obtained.

Machine learning

Thirty-nine FT genes (related to flowering time) from the literature were collected as the positive training data, and 39 non-FT genes selected from the Classical Genes of the MaizeGDB database (<https://www.maizegdb.org/>) were used as the negative training data (Supplementary Table 9). Training data included the TPM of 26 tissues or stages in the transcriptome, the TPM of 20 tissues or stages in the translome, and the SD values of the 78 genes in the proteome. A matrix was assembled as the input file, in which the first row listed the genes to predict and the first column listed the 124 training features. The Scikit-learn package (version 0.21.2; <https://scikit-learn.org/stable/>) was used to identify FT and non-FT genes with default parameters. Three models of machine learning were used to train and extract the data attributes: LogisticRegression, Support Vector Machine, and adaboost (<https://scikit-learn.org/stable/>). Five-fold cross-validation was used to assess the prediction accuracy, and ROC-AUC (Receiver-Operating Characteristic Area Under the Curve), a standard performance evaluation criterion used for two-class pattern classification problems, was used to compare different classification algorithms.

Phenotypic verification

CRISPR-based genome editing

764 Mutants of *Zm00001d006918*, *zmspl6* (*Zm00001d042319*) *Zm00001d017241*, and
765 *Zm00001d021291* were generated using a high-throughput genome-editing system³⁰. In
766 brief, line-specific sgRNAs (single guide RNAs) were designed based on the assembled
767 pseudo-genome of the receptor KN5585. A double sgRNAs pool (DSP) approach was used
768 for vector construction. The receptor KN5585 was transformed with the vectors, and the
769 targets of each T0 plant were examined by barcode-based sequencing. The genotype of
770 each gene-edited line was identified by PCR amplification and Sanger sequencing using
771 target-specific primers (Supplementary Table 15). The flowering times of the gene-edited
772 lines and the receptor KN5585 were investigated, including days to tasselling (DTT), days
773 to anthesis (DTA), and days to silking (DTS).

774 Positive transgenic lines of the four genes were evaluated for flowering time,
775 including *Zm00001d006918*, *ZmSPL6* (*Zm00001d042319*), *Zm00001d017241*, and
776 *Zm00001d021291*. In the summer of 2018 in Gongzhuling city, Jilin province (43°30'N
777 124°49'E), the homozygous transgene-positive lines of *Zm00001d006918* (T2 generation)
778 and *Zm00001d017241* (T1 generation) and their negative counterparts were evaluated for
779 flowering time. In the summer of 2018 in Gasa town, Xishuangbanna dai autonomous
780 prefecture, Yunnan province (21°57'N 100°45'E), the homozygous transgene-positive
781 lines of *ZmPRR95a* (T1 generation) and its negative counterparts were evaluated for
782 flowering time. In the summer of 2019 in Gongzhuling city, Jilin province (43°30'N
783 124°49'E), the homozygous transgene-positive lines of *zmspl6* (T4 generation) and
784 *Zm00001d021291* (T4 generation) and their negative counterparts were evaluated for
785 flowering time. The homozygous transgene-positive lines of *Zm00001d006918* (T6
786 generation), *zmspl6* (T5 generation), and *Zm00001d017241* (T5 generation) and their

negative counterpart were evaluated for flowering time in the Winter of 2019 in Foluo town, Sanya City, Hainan Province (18°34'N 108°43'E).

EMS mutants

To verify the functions of the candidate genes, 54 mutants with SNP mutations leading to early translational stops, one mutant with an SNP mutation that changed the start codon, and one mutant with an SNP mutation at the splice site from a maize EMS mutant library⁷³ were obtained. The mutants were planted in Sanya, Hainan, Zibo, Shandong, and Langfang, Hebei, China. DNA was extracted from mutant seedlings and used to genotype each plant and to identify the homozygous mutant plants and their wild-type counterparts. DTT, DTA, and DTS of the mutant and wild-type plants were evaluated, and the results were subjected to one-tailed *t* test to compare the significance of their differences.

Other analysis

Four attributes, including Degree, Shortest distance, Average distance, and Transitivity were calculated for each network using the R package igraph-1.2.4.2 (<https://igraph.org/r/>). Gene ontology analysis was performed using AGRIGO version2 (<http://systemsbiology.cau.edu.cn/agriGOv2/index.php>). Interactions in the network were clustered using the Markov Cluster Algorithm (<https://github.com/JohannesBuchner/mcl>). Gephi (<https://gephi.org/>) was used for visualization and feature extraction of the co-expression networks.

