bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A co-expression network in hexaploid wheat reveals mostly balanced expression and lack of significant gene loss of homeologous meiotic genes upon polyploidization.

Abdul Kader Alabdullah[1*], Philippa Borrill[2], Azahara C. Martin[1], Ricardo H. Ramirez-Gonzalez[1], Keywan Hassani-Pak[3] Cristobal Uauy[1], Peter Shaw[1] and Graham Moore[1].

[1] John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom

[2] School of Biosciences, University of Birmingham, Birmingham B15 2TT, United Kingdom

[3] Rothamsted Research, Harpenden AL5 2JQ, United Kingdom

* AbdulKader.Alabdullah@jic.ac.uk

Abstract

Polyploidization has played an important role in plant evolution. However, upon polyploidization the process of meiosis must adapt to ensure the proper segregation of increased numbers of chromosomes to produce balanced gametes. It has been suggested that meiotic gene (MG) duplicates return to a single copy following whole genome duplication to stabilise the polyploid genome. Therefore, upon the polyploidization of wheat, a hexaploid species with three related (homeologous) genomes, the stabilization process may have involved rapid changes in content and expression of MGs on homeologous chromosomes (homeologs). To examine this hypothesis, sets of candidate MGs were identified in wheat using co-expression network analysis and orthology informed approaches. In total, 130 RNA-Seq samples from a range of tissues including wheat meiotic anthers were used to define co-expressed modules of genes. Three modules were significantly correlated with meiotic tissue samples but not with other tissue types. These modules were enriched for GO terms related to cell cycle, DNA replication and chromatin modification, and contained orthologs of known MGs. Overall 74.4 % of genes within these meiosis-related modules had three homeologous copies which was similar to other tissue-related modules. Amongst wheat MGs identified by orthology, rather than co-expression, the majority (93.7 %) were either retained in hexaploid wheat at the same number of

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

copies (78.4 %) or increased in copy number (15.3%) compared to ancestral wheat species. Furthermore, genes within meiosis-related modules showed more balanced expression levels between homeologs than genes in non-meiosis-related modules. Taken together our results do not support extensive gene loss nor changes in homeolog expression of MGs upon wheat polyploidization. The construction of the MG co-expression network allowed identification of hub genes and provided key targets for future studies.

Author summary

All flowering plants have undergone a polyploidization event(s) during their evolutionary history. One of the biggest challenges faced by a newly-formed polyploid is meiosis, an essential event for sexual reproduction and fertility. This process must adapt to discriminate between multiple related chromosomes and to ensure their proper segregation to produce fertile gametes. The meiotic mechanisms responsible for the stabilisation of the extant polyploids remain poorly understood except in wheat, where there is now a better understanding of these processes. It has been proposed that meiotic adaptation in established polyploids could involve meiotic gene loss following the event of polyploidization. To test this hypothesis in hexaploid wheat, we have computationally predicted sets of hexaploid wheat meiotic genes based on expression data from different tissue types, including meiotic anther tissue, and orthology informed approaches. We have calculated homeolog expression patterns and number of gene copies for the predicted meiotic genes and compared them with proper control gene sets. Our findings did not support any significant meiotic gene loss upon wheat polyploidization. Furthermore, wheat meiotic genes showed more balanced expression levels between homeologs than non-meiotic genes.

Introduction

Meiosis is a specialized mode of cell division which generates haploid gametes. Prior to meiosis, chromosomes are replicated. On entry into meiosis, homologous chromosomes (homologs) locate each other, and intimately align (synapse) along their length. Within this paired structure,

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

chromosomes recombine and crossover before being accurately segregated **[1-3]**. This complex and dynamic process is essential to maintain genome stability and integrity over sexual life cycles and to generate genome variation, which is a major evolutionary driving force **[4,5]**. The genetic variation created by meiotic recombination underpins plant breeding to improve crop species **[6-8].**

Polyploidization has played an important role in the evolution and speciation of flowering plants **[9,10]**, although the resultant multiplicity of related genomes poses a major challenge for the meiotic process. Segregation of the chromosomes to produce balanced gametes requires correct pairing, synapsis and recombination between only true homologs, rather than any of the other highly related chromosomes (homeologs) **[10-12]**.

In the last two decades, there have been significant advances in our understanding of plant meiosis. Since the isolation of the first meiosis-specific cDNA from lily in the mid-1990s **[13,14]**, more than 110 plant meiotic genes (MGs) have been identified, mainly from studies of the model diploid plants Arabidopsis and rice **[2,15,16]**. Although 25-30% of flowering plants are extant polyploids **[9]**, the meiotic mechanisms responsible for their stabilisation remain poorly understood. An exception is hexaploid wheat (*Triticum aestivum* L.), where there is now better understanding of these processes **[17]**. Despite possessing multiple related genomes, durum wheat, a tetraploid (AABB) and bread wheat, a hexaploid (AABBDD) behave as diploids during meiosis. Thus, most of the meiotic studies conducted in hexaploid wheat have focused on providing better understanding of the meiotic processes required to stabilise this polyploid species **[18-22]**. An emphasis has been to characterise the role of the *Ph1* locus in the suppression of recombination between homeologs **[23-31]**. Recent studies have defined this phenotype to a *ZIP4* gene which duplicated and diverged on polyploidization **[31-33]**. This event resulted in the suppression of homeologous crossover, and promotion of homologous synapsis.

Although all flowering plants have undergone at least one event of whole genome duplication during their evolutionary history **[34]**, it has been suggested that MG duplicates return to a single copy following whole genome duplication, more rapidly than the genome-wide average **[35]**. Therefore, it has been assumed that the stabilisation process upon the polyploidization of wheat also involved rapid

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

changes in the content and expression of the genes on homeologs. This process would facilitate the correct pairing and synapsis of homeologs during meiosis. The recent development of an expression atlas for hexaploid wheat revealed that 70% of homeologous genes in syntenic triads showed balanced expression **[36]**; however, this study did not include analysis of the genes expressed during meiosis.

Here, we assessed whether the level of expression of all genes in triads was balanced between homeologs during meiosis. Analysis indicated similar balanced expression to that observed in other wheat tissues. However, it could be argued that only meiotic specific genes might show differential expression between homeologs. Sets of candidate MGs were identified using co-expression network analysis and orthology informed approaches, allowing us to evaluate the effect of polyploidization on wheat MG copy number and expression. The combination of co-expression network analysis, in conjunction with orthologue information, will now contribute to the discovery of new MGs and greatly empower reverse genetics approaches to validate the function of candidate genes in wheat.