Supplementary References

48. Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L. & Rice, P. M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* **38**, 1767–1771 (2010).
49. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
50. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
51. Wang, B. *et al.* Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* **7**, 1–13 (2016).
52. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
53. Li, L. *et al.* Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.* **15**, R40 (2014).
54. Frith, M. C. *et al.* The Abundance of Short Proteins in the Mammalian Proteome. *PLoS Genet.* **2**, e52 (2006).
55. Kang, Y.-J. *et al.* CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**, W12–W16 (2017).
56. Wang, K. *et al.* MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* **38**, e178–e178 (2010).
57. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).
58. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
59. Xu, H. *et al.* FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads. *PLoS One* **7**, e52249 (2012).
60. Gao, Y., Wang, J. & Zhao, F. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.* **16**, 4 (2015).
61. Jiao, Y. *et al.* Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017).
62. Zhang, X.-O. *et al.* Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* **26**, 1277–1287 (2016).

- 839 63. Wang, J., Yao, W., Zhu, D., Xie, W. & Zhang, Q. Genetic basis of sRNA quantitative
840 variation analyzed using an experimental population derived from an elite rice hybrid. *eLife* **4**,
841 e03913 (2015).
- 842 64. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient
843 alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
- 844 65. Crisp, P. A. *et al.* Variation and Inheritance of Small RNAs in Maize Inbreds and F1 Hybrids.
845 *Plant Physiol.* **182**, 318–331 (2020).
- 846 66. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–
847 2079 (2009).
- 848 67. Langmead, B. Aligning Short Sequencing Reads with Bowtie. *Current Protocols in*
849 *Bioinformatics* **32**, 11.7.1–11.7.14 (2010).
- 850 68. A. Bogdanova, E., A. Shagin, D. & A. Lukyanov, S. Normalization of full-length enriched
851 cDNA. *Mol. Biosyst.* **4**, 205–212 (2008).
- 852 69. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
853 *EMBnet.journal* **17**, 10–12 (2011).
- 854 70. Ma, C. & Wang, X. Application of the Gini Correlation Coefficient to Infer Regulatory
855 Relationships in Transcriptome Analysis. *Plant Physiol.* **160**, 192–203 (2012).
- 856 71. Obayashi, T. & Kinoshita, K. Rank of Correlation Coefficient as a Comparable Measure for
857 Biological Significance of Gene Coexpression. *DNA Res.* **16**, 249–260 (2009).
- 858 72. Wisecaver, J. H. *et al.* A Global Coexpression Network Approach for Connecting Genes to
859 Specialized Metabolic Pathways in Plants. *Plant Cell* **29**, 944–959 (2017).
- 860 73. Lu, X. *et al.* Gene-Indexed Mutations in Maize. *Mol. Plant.* **11**, 496–504 (2018).
- 861 74. Qiao, X. *et al.* Gene duplication and evolution in recurring polyploidization–diploidization
862 cycles in plants. *Genome Biol.* **20**, 38 (2019).
- 863 75. Freeling, M. Bias in Plant Gene Content Following Different Sorts of Duplication: Tandem,
864 Whole-Genome, Segmental, or by Transposition. *Annu. Rev. Plant Biol.* **60**, 433–453 (2009).
- 865 76. Shao, H., Wang, H. & Tang, X. NAC transcription factors in plant multiple abiotic stress
866 responses: progress and prospects. *Front. Plant Sci.* **6**, (2015).

Supplementary

Supplementary Text

Regulatory divergence between duplicate genes

Speciation and domestication were accompanied by whole-genome duplication (WGD), tandem duplication (Tandem), proximal duplication (Proximal), transposed duplication (Transposed) mediated by transpositions, and dispersed origination (Dispersed)^{74,75}. In our interactome, gene interaction modules contain genes with functionally conserved roles, and Shortest Distance (SD) (<https://igraph.org/r/doc/distances.html>), which represents the shortest path between a pair of nodes, describes the functional divergence for any pair of duplicate genes. As expected, all duplicates had significantly higher functional conservation and lower functional divergence compared to randomly selected genes, supporting the validity of our first-generation interactome of maize. When we compared duplicated genes, we observed consistent tendencies for both functional conservation and divergence (Supplementary Fig. 8a-8d): the functions of Tandem and Proximal genes were more likely to be conserved, whereas the functions of Transposed genes were least likely to be conserved (Supplementary Fig. 8a-8d). WGD, Tandem, and Proximal genes had the lowest SD values and the least functional divergence, while Transposed and Dispersed genes had the highest SD values and the most functional divergence (Supplementary Fig. 8a-8d). To dissect the regulatory divergence of the maize subgenomes, we compared the degree (number of connections with its directly adjacent edges in the interactome) values of genes at different omic layers. The genes were classified into four groups, Maize1 genes with retained

Maize2 orthologs, Maize1 genes without retained Maize2 orthologs, Maize2 genes with retained Maize1 orthologs, and Maize2 genes without retained Maize1 orthologs. The degree variations of four different gene groups have been calculated and compared (Supplementary Fig. 8g,8h). Additionally, the dominant subgenome bins were obtained through Chi-square test on the degree to which the proportion of genes in the subgenome (out of 100) in a sliding window is significantly higher than that of the corresponding subgenome genes displayed based on the gene order along sorghum chromosomes (Supplementary Fig. 9).

Example of hub genes suffering from selection

The transcriptome-level Maize2 hub gene *ZmNAC75*, encoding a NAC domain transcription factor, interacted with four regulatory modules. Genes from three *ZmNAC75*-regulated modules were significantly enriched in the gene ontology (GO) terms stomatal movement, ion transport, and ABA-activated pathway. Many stress-responsive NAC TFs have been used to improve stress tolerance in crops⁷⁶. The regulatory networks of *ZmNAC75* reveal the molecular basis of stress responses to some extent (Supplementary Fig. 11a). Analysis of the nucleotide diversity of wild maize relatives, landraces, and modern inbreds indicated that *ZmNAC75* has been a selection target during modern maize improvement (Supplementary Fig. 11b).

Supplementary Fig. 1. Expression-level variation across 26 different stages/tissues for all different types of transcripts. a. Sampling of 26 stages/tissues spanning the maize whole life period. b. Transcriptome landscape of all 26 different stages/tissues for all different types of transcript isoforms. c-f. Stage/tissue-specific expression of protein-coding, long noncoding, fusion, and circular transcripts. g. Stage/tissue-specific abundance of ribosome-imprinted transcripts. h. Genome coverage for small RNAs at different cutoffs.

Supplementary Fig. 2. The pipeline of high-throughput yeast-2-hybrid for the identification of protein-protein interactions. This pipeline consists of three major steps, yeast cDNA library construct, yeast mating and data analysis. Detailed methods can be found in the Method part.

Supplementary Fig. 3. Methods for the construction of transcriptomic and translatomic interactions. a-d. Saturation-levels across different mutual ranks of co-expression relationships and cutoffs for the identification of high-confidence, middle-confidence, and low-confidence co-expression networks.

Supplementary Fig. 4. Overlapping of hub genes and edges between different omic levels at different confidences. a-b. Number of total and overlapping hub genes among transcriptome, translatome, and proteome at middle- and low-confidence levels. c-h. Shared edges between transcriptome, translatome, and proteome in our study and the randomly generated interaction datasets at all different confidence levels. The sharing rates of edges in our study (c, d, e) is significantly higher than that in the randomly generated interaction datasets (f, g, h).

Supplementary Fig. 5. The numbers of modules which contain types of functional elements in transcriptome and integrative networks. The x-axis represents how many different types of functional elements are contained in the module and the total number of modules.

Supplementary Fig. 6. Overlapping between reported interactions and our current study and the validation rate by BiFC of protein-protein interactions, and saturation curve for each library across different number of matings.

Supplementary Fig. 7. Validation of randomly selected PPIs by BiFC.

953

954

955

956

957

958

959

960

961

962

963

964

965

966

Supplementary Fig. 8. Functional divergence between duplicate genes of different origination types and progressive differentiation between two maize subgenomes at middle- and low- confidence levels. a-f, The proportion of different types of duplicate genes that are functionally clustered (left) in the same module across different regulatory levels of high- (a, b), middle- (c, d) and low-confidences (e, f) and distribution of shortest distance (right) in different regulatory networks for different types of duplicate genes of high-, middle- and low-confidences. g-h. Degree variation of maize subgenome genes across different regulatory levels of middle- and low-confidences. Maize1 and Maize2 represent homologous genes from subgenome maize1 and maize2, respectively. Maize1 without maize2 indicates subgenome maize1 genes, of which the corresponding maize2 genes lost. Maize2 without maize1 indicates subgenome maize2 genes, of which the corresponding maize1 genes lost.