Results and Discussion

An initial assessment of the homeolog expression pattern in triads during meiosis in hexaploid wheat was undertaken. Relative expression abundance of 19,801 triads (59,403 genes) was calculated for 8 tissues, including meiotic anther tissue, according to published criteria **[36]**. This analysis revealed that the percentage of balanced triads was slightly higher in meiotic anther tissue (77.3%) than in other type of tissues (ranging from 67.3% in floral organs and 76.6% in leaves) (**S1 Fig**). The copy number of genes expressed during meiosis was also investigated. This involved the definition of 19801 triads (59403 genes), 7565 duplets (15130 genes), 15109 monads (single-copy genes) and 18250 genes from the "Others" group with various copy numbers, based on the *Ensembl*Plants database for the HC genes of hexaploid wheat **[37]** (IWGSC v1.1 gene annotation; **S1 Table**). Comparison of copy number of genes expressed in the 8 different tissues showed that 70.9% of the genes expressed during meiosis belonged to triads. This percentage ranged between 66.5% and 72.5% for the genes expressed in floral organs and stem tissues, respectively (**S2 Fig**). These results were

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

112 consistent with a previous study reporting significant balanced expression between homeologous
113 genes in tissues other than meiotic anthers **[36]**. However, the observations were not in agreement
114 with the hypothesis that stabilisation of polyploidization involved significant changes in gene content
115 and expression between homeologs **[35,38,39]**. Considering that not all genes expressed in meiotic
116 anther tissue are directly involved in the meiosis process, it is possible that meiotic specific genes
117 exhibit a different pattern. Therefore, a co-expression gene network was developed to compare the
118 expression pattern of homeologous genes in meiosis-related modules, which potentially represent
119 meiosis specific genes, and other tissue-related modules.
120
121 **Weighted co-expression network construction**
122 Network-based approaches have been proved useful in systems biology, to mine gene function from
123 high-throughput gene expression data. Gene co-expression analysis has become a powerful tool to
124 build transcriptional networks of genes involved in common biological events in plants **[40-45]**. The
125 use of co-expression networks has uncovered candidate genes to regulate biological processes in
126 many plants including wheat **[37]**, rice **[46,47]** and *Arabidopsis thaliana* **[48,49]**. The recently
127 published high-quality genome reference sequence **[37]** and a developmental gene expression atlas
128 **[50,36]**, together with the gene expression data collected from meiotic samples, were used to build a
129 co-expression gene network. One hundred and thirty samples from different tissue types were
130 included in this co-expression analysis (**S2 Table**; **Fig 1A**). A set of 60,379 genes out of the total
131 107,892 HC genes was considered expressed during meiosis (transcript per million (TPM) > 0.5 in at
132 least one meiosis sample) and used to run the co-expression analysis. Using the "WGCNA" package
133 in R **[51,52]**, genes with similar expression patterns were grouped into modules via the average
134 linkage hierarchical clustering of normalized count expression values (**Fig 1E**). The power of $\beta = 7$
135 (scale free topology $R^2 = 0.91$) was selected as the soft threshold power to emphasize strong
136 correlations between genes and penalize weak correlations to ensure a scale-free network (**Fig 1B-C**).
137 Based on this analysis, 50,387 out of 60,379 genes (83.5% of expressed genes) could be assigned to
138 66 modules. Module size ranged from 52 to 7541 genes (mean 763 genes; median 429 genes) (**Fig**
139 **1D**). Detailed information about gene number and gene IDs in each module is shown in **S3 Table**.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The expression patterns of all genes within a single module were summarized into a Module Eigengene (ME; representative gene of the module) to minimize data size for subsequent analyses. Expression patterns of modules are shown as a heatmap by plotting ME values in relation to tissue samples (**Fig 2**).

Fig 1. The weighted gene co-expression networks analysis (WGCNA). (A) Clustering dendrogram of 130 samples from different types of tissues. The sample clustering was based on the expression data of the genes expressed in at least one meiosis sample. **(B)** Determination of soft-thresholding power (β). The soft power threshold used in constructing the weighted gene co-expression networks was chosen as the first power to exceed a scale-free topology fit index of 0.9 (then $\beta = 7$). **(C)** Analysis of the mean connectivity for various soft-thresholding powers. **(D)** Number of genes in the modules with their frequency. **(E)** Blockwise dendrogram of the analysed genes (60,379 genes) clustered based on a dissimilarity measure of topological overlap matrices (TOM).

Fig 2. Heatmap plotting of MEs values in relationship to tissue samples. *n* indicates number of samples.

Identification of meiosis-related modules

A correlation analysis was conducted using the 66 MEs and the 8 different tissue types. A module was considered as meiosis-related when there was a strong correlation (*r*) with the 17 meiosis samples, and a weak or negative correlation with other tissue types. Accordingly, three meiosis-related modules were identified: module 2 (containing 4,940 genes); module 28 (544 genes); and module 41 (313 genes). Module 41 showed the strongest correlation with meiotic tissue ($r = 0.73$, FDR $=2.7 \times 10^{-20}$), compared to module 2 ($r = 0.61$, FDR $=9.2 \times 10^{-13}$) and module 28 ($r = 0.52$, FDR $=2.1 \times 10^{-8}$). (**Fig 3**).

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig 3. Co-expression network modules in relationship to tissues samples. Each row corresponds to a module; each column corresponds to a tissue type; Each cell contains the correlation value and, in parentheses, its corresponding FDR adjusted *P* value. *n* indicates number of samples. Only modules that have correlation value > 0.5 with meiotic anther tissue are shown.

Two other modules (modules 11 and 25) also showed significantly positive correlation with meiotic samples (r = 0.68 and 0.65, respectively), however they were not considered meiosis-related because they also correlated with samples from floral organs (at stages other than meiosis) and spike tissues, as shown in **Fig 3**. Therefore, our analysis focused on the three modules (2, 28 and 41), exhibiting a strong correlation with meiotic tissues and not with other floral organs, while modules 11 and 25 were considered as non-meiosis specific modules (referred to in this paper as non-meiotic modules). Other tissue-related modules (the top three correlated modules) were also identified to be used as controls for the meiosis-related modules in the subsequent analysis. These modules were: grain-related modules 5, 13 and 32 (r = 0.89, 0.89 and 0.85, respectively); leaves-related modules 1, 45 and 60 (r = 0.72, 0.68 and 0.71, respectively); and roots-related modules 7, 9 and 64 (r = 0.70, 0.76 and 0.85, respectively) **(S3 Fig)**.