Supplementary Fig. 9. Biased regulatory fractionation at transcriptomic, translomic, and proteomic levels is observed for the reconstructed, or “sorghumized,” pair of maize ancestral chromosome2 to chromosome10.

Supplementary Fig. 10. Number of sliding window bins (at different step sizes of 10 and 50) showing significant functional divergence between two subgenomes.

Supplementary Fig. 10. Maize2 hub gene *ZmNAC75* was subject to selection during maize improvement. a, A regulatory network of maize2 hub gene *ZmNAC75* at the transcriptome level. *ZmNAC75* cross-talks to four regulatory modules, of which three have significant gene ontology enrichment for ABA activated pathway, stomatal movement, and ion transport, respectively. b, Nucleotide diversity across *ZmNAC75* calculated using maize HapMapV3 data. Nucleotide diversity (π) for teosinte (blue), maize (red), and maize landrace (green) was calculated using a 100-bp sliding window with 25-bp step.

Supplementary Fig. 12. Mutagenesis in *zmalog1*, *zmalog2*, *nam1* and *nam2*.

Supplementary Fig. 13. The prediction rate of regulatory network extracted from the draft interactome by a randomly selected subset of well-known kernel functional genes (Kernel) and by a randomly selected genome-wide genes (Random) as queries. a, The constructed subnetworks of 40 randomly selected kernel genes (1000 time simulations) could predict up to 40% of the validation kernel genes, which is significantly higher ($P = 0.00$) that of randomly selected function-unrelated genes. b, Mutagenesis in *Zm00001d003047*.

Supplementary Fig. 14. Reconstruction of the regulatory network for well-known pathways in maize. a, Reconstruction of the regulatory network for starch synthesis pathway in maize. b, Reconstruction of the regulatory networks for meristem developmental pathway. c, Reconstruction of the regulatory networks for abiotic stress related genes.

Supplementary Fig. 15. Percentage of predicted flowering genes supported by other related flowering time evidence. Homolog means genes that are homologs to Arabidopsis FT genes identified in flowering time database (<http://www.phytosystems.ulg.ac.be/florid/>). SNPs-GWAS means genes with FT association signals identified in maize by Liu et al., 2020³⁴. Others represent genes without known functional evidence.

Zm00001d006236

Supplementary Fig. 16. Expression of predicted FT gene Zm00001d006236.

Supplementary Fig. 17. Functionally validation of predicted FT genes *Zm00001d024126* by EMS mutant. a. An early stop codon was introduced in the exon region of the gene *Zm00001d024126*. b. Expression-level variation of the FT gene at both transcriptome and translome levels. c. Significant FT variation between wildtype and mutant at growing locations and years. The x-axis is displayed by year.location.season. C means spring. WH means Wuhan, China. DTT means days to tassels. DTA means days to anthesis. DTS means days to silking.

Supplementary Fig. 18. Functionally validation of predicted FT genes *Zm00001d013402* by EMS mutant. a. An early stop codon was introduced in the exon region of the gene *Zm00001d013402*. Different color boxes represent different functional domain annotated by Interproscan. b. Expression-level variation of the FT gene at both transcriptome and translome levels. c. Flower time differs between wildtype and mutant. d. Significant FT variation between wildtype and mutant at different growing locations and years. The x-axis is displayed by year.location.season. X means summer. LF means Langfang, China. SD means Shandong, China. DTT means days to tassle. DTA means days to anthesis. DTS means days to silking.

Supplementary Fig. 19. Functionally validation of predicted FT genes *Zmphyb1* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 20. Functionally validation of predicted FT genes *Zm00001d033799* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 21. Functionally validation of predicted FT genes *ZmphyC1* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 22. Functionally validation of predicted FT genes *Zm00001d045944* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 23. Functionally validation of predicted FT genes *Zm00001d021291* by CRISPR gene editing. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 24. Functionally validation of predicted FT genes *Zm00001d006212* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 25. Functionally validation of predicted FT genes *Zm00001d017241* by CRISPR gene editing. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 26. Functional validation of predicted FT genes *Zm00001d049543* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 27. Functionally validation of predicted FT genes *Zm00001d034313* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 28. Functionally validation of predicted FT genes *Zm00001d006198* by EMS mutant. Legends are same to Supplementary Fig. 16.

Supplementary Fig. 29. Functional validation of predicted FT genes *Zm00001d010625* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 30. Functionally validation of predicted FT genes *Zm00001d011748* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 31. Functionally validation of predicted FT genes *Zm00001d017193* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 32. Functionally validation of predicted FT genes *Zm00001d032894* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 33. Functionally validation of predicted FT genes *Zm00001d006918* by CRISPR gene editing. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 34. Functionally validation of predicted FT genes *Zm00001d006131* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 35. Functionally validation of predicted FT genes *Zm00001d045481* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 36. Functionally validation of predicted FT genes *Zm00001d033680* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 37. Functionally validation of predicted FT genes *Zm00001d042319* by CRISPR gene editing. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 38. Functionally validation of predicted FT genes *Zm00001d013863* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 39. Functionally validation of predicted FT genes *Zm00001d020157* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 40. Functional validation of predicted FT genes *Zm00001d034964* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 41. Functionally validation of predicted FT genes *Zm00001d036801* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 42. Functionally validation of predicted FT genes Zm00001d038644 by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 43. Functionally validation of predicted FT genes *Zm00001d039589* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 44. Functionally validation of predicted FT genes *Zm00001d049305* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 45. Functionally validation of predicted FT genes *Zm00001d030968* by EMS mutant. Legends are same to Supplementary Fig. 17.

Supplementary Fig. 46. Percentage of genes from eight pathways enriched to selection target groups of domestication and improvement in maize. *, and ** indicate P-values < 0.05, 0.01 by Chi-square enrichment test, respectively.

1379 **Supplementary Tables**

1380 **Supplementary Table 1.** Detailed transcriptome dataset collected on the whole lifespan
1381 of reference maize inbred B73 in our study.

1382 **Supplementary Table 2.** Number of interactions at different layers of genetic information
1383 flow for different confidence levels.

1384 **Supplementary Table 3.** Eight distinct tissues for the construction of 8 bait and 8 prey
1385 cDNA libraries were applied for interaction screens of library vs. library mating using the
1386 RLL-Y2H system.

1387 **Supplementary Table 4.** All protein-protein interactions detected across different matings
1388 and different tissues.

1389 **Supplementary Table 5.** Detailed information of protein-protein interactions validated by
1390 BiFC.

1391 **Supplementary Table 6.** Detailed information of kernel related genes in maize.

1392 **Supplementary Table 7.** Well-known meristem development related genes in maize.

1393 **Supplementary Table 8.** Genes in a well-known abiotic stress related pathway.

1394 **Supplementary Table 9.** Detailed information of training positive and negative FT genes.

1395 **Supplementary Table 10.** All the predicated FT genes in maize.

1396 **Supplementary Table 11.** Statistical summary of FT variation for the validated FT genes.

1397 **Supplementary Table 12.** Eight well-curated FT pathways were uncovered in our study.

1398 **Supplementary Table 13.** Functional annotation of eight FT pathways.

1399 **Supplementary Table 14.** Primer information for the high throughput protein-protein
1400 interaction screening.

1401 **Supplementary Table 15.** Primer information for the genotyping of CRISPR gene editing
1402 mutations.

1403

1404

Figures

Figure 1

The interactome uncovers the genome-wide regulatory landscape in maize. a, Data collection. Complete omics data covering all layers of genetic information flow were collected from maize inbred B73 throughout its entire lifecycle. The numbers in the cells of the matrix indicate the number of tissues collected for each of the four types of high-throughput data. b, Distribution and proportion of different types of expressed RNAs across the maize genome. c–f, Genome-wide landscape of interactions in the genome, transcriptome, translatome, and proteome, respectively. Numbers in each panel (from top to bottom) represent the number of functional elements, interactions (edges), and modules. g, Total number of hub genes and number of overlapping hub genes (nodes above the 90th percentile for the number of edges) among the transcriptome, translatome, and proteome. h, Genome-wide landscape of integrative

interactome based on multi-omics. i, An example of an interaction module for the well-known gene *tga1* involving lncRNAs (orange long line), circRNAs (orange circle), small RNAs (orange short line), and protein-coding mRNAs (orange oval). Solid and dotted lines represent direct and indirect interactions, respectively.