Biological significance of expression similarity in modules

Several approaches were undertaken to validate the meiosis-related modules. The three modules (2, 28 and 41), strongly correlated with meiotic tissue expression, were found to be significantly enriched with the gene ontology (GO) slim terms "cell cycle", "DNA metabolic process", "nucleobase-containing compound metabolic process" and "nucleus" (**Fig 4**). Among the top five enriched GO slim terms in each of the 66 modules, the term "cell cycle" was significant only in the three meiosis-related modules, suggesting this was not a general property of all modules, and was instead specific to the meiosis modules. Module 2 in particular was significantly enriched with GO terms related to many biological processes occurring during meiosis such as "DNA replication", "histone methylation", "cytokinesis", "nucleosome assembly" and "chromatin silencing" (**Table 1; S4 Table**). The term "double-strand break repair via homologous recombination", an important process during

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

meiosis, was the primary enriched Biological Processes GO term in module 41 (FDR < 0.05). The biological processes mediated by genes in module 28 included "protein deneddylation", "positive regulation of G2/M transition of mitotic cell cycle", "COP9 signalosome" and other terms related to protein deneddylation and cell cycle control (**Table 1**). GO terms of meiosis-related modules were compared with those of modules highly correlated with other tissues. The GO terms "chloroplast", "plastids", "thylakoid", "photosystem" were significantly enriched in module 1, the most highly correlated module with leaves. The terms related to protein ubiquitination and protein binding were enriched in module 5 (the most highly correlated module with grain), while the terms "lignin biosynthetic process", "phenylpropanoid metabolic process" and "response to wounding" were enriched in module 7, the largest module correlated with roots (**S4 Fig**). This indicated that our co-expression module-tissue correlation was meaningful both from the biological and physiological point of view. Detailed information of the enriched GO and GO slim terms in all modules is listed in the supplementary table **S4 Table**. In summary, GO analysis confirmed that the three modules (2, 28 and 41) were enriched for genes associated with meiotic processes.

Fig 4. Enriched GO slim terms in the meiosis-related modules. Top 5 enriched GO terms in each module are shown. BP indicates Biological Processes and CC Cellular Component. No Molecular Function (MF) GO terms appear among the top 5 GO slim terms. Black bars indicate the number of genes in the GO term.

Table 1. Top 5 enriched GO terms in the meiosis-related modules for each ontology group.

	GO term (FDR adjusted *P* value)		
	Biological Process	**Molecular Function**	**Cellular Components**
Module 2	cell proliferation (2.2×10^{-222})	protein heterodimerization activity (1.5×10^{-93})	nucleosome (9.3×10^{-162})
	DNA replication (4.8×10^{-151})	DNA binding (1.8×10^{-69})	nuclear chromatin (2×10^{-100})

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	histone H3-K9 methylation (1.6×10^{-137})	microtubule binding (1.3×10^{-41})	chromosome (3.5×10^{-95})
	DNA-dependent DNA replication (5.4×10^{-135})	DNA-dependent ATPase activity (1.6×10^{-32})	pericentric heterochromatin (3.2×10^{-88})
	regulation of DNA replication (2.4×10^{-127})	motor activity (4.4×10^{-31})	heterochromatin (6.7×10^{-84})
Module 28	protein deneddylation (3.6×10^{-16})	apurinic or apyrimidinic site) endonuclease activity (3.3×10^{-05})	COP9 signalosome (3.3×10^{-21})
	positive regulation of G2/M transition of mitotic cell cycle (5.4×10^{-11})	RNA cap binding (1.8×10^{-03})	nucleus ($6.8M \times 10^{-10}$)
	COP9 signalosome assembly (1.2×10^{-7})	NADH activity (3.2×10^{-3})	nuclear cap binding complex (1.5×10^{-6})
	mitotic recombination (1.5×10^{-7})	enoyl-[acyl-carrier-protein] reductase activity (3.2×10^{-3})	protein-containing complex (1.9×10^{-6})
	photomorphogenesis (1.7×10^{-7})	signaling receptor activity (3.9×10^{-3})	cortical cytoskeleton (1.5×10^{-3})
Module 41	double-strand break repair via homologous recombination (3.9×10^{-5})	methyl-CpG binding (1×10^{-4})	nuclear euchromatin (7.3×10^{-5})
	somatic cell DNA recombination (1.1×10^{-4})	siRNA binding (1.6×10^{-3})	nucleus (1×10^{-4})
	megasporocyte differentiation (4.3×10^{-4})	SUMO transferase activity (2.1×10^{-3})	RNA polymerase IV complex (1.1×10^{-3})
	gene silencing by RNA (5.5×10^{-4})	DNA binding (5.7×10^{-3})	RNA polymerase II, core complex (4×10^{-3})
	positive regulation of sulfur metabolic process (1.1×10^{-3})	cytosine C-5 DNA demethylase activity (1×10^{-2})	proteasome regulatory particle, base subcomplex (4.4×10^{-3})

Enrichment of meiosis-related modules for wheat orthologs of known MGs.

An assessment was undertaken to confirm that the meiosis-related modules were enriched for wheat orthologs of known MGs. Although the first wheat meiotic cDNA clones were isolated concurrently

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

with the early discoveries of MGs in other plants **[53]**, the identification of MGs in wheat has been hampered by the large wheat genome size, its polyploid nature and the absence of a complete genome sequence. Thus, in comparison to model plants (Arabidopsis and rice), few MGs have been functionally characterized in wheat. Characterised wheat MGs include *TaASY1* **[54,55]**, *TaMSH7* **[56]**, *TaRAD51* **[57,58]**, *TaDMC1* **[57,58]**, *TaPSH1* **[59]**, *TaZIP4* **[31-33]** and RecQ-7 **[60]**. Given that, the assessment of whether the three modules contain known MGs was undertaken using orthology informed approaches. A set of 1063 candidate MGs in wheat was identified and categorized based on the method used to identify the genes: the "Orthologs" group contained 407 genes (**S5 Table; S6 Table)** that correspond to wheat orthologs of 103 functionally characterized MGs in model plant species, and the "Meiotic GO" group contained 927 wheat genes annotated with one or more meiotic GO terms (**S7 Table**). There were 271 genes overlapping between the two groups (**Fig 5A),** that were considered in the "Orthologs" group when undertaking gene enrichment analysis. The presence of each gene in the different modules was determined. A set of 848 genes was assigned to modules in the co-expression network (**Fig 5B**), including 340 genes in meiosis-related modules. Genes from both groups were significantly over-represented ($P < 0.05$) in four modules, including two meiosis-related modules (2 and 28). Module 2 in particular, was the most enriched for these genes, possessing more than one third of the total candidate MGs assigned to modules. Module 2 had 142 wheat orthologs of MGs and 155 genes with meiotic GO terms, compared to the expected number (based on module size) of 27 and 42, respectively. Module 41 (the third meiosis-related module) was enriched only with genes from the "Orthologs" group, having 15 orthologs of known MGs, whereas the expected number was 2 (**Fig 5C**). Consistent with this, genes from the "Orthologs" and/or "Meiotic GO" groups, were significantly under-represented in modules strongly correlated with other types of tissue (modules 1, 7 and 9), and in modules with negative or no correlation with meiotic tissue (modules 3, 8, 14, and 17) (**S8 Table**).

Fig 5. Enrichment of meiosis-related genes in the co-expression network modules. (**A**) Venn diagram of overall genes in the meiosis-related modules (modules 2, 28 and 41), wheat orthologs of MGs in model plant species and genes with meiotic GO terms. *n* indicates number of genes in each

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

group. Numbers in brackets refer to number of genes not included in the WGCNA analysis because they are not expressed in meiotic anther tissue. (**B**) Total number of genes assigned to modules from orthologs of MGs and genes with meiotic GO terms. (**C**) Gene enrichment in modules. Statistical significance of gene enrichment in a module is colour coded (Red indicates over-represented, blue under-represented and grey not significant; $P < 0.05$). Rhombus shape indicates the expected number of genes in module.

In this study three co-expression gene modules were identified that are strongly correlated to meiotic anther tissue and highly enriched with GO terms related to many processes occurring during meiosis, orthologs of known MGs and genes having meiosis-specific GO terms. Although 67 (65%), out of the 103 wheat orthologs, had at least one gene copy assigned to one of the three meiosis-related genes, there were 36 orthologs whose gene homeologs were assigned to other modules (**S9 Table**). Some of those genes have essential meiotic functions, like: *ASY1*, which encodes a protein essential for homologous chromosome synapsis **[54,61,62]**; and *DMC1*, a gene encoding a recombination protein that acts only in meiosis **[58,63]**. Others are known to have both meiotic and mitotic functions, like: *BRCA2*, a DNA repair gene required for double strand breaks repair by homologous recombination **[64]** and *SMC1* and *SMC3*, chromosome cohesion genes **[65]**, thus they are expressed in both reproductive and non-reproductive tissues. Assessment of these 36 orthologs showed that the expression patterns of their gene copies did not allow them to be clustered in any of the meiosis-related modules (or allocated to module 0 that is composed of genes not forming part of a co-expressed module), either because they were expressed in most samples from all types of tissues, or because they were expressed in a few samples of a specific tissue type (like meiotic anther tissue). The expression values (TPMs) of all gene copies of those 36 orthologs are summarized in **S10 Table**. The number of meiotic anther samples (17) used in the present study, might not be enough to identify all MGs being expressed in a specific meiotic stage. Such genes might be identified by WGCNA analysis when a larger number of meiotic samples becomes available. However, the analysis confirmed that the meiosis-related modules were indeed enriched for orthologs of known MGs, and for GO terms associated with processes involved in meiosis.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

276

277

278 **Copy number of MGs**

279 It has previously been suggested that MG duplicates return to a single copy following whole genome
280 duplication more rapidly than the genome-wide average in angiosperms **[35]**. The analysis of 19
281 meiotic recombination genes in hexaploid wheat and oilseed rape showed no evidence of gene loss
282 after polyploidization. However, a recent study in tetraploid oilseed rape showed that reducing the
283 copy number of *MSH4*, a key meiotic recombination gene involved in the ZMM pathway, prevents
284 meiotic crossovers between non-homologous chromosomes **[66]**. This led to the suggestion that
285 meiotic adaptation in polyploids could involve 'fine-tuning' the progression or the effectiveness of
286 meiotic recombination, which could be achieved through the loss of MG duplicates in the newly-
287 formed polyploids **[35,66]**. This hypothesis was evaluated in hexaploid wheat. The gene copy number
288 was assessed for the genes in the three meiosis-related modules and compared with genes in all
289 modules and in other tissue-related modules. Analysis showed that the percentage of genes belonging
290 to triads was 74.4% in the meiosis-related modules, which was similar to this percentage in other
291 tissue-related modules (72.5%, 74.0% and 76.1% in leaves-, grain- and roots-related modules,
292 respectively); however, it was significantly higher than those of the non-meiotic modules (57.4%).
293 The highest percentage of genes with three homeologs (83.3%) and lowest percentage of genes with
294 single copy (2.7%) were observed in the group of genes identified as MG orthologs and/or possessing
295 a meiotic GO term (**Fig 6A**).

296

297 **Fig 6. Copy number and homeolog expression pattern for genes from meiosis-related and other**
298 **tissue-related modules.** (**A**) Proportion of genes in each copy number category (triads, duplets,
299 monads and others) for different sets of expressed genes during meiosis including: "Meiotic modules"
300 refers to the three meiosis-related modules 2, 28 and 41; "Non-meiotic modules" refers to the modules
301 11 and 25 that showed high correlation with meiotic anther but were not considered meiosis-related
302 because they were also correlated with spike and floral organs tissues; the top 3 correlated modules
303 with each of leaves (modules 1, 45 and 60), grain (modules 5, 13 and 32) and roots (modules 7, 9 and

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

64) tissues; "All modules" contains all genes assigned to modules in the co-expression network and "Orthologs & Meiotic GO" refers to the set of genes that are orthologs of known MGs in other plant species and/or have meiotic GO terms. *n* number of genes in each set. (**B**) Proportion of genes from each homeolog expression pattern category (balanced, dominant and suppressed) calculated for triads in the previously mentioned sets of genes. *n* number of genes in each set. (**C**) Ternary plot showing relative expression abundance in meiotic anther tissue of 2,366 triads to which the genes of meiosis-related modules (2, 28 and 41) belong. Each circle represents a gene triad with an A, B, and D coordinate consisting of the relative contribution of each homeolog to the overall triad expression. Triads in vertices correspond to single-subgenome-dominant categories, whereas triads close to edges and between vertices correspond to suppressed categories. Box plots indicate the relative contribution of each subgenome based on triad assignment to the seven categories (Balanced, A dominant, B dominant, D dominant, A suppressed, B suppressed, D suppressed). Percentages between brackets indicate the percentage of triad number in each category to the total number of triads.

The high percentage of meiosis-related genes present as triads provides evidence that polyploid wheat did not experience significant gene loss (gene erosion) after polyploidization. However, this assumes that these genes were originally present as single copy genes in each of the A-, B- and D-genome progenitor species which gave rise to polyploid wheat. Therefore, the copy number of the 103 wheat MG orthologs in wheat progenitor species was investigated. All possible orthologs (high and low confidence predicted orthologs) were retrieved from *Ensembl* Plants Genes 43 database for *Triticum urartu* (ASM34745v1; **[67]**), the diploid progenitor of the wheat A-genome, the D-genome ancestor *Aegilops tauschii* (Aet_v4.0; **[68]**), the diploid progenitor of the wheat D-genome and *Triticum dicoccoides* (WEWSeq_v.1.0; **[69]**), the tetraploid progenitor of the hexaploid wheat (genome AABB). There was no change in copy number of 78.4% of genes, while 6.3% and 15.3% of genes had a lower and greater number of copies, respectively (**Table 2** and **S11 Table**). Regardless of genome of origin, the percentage of MGs with more copies was always greater than the percentage of genes with fewer copies. Comparing the A-genome MGs copy number in hexaploid wheat with the relevant

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

orthologs copy number in the corresponding A-genome ancestor, 86 genes (84.5%) had the same gene copy number in *T. dicoccoides* as in hexaploid wheat, while only 64 genes (63.1%) had the same gene copy number in *T. urartu*. This is consistent with the evolutionary history of hexaploid wheat, with *T. dicoccoides* being a more recent wheat progenitor (~10,000 years) than *T. urartu* (> 5 million years) **[70]**.