Figure 2

Asymmetric regulatory divergence indicates differences in evolutionary selection between the two maize subgenomes. a, Degree of variation of genes in subgenomes Maize1 and Maize2 across different regulatory levels. Maize1 and Maize2 represent homologous genes from subgenome Maize1 and Maize2, respectively. Maize1 without Maize2 indicates Maize1 genes whose corresponding Maize2 genes were lost. Maize2 without Maize1 indicates Maize2 genes whose corresponding Maize1 genes were lost. b, Biased regulatory fractionation at the transcriptome (top), translatome (middle), and proteome (bottom) levels is observed for the reconstructed or “sorghumized” pair maize ancestral chromosome–chromosome 1. Dominant bin is defined as a genome region in which more genes in one subgenome have a significantly higher degree of bias than the corresponding homologous genes in the other subgenome. Lines are LOESS regression lines. The pie chart to the right represents the number of dominant bins from each subgenome. c, Enrichment of Maize1 and Maize2 hub genes at the transcriptome (top), translatome (middle), and proteome (bottom) levels at loci targeted for domestication and improvement. *, **, and *** indicate P-values < 0.05, 0.01, and 0.001 by Chi-square test, respectively. Grey, yellow, and blue bars represent genome-wide, Maize1, and Maize2, respectively.

Figure 3

The interactome can predict gene function and re-construct regulatory networks underlying complex traits. a, A network showing that the well-known tillering-related genes *tb1*, *tru1*, and *gt1* are connected to two novel genes (*ZmALOG1* and *ZmALOG2*). b, Knockout mutants of *zalog1* and *zalog2* show altered tillering phenotypes. c, Protein interactions between *ZmALOG1* and 2 and *TRU1*, *TB1* were confirmed by Yeast-2-Hybrid assays. d, A network showing that *CUC3*, a gene known to function in lateral tissue

development, is connected to two novel genes, NAM1 and NAM2. e, CRISPR/Cas9 knockout mutants of NAM1 and NAM2 show suppressed development of lateral meristem. f, Construction of regulatory networks to uncover the molecular mechanism conferring kernel development. The nodes contain genes with previously reported functions in kernel development. Different colours represent different modules associated with kernel variation. g, A gene of unknown function (Zm00001d003047) was identified in the newly constructed regulatory network based on the interactome of well-known kernel genes. h, The shrunken kernel phenotype of the knockout mutant (Zm00001d003047) co-segregated with the genotypes, whose phenotype and genotypes fit a 3:1 segregation ratio.

Figure 4

The interactome facilitates systematic dissection of the molecular mechanisms underlying flowering time in maize. a, Distribution of the AUC (Area Under Curve) values of different layers of the interactome for machine learning to predict flowering time (FT)-related genes. b, Distribution of SD values of predicted FT genes and randomly selected genes. SD represents the shortest distance to well-known FT genes. c, The percentage of genes with validated functions in loss-of-function mutants. d–f, Validation of the function

of Zm00001d006236. d, An early stop codon mutation occurred in the last exon of Zm00001d006236. e. Zm00001d006236 mutants show delayed flowering time compared to the wild type. f, Statistical comparison of FT variation between the wild type and Zm00001d006236 mutants show that the mutant has significantly delayed FT compared to wild-type plants. **, and *** indicate P-values < 0.01, and 0.001 by Student's T-test, respectively. g, The 30 FT genes were shown to be significantly associated with FT variation in multiple years or locations. The label in y-axis shows year.location.season.trait. LF: Langfang, Hebei, China; SD: Zibo, Shandong, China; JL: Gongzhuling, Jilin, China; YN: Xishuangbanna, Yunnan, China; HN: Sanya, Hainan, China; C: spring; X: summer; D: winter; T: DTT (days to tasselling); A: DTA (days to anthesis), S: DTS (days to silking). h, All validated FT genes (including well-known genes and newly validated genes in this study) and their interaction genes could be classified into eight pathways. The 30 validated genes were highlighted in red. i, Genetic effects of 21 FT genes in the same growing location. Different colours of gene ID on the x-axis represent different pathways in Fig. 4h.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.xls](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable4.xlsx](#)
- [SupplementaryTable5.xlsx](#)
- [SupplementaryTable6.xlsx](#)
- [SupplementaryTable7.xlsx](#)
- [SupplementaryTable8.xlsx](#)
- [SupplementaryTable9.xlsx](#)
- [SupplementaryTable10.xlsx](#)
- [SupplementaryTable11.xlsx](#)
- [SupplementaryTable12.xlsx](#)
- [SupplementaryTable13.xlsx](#)
- [SupplementaryTable14.xlsx](#)
- [SupplementaryTable15.xlsx](#)