Table 2. Changes in copy number of wheat MGs in comparison with their orthologs in wheat progenitors. Comparison was done for two sets of genes: wheat orthologs of MGs (n = 103) and wheat meiotic recombination genes (n = 64).

	Number of meiotic genes (%)			Number of meiotic recombination genes (%)		
	Lower copy number	**Greater copy number**	**Equal copy number**	**Lower copy number**	**greater copy number**	**Equal copy number**
A genome (*T. urartu*)	9 (8.7%)	29 (28.2%)	64 (63.1%)	3 (4.7%)	21 (32.8%)	40 (62.5%)
A genome (*T. dicoccoides*)	5 (4.9%)	11 (10.7%)	86 (84.5%)	3 (4.7%)	9 (14.1%)	52 (81.3%)
B genome (*T. dicoccoides*)	6 (5.8%)	12 (11.7%)	84 (82.5%)	3 (4.7%)	6 (9.4%)	55 (85.9%)
D genome (*Ae. tauschii*)	6 (5.8%)	11 (10.7%)	85 (83.5%)	4 (6.3%)	5 (7.8%)	55 (85.9%)
Average per all genomes	7 (6.3%)	16 (15.3%)	81 (78.4%)	3 (5.1%)	10 (16.0%)	51 (78.9%)

An analysis on a subset of wheat genes, which were expected to be involved in meiotic recombination based on the function of their orthologs in model plants (64 genes; **S11 Table**), was conducted. Again, results showed that the majority (94.9%) of those genes had greater or no change in number of copies (**Table 2**). Given it has been suggested that the reduction in the copy number of ZMM pathway genes could stabilize meiosis in Brassica **[66],** the copy number of the wheat orthologs of seven ZMM genes

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

was evaluated. Five of the seven ZMM genes (*MER3*, *MSH5*, *ZIP4*, *PTD* and *SHOC1*) had equal or greater number of copies. However, *TaMSH4* gained one A-genome copy (comparing with *T. urartu*) and lost one D-genome copy (compared with *Ae. tauschii*), while *TaHEI10* lost A-genome copy and gained a D-genome copy (**S11 Table**). In conclusion, our findings did not support any significant gene loss upon the polyploidization of hexaploid wheat, as suggested for other polyploids **[35,66,71]**.

Homeolog expression patterns in triads of MGs

Initial analysis revealed that most genes expressed during meiosis showed balanced expression between homeologs (**S1 Fig**). The analysis was repeated using gene expression within the validated meiosis modules. Genes from all modules were assigned to three categories (balanced, dominant and suppressed) (see Materials and Methods; **S12 Table**). Homeolog expression patterns in triads showed that meiosis-related modules 2, 28 and 41 had the highest percentage (87.3%) of genes with balanced expression (belong to balanced triads), compared to the top three tissue-related modules for grain, leaves and roots (**Fig 6B**). Surprisingly, the group of candidate MGs selected for being orthologs of known MGs in other plant species and/or having meiotic GO terms had a higher percentage of genes from balanced triads (88.3%); whereas the modules not considered meiosis-specific (having high correlation with meiotic anther tissue and with spike and floral organs) contained only 68.3% genes with balanced expression (**Fig 6B**). The majority (84.19%) of triads with genes in meiosis-related modules (2,366 triads) showed balanced expression in meiotic anther tissue (**Fig 6C**).

In wheat, meiotic recombination and gene evolution rates are strongly affected by chromosome position, with relatively low recombination rates in the interstitial and proximal regions (genomic compartments R2a, C and R2b) but notably higher rates toward the distal ends of the chromosomes (genomic compartments R1 and R3) **[72,73]**. The lack of significant changes in gene content and more balanced expression between homeologs, suggested that these genes might be more prevalent in the proximal genomic compartments **[36,37]**. The distribution of MGs was therefore assessed across the genomic compartments compared with the distribution of all HC genes across chromosomes (**S13 Table**). Analysis showed that genes from the meiotic modules (modules 2, 28 and 41), were

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

significantly over-represented in the genomic compartments R2a, C and R2b (P = 2.4 x 10^{-5}, 3.1 x 10^{-6} and 1 x 10^{-5}, respectively), while they were under-represented in the R1 and R3 genomic compartments (P = 1.7 x 10^{-8} and 1.7 x 10^{-10}, respectively) (**Fig 7A**). Enrichment in the R2a genomic compartment region was not observed for genes from any of the other top three tissue-related modules (**Fig 7B**), since 21.7% of genes from the meiosis-related modules were assigned to R2a, while this percentage ranged between 18.2% and 19.5% in other tissue-related modules (**Table 3**). Interestingly, the set of the genes identified through orthology approaches had also similar high percentage (21%) of genes assigned to R2a genomic compartment (**Table 3**).

Fig 7. Enrichment of genes from different tissue-related modules in the wheat genomic compartments. **(A)** Number of genes (actual and expected) from the three meiosis-related modules in each genomic compartment. **(B)** Comparison of number of genes from different tissue-related modules in genomic compartments. Statistical significance of gene enrichment in modules is colour coded (Red indicates enriched, blue depleted and grey not significant; $P < 0.05$). Black dots indicate the expected number of genes in groups.

Table 3. Number of genes from different groups in the wheat genomic compartments.

Modules	R1		R2a		C		R2b		R3	
	No.	%	No.	%	No.	%	No.	%	No.	%
Meiotic modules	625	10.9%	1246	21.7%	531	9.3%	1823	31.8%	1515	26.4%
Non-meiotic modules	158	10.5%	292	19.5%	123	8.2%	469	31.3%	456	30.4%
Grain modules	338	10.2%	649	19.5%	287	8.6%	1144	34.4%	903	27.2%
Leaves modules	838	10.8%	1466	18.9%	853	11.0%	2557	33.0%	2037	26.3%
Roots modules	270	10.3%	479	18.2%	177	6.7%	951	36.1%	754	28.7%
All modules	5147	10.3%	9880	19.9%	4507	9.1%	16491	33.1%	13738	27.6%
Orthologs & Meiotic GO	78	6.5%	253	21.0%	164	13.6%	387	32.1%	324	26.9%

Our analysis reveals that homeologous MGs on homeologs mostly show balanced expression and lack a significant change in MG content following polyploidization. The majority of homeologous genes (not only MGs) on homeologs also show over 95% sequence identity to each other **[37,74]**. Given

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

these observations, such homeologs could synapse and recombine during meiosis. However, in allohexaploid wheat, homologs rather than homeologs synapse and recombine during meiosis ensuring the stability and fertility of this species, and the *Ph1* locus, in particular the *TaZIP4* gene copy inside this locus, has been identified as the main locus controlling this process. The wheat *ZIP4*, an ortholog of *ZIP4/Spo22* in *A. thaliana* and rice, is a member of the ZMM genes involved in the synaptonemal complex formation and class I crossovers pathway **[75,76].** Moreover, wheat lacking *Ph1* exhibits extensive genome rearrangements, including translocation, duplications, and deletions **[33]**. Thus, the evolution of *Ph1* during wheat polyploidization may explain why wheat has largely maintained a similar gene content and balanced expression of its homeologs. How meiosis has adapted to cope with allopolyploidy in other species is still to be resolved; however, it has been suggested that reduction in the copy number of meiotic genes may stabilize the meiotic process after polyploidization **[35,66].** The present study shows that this is not the case in wheat. It is possible that the presence of *Ph1* in wheat enabled the retention of multiple copies of meiotic genes as an alternative mechanism to ensure proper segregation of chromosomes during meiosis. In any case, the identification of the *TaZIP4* gene within the *Ph1* locus as the gene responsible for the *Ph1* effect on recombination and the observed effects of *Ph1* in wheat suggests that it may have more of a central role in meiosis than originally suspected from studies on model systems **[75,76]**. It has recently been suggested that *ZIP4* might acts as a scaffold protein facilitating physical interactions and assembly of different proteins complexes **[75,76,77]**. Therefore, our co-expression network was used to identify the wheat orthologs of known MGs connected with *TaZIP4*. The analysis indicates that the three *TaZIP4* homeologs on group 3 (TraesCS3A02G401700, TraesCS3B02G434600 and TraesCS3D02G396500) were clustered in module 2, the largest meiosis-related module, and strongly connected to many orthologs of MGs with various meiotic functions (**Fig 8**). However, the *TaZIP4* copy responsible for *Ph1* phenotype (TraesCS5B02G255100) did not cluster in the same module, reflecting its different expression profile from the other homeologs, being expressed in most tissues **[31-33]**. *TaZIP4* was connected to wheat orthologs of genes known to be involved in crossover formation such as *MSH2*, *SHOC1*, *FANCM*, *FLIP*, *EME1B*, *MUS81* **[2]**. This suggests that there may be an interplay between *TaZIP4* and genes from the anti-crossover pathway. The *TaZIP4* sub-network

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

supports a more central role of *ZIP4* in meiosis than originally suspected from studies on model species.

Fig 8. The wheat MG orthologs connected to *TaZIP4*. The alluvial diagram shows the connected genes to the *TaZIP4* homeologs *TaZIP4-A1*, *TaZIP4-B1* and *TaZIP4-D1*. Edge weight > 0.05 was used as threshold to visualise connected genes. Black bars indicate the number of homeologs for each connected gene.

Further characterisation of the wheat meiotic co-expression network

1) Identification of hub genes in the meiosis-related modules

Hub genes were identified within our meiosis-related modules by calculating the correlation between expression patterns of each gene and the module eigengene: the most highly correlated genes to the eigengene being the hub genes. The top 10 hub genes of each module with their functional annotation are shown in **Table 4**. The top 10 hub genes in module 2 were core histone genes, supporting the strong contribution of histones in this meiosis-related module. For further verification of histone involvement in module 2 and other modules in general, all wheat genes annotated as core histones or having GO terms related to histone modification, were retrieved for enrichment analysis. Analysis showed that the five types of histones (H1, H2A, H2B, H3 and H4) were enriched only in module 2 (P = 3.6 x 10^{-4}, 1.2 x 10^{-22}, 1.1 x 10^{-19}, 9.4 x 10^{-21}, and 3.3 x 10^{-26}, respectively), having 433 genes (85% of all core histone genes in all modules), compared to an expected number of genes of 39 (**Fig 9**). Similar results were obtained for histone modification genes. Module 2 was the most enriched module with this group of genes (P = 9.3 x 10^{-52}), containing 438 genes (30% of all histone modification genes in all modules). The histone modification genes were also enriched in 11 other modules, including the other meiosis-related modules (modules 28 and 41), however, with much lower numbers of enriched genes (**Fig 9**). Detailed information about genes included in this analysis is provided in **S14 Table**. The strong enrichment of histone modification genes in module 2 (the largest meiosis-related module) supports the important role of histone modifications in meiosis **[78-83]**.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig 9. Histone genes enrichment in the gene co-expression network modules. The analysis included all the genes annotated as core histones (H1, H2A, H2B, H3 and H4) in the wheat genome and the genes with GO terms related to histone modification. Statistical significance of gene enrichment in a module at $P < 0.05$ is colour coded (Red indicates enriched, blue depleted and grey not significant). Rhombus shape indicates the expected number of genes in module.

Table 4. The top 10 hub genes of each meiosis-related module with their functional annotation.

	Gene	Functional annotation	GO terms
Module 2	TraesCS1B02G192500	Histone H2B	nucleosome; DNA binding; protein heterodimerization activity
	TraesCS1D02G286600	Histone H4	nucleosome; DNA binding; nucleosome assembly; protein heterodimerization activity
	TraesCS3A02G534100	Histone H2A	nucleosome; DNA binding; protein heterodimerization activity
	TraesCS6A02G034300	Histone H2A	nucleosome; DNA binding; protein heterodimerization activity
	TraesCS6A02G034600	Histone H2A	nucleosome; DNA binding; protein heterodimerization activity
	TraesCS6B02G048300	Histone H2A	nucleosome; DNA binding; protein heterodimerization activity
	TraesCS6B02G049000	Histone H2A	nucleosome; DNA binding; protein heterodimerization activity
	TraesCS6D02G326800	Histone H2B	nucleosome; DNA binding; protein heterodimerization activity
	TraesCS7B02G408400	Histone H2A	nucleosome; DNA binding; nucleus; protein heterodimerization activity
	TraesCSU02G095700	Histone H3	nucleosome; DNA binding; protein heterodimerization activity
Module 28	TraesCS1A02G263400	Zinc finger CCCH domain protein	DNA binding; protein binding; zinc ion binding
	TraesCS2B02G281900	Receptor kinase	protein kinase activity; protein binding; ATP binding; protein phosphorylation

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	TraesCS2D02G000100	Histone deacetylase complex subunit SAP30	protein binding
	TraesCS2D02G263500	Receptor kinase	protein kinase activity; protein binding; ATP binding; protein phosphorylation
	TraesCS5B02G379400	Vacuolar protein sorting-associated 2-2-like protein	vacuolar transport
	TraesCS5D02G070800	Aminotransferase	catalytic activity; biosynthetic process; pyridoxal phosphate binding
	TraesCS5D02G386000	Vacuolar protein sorting-associated 2-2-like protein	vacuolar transport
	TraesCS6A02G253900	High mobility group protein	chromatin assembly or disassembly; chromatin binding; chromatin remodelling; DNA binding
	TraesCS6B02G271600	High mobility group protein	chromatin assembly or disassembly; chromatin binding; chromatin remodelling; DNA binding
	TraesCSU02G072600	Vacuolar protein sorting-associated 2-2-like protein	vacuolar transport
Module 41	TraesCS1D02G003800	Serine/threonine-protein kinase ATM	NA
	TraesCS1D02G072800	Chaperone protein dnaJ	cytoplasm ; protein folding; unfolded protein binding; heat shock protein binding; response to stress
	TraesCS2A02G120400	adenine nucleotide transporter 1	NA
	TraesCS2D02G401100	High mobility group family	Nucleus; DNA binding; transcription factor activity; sequence-specific DNA binding
	TraesCS3A02G162200	Zinc finger CCCH domain-containing protein 4	metal ion binding
	TraesCS3D02G127500	Ankyrin repeat protein-like	protein binding
	TraesCS4D02G284200	F-box family protein	protein binding
	TraesCS5A02G170400	F-box protein	protein binding
	TraesCS7A02G233900	Poor homologous synapsis 1 protein	Nucleus; synapsis; intracellular signal transduction; kinase activity
	TraesCS7D02G181800	Interleukin-6	NA

Hub genes such as "*Poor homologous synapsis 1"* (*PHS1*) were also identified with module 41, the module most highly correlated to meiotic samples. This gene has been previously reported to have a key role in homologous chromosome pairing, synapsis, DNA recombination and accurate chromosome segregation during meiosis in maize **[84]**, Arabidopsis **[85]** and wheat **[59]**. Other hub

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

genes identified in modules 28 and 41 encoded for the high mobility group proteins **[86-90]**, histone deacetylase **[91-94]**, and F-box proteins **[95]**.

2) Analysis of Transcription factors within the meiosis-related modules

Many transcription factors (TFs) have been reported as key regulators of meiosis from studies on animals **[96,97]**, yeast **[98-101]** and protozoa **[102]**. However, very little is known about the involvement of TFs in plant meiosis. The meiotic co-expression network was therefore exploited to identify potential meiosis-specific TFs. An assessment was undertaken of the enrichment of previously identified TF families in hexaploid wheat in the meiosis-related modules 2, 28 and 41. A total of 4889 high confidence genes belonging to 58 TF families were predicted from the annotation of the wheat genome sequence. Of these, 2439 TFs from 57 families could be assigned to the 66 modules in the gene co-expression network. Modules 2, 28 and 41 (meiosis-related modules) had 225, 25 and 17 TFs belonging to 31, 13 and 9 TF families, respectively (S15 Table). Compared to the expected number of TF families genes in each module, only 5 TF families were significantly enriched in module 2: Mitochondrial transcription termination factor (mTERF); Growth-Regulating Factor (GRF); abscisic acid-insensitive protein 3/Viviparous1 (ABI3/VP1); Forkhead-associated domain (FHA) and E2F/ Dimerization Partner (DP). On the other hand, 4 TF families were significantly depleted (**Fig 10**). The TF family NAC was the only TF family significantly enriched in module 41, containing 5 NAC genes (expected number 0.6; $P < 0.05$). Module 28 was not enriched with any TF family, although E2F/DP transcription factors were enriched in this module with borderline statistical significance ($P = 0.06$), with 4 genes in this module (the expected number was 0.2). Except in module 2, E2F/DP and FHA transcription factors families were not enriched in any other modules in the gene co-expression network (**S16 Table**). E2F/DP plays an important role in regulating gene expression necessary for passage through the cell cycle in mammals and plants **[103-105]**. Members of FHA contain the forkhead-associated domain, a phosphopeptide recognition domain found in many regulatory proteins. Genes belonging to the FHA group are reported to have roles in cell cycle regulation **[106-108]**, DNA repair **[109-112]** and meiotic recombination and chromosome segregation **[113-115]**. A previous meiotic transcriptome study identified up-regulation of TFs belonging to the

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

MADS-box, bHLH, bZIP, and NAC families in Arabidopsis and maize meiocytes at early meiosis **[116]**. Zinc finger-like TFs have also been suggested to be regulators of maize MG expression **[117]**. The present study indicates that TF families known to have roles in cell cycle and meiosis processes are over-represented in the meiosis-related modules (module 2 in particularly). Those TF families contain about 20 meiosis-specific candidate TF genes whose function can be validated using the available reverse genetics resources in polyploid wheat **[118]**.

Fig 10. Transcription factor families in the meiosis-related modules. Statistical significance of gene enrichment in the modules is colour coded (Red indicates over-represented, blue under-represented and grey not significant; $P < 0.05$). Rhombus shape indicates the expected number of genes in module.

3) Visualisation of networks and identification of candidate MGs

Having identified meiosis-related modules, the Networks within such modules can be visualised, highlighting genes for future studies. Edge files were created with gene annotation for the three meiosis-related modules 2, 28 and 41. Those files can be used to investigate the relation between orthologs of MGs within a module and ranked based on the strength of the connection (weight value). Another application of co-expression networks is the identification of previously uncharacterised genes regulating biological processes **[37,40,42-49]**. Cytoscape 3.7.1 software **[119]** was therefore used to visualise our network and to show connections between different orthologs of MGs in meiosis-related modules. Wheat MG orthologs in meiosis-related modules were used as "guide genes" to construct co-expression subnetworks containing only genes with direct connections to the guide genes. One such subnetwork is shown in **Fig 11**, where the following wheat orthologs of MGs in module 41 were selected and used to construct a meiotic subnetwork: *Poor Homologous Synapsis 1* (*TaPHS1;* **[59]**); *Argonaute* (AGO9/AGO104; **[120,121]** *Replication protein A2c* (*OsRPA2c*; **[122]**; *Meiotic nuclear division protein 1* (*AtMND1*; **[123,124]**); *MMS and UV Sensitive 81* (*AtMUS81*; **[125,126]**); and *Parting Dancers* (*AtPTD*; **[127,128]**) (guide genes; red circles in **Fig 11**). The

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

network complexity was reduced using an edge weight > 0.05. The visualized subnetwork contained 53 gene IDs including 9 guide gene copies. The gene TraesCS7A02G233900 (*TaPHS1*), a hub gene in module 41, was central in the network having the highest number of direct edges (41 direct edges; connected with 77.4% of the genes in the subnetwork). This subnetwork allowed identification of other genes with putative roles in meiosis (pink circles): a) RNA recognition motifs-containing gene (TraesCS5A02G319000) similar to Mei2, a master regulator of meiosis and required for premeiotic DNA synthesis as well as entry into meiosis in *Schizosaccharomyces pombe* **[129,130]**; b) the gene TraesCS4D02G050000 showed similarity to Male meiocyte death 1 (*MMD/DUET*), a PHD-finger protein plays role in chromatin structure and male meiotic progression in *A. thaliana* **[131,131]**; c) the gene TraesCS5D02G454900, a possible TF belonging to the FHA family known to have function in cell cycle regulation **[106-108]**, DNA repair **[109-112]** and meiotic recombination and chromosome segregation **[113-115]**. The meiotic subnetwork contained genes with similarity to cell cycle like F-box family proteins, high mobility family proteins, and chromatin remodelling genes. The subnetwork also contained a group of genes connected to most of our guide genes, which thus might be involved with them in similar biological processes. Examples of such genes are TraesCS3A02G101000, TraesCS1A02G292700 and TraesCS1D02G291100 which encode for zinc finger CCCH domain-containing proteins (**Fig 11**). Other meiotic subnetworks were also constructed using other guide genes from modules 2 and 28.

Fig 11. A meiotic co-expression subnetwork in hexaploid wheat. This subnetwork was constructed using 9 guide genes in module 41. Guide genes are wheat orthologs of MGs in other plant species (red circles); pink circles represent genes with putative meiosis function. Edge weight 0.05 was used as threshold to visualise genes in the subnetwork using Cytoscape 3.7.1 software.

4) The meiotic co-expression network is accessible in a larger biological contest

Our WGCNA co-expression network and GO enrichment data has been integrated with the wheat knowledge network **[132]** to make it publicly accessible in a larger biological context and to make it searchable through the KnetMiner web application (http://knetminer.rothamsted.ac.uk; **[133]**.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

KnetMiner can be searched with keywords (incl. module ID and GO terms) and wheat gene identifiers. The gene knowledge graphs generated contain many additional relation types such protein-protein interactions, homology and links to genome wide association studies and associated literature placing the co-expression networks generated here in a wider context.

Conclusion

In summary, the present study shows that most MGs in wheat are retained as three homeologous genes, which are expressed during meiosis at similar levels, suggesting that they have not undergone extensive gene loss nor sub/neo-functionalisation. Meiosis-related modules have been used to create networks and identify hub genes providing targets for future studies. The network containing the *ZIP4* gene, recently defined as *Ph1* **[31-33]** for example, highlights potential interacting partners. Finally, the networks highlight genes such as *ZIP4* and "*Poor homologous synapsis 1*", which may play a more central role in meiosis than previously thought. The co-expression network analysis combined with orthologue information will contribute to the discovery of new MGs and greatly empowers reverse genetics approaches to validate the function of candidate genes **[118]**. Ultimately this will lead to better understanding of the regulation of meiosis in wheat (and other polyploid plants) and subsequently improve wheat fertility.

Materials and Methods

RNA-Seq data collection

For co-expression network analysis we included 130 samples, containing 113 samples previously described in Ramírez-González et al. **[36]** and 17 samples from anthers during meiosis (9 samples from Martin *et al*. **[33]**), and 8 samples downloaded from (https://urgi.versailles.inra.fr/files/RNASeqWheat/Meiosis/). Samples were selected to represent all main tissue types: grain (n = 37 samples), leaves (n = 21 samples), roots (n = 20 samples), anther at

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

meiosis (n = 17 samples), spike (n = 12 samples), floral organs (anther, pistil and microspores) at stages other than meiosis (n = 10 samples), stem (n = 7 samples) and shoots (n = 6 samples). All samples were under nonstress conditions and mostly from the reference accession Chinese Spring. Detailed information about the used samples are listed in the supplementary materials (**S2 Table**).

Mapping of RNA-Seq reads to reference

Kallisto v0.42.3 **[134]** was used to map all RNA-Seq samples to the Chinese Spring transcriptome reference IWGSC RefSeq Annotation v1.1 **[37]**, following default parameters previously shown to result in accurate homeolog-specific read mapping in polyploid wheat **[36,50]**. Tximport v1.2.0 was then used to summarise expression levels from transcript to gene level **(S1 Text)**.

Co-expression network construction

The WGCNA package in R **[51,52]** was used to construct the scale-free co-expression network. Metadata for all samples were assigned with 8 tissue types (average 16.25; median 14.5 replicates per factor). Only high confidence (HC) genes **[37]** with expression > 0.5 TPM in at least one meiosis sample were retained for co-expression network construction using the R Package WGCNA (version 1.66) **[51,52]**. Using the varianceStabilizingTransformation() function from DESeq2 **[135]**, the count expression level of selected genes was normalised to eliminate differences in sequencing depth between different RNA-Seq studies **(S2 Text)**. To select a soft power threshold (β) for adjacency calculation (as $a_{ij} = |s_{ij}|^{\beta}$; where s_{ij} is the correlation between gene i and gene j), the Scale-free Topology Criterion was used **[136]**. Using the pickSoftThreshold() function to calculate β values, the soft power threshold emphasising strong correlations between genes and penalising weak correlations was selected as the first power to exceed a scale-free topology fit index of 0.9 **[36] (S3 Text)**. The correlation type used to calculate adjacency matrices was biweight midcorrelation (bicor). The adjacency matrices were transformed into a topological overlap matrix (TOM), measuring the network connectivity of a gene defined as the sum of its adjacency with all other genes for network generation. The blockwiseModules() function was used to calculate matrices and construct blockwise networks considering the following parameters: network type (networkType) = “signed hybrid”,

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

maximum block size (maxBlockSize) = 46,000 genes, soft power threshold (power) = 7, correlation type (corType) = "bicor" (biweight midcorrelation with maxPOutliers set to 0.05 to eliminate effects of outlier samples), topological overlap matrices type (TOMType) = "unsigned" with the mergeCutHeight = 0.15 and the minModuleSize = 30 to classify genes with similar expression profiles into gene modules using average linkage hierarchical clustering, according to the TOM-based dissimilarity measure with a minimum module size of 30 genes **(S4 Text)**. Module Eigengene (MEs), summarising the expression patterns of all genes within a given module into a single characteristic expression profile, were calculated as the first principal component in the Principal Component Analysis (PCA) using the moduleEigengenes() function **(S4 Text)**.

Identifying meiosis-related modules

The module eigengenes was used to test correlations between gene modules and traits (8 tissue types) using the cor() function. To assess the significance of correlations, Student asymptotic *P* values for correlations were calculated using the function corPvalueStudent(), and corrected for multiple testing by calculating FDR (false discovery rate) using a p.adjust() function following the Benjamini & Yekutieli **[137]** method. We considered a module meiosis-related when its correlation was strong with meiosis samples ($r > 0.5$ and FDR < 0.05) and weak ($r < 0.3$) or negative with other type of tissues **(S5 Text)**.

Analysis of GO term enrichment in modules

GO term enrichment was calculated using the "goseq" package **[138]**. Gene ontology (GO) annotations of IWGSC RefSeq v1.0 genes were retrieved from the file "FunctionalAnnotation.rds" in https://opendata.earlham.ac.uk/wheat/under_license/toronto/Ramirez-Gonzalez_etal_2018-06025-Transcriptome-Landscape/data/TablesForExploration/FunctionalAnnotation.rds **[36]** by filtering for ontology "IWGSC+Stress". GO data was then converted to IWGSC RefSeq Annotation v1.1 by replacing "01G" by "02G" in the IWGSC v1.0 gene IDs and retaining only genes > 99 % similar with > 90% coverage in the v1.0 and v1.1 annotation versions (as determined by blastn of the cDNAs) (called "all_go"). *P* values for GO term enrichment were calculated using the goseq() function (using

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

631 the following parameters: the pwf object was created using the nullp() function which calculated a
632 Probability Weighting Function for the genes v1.1 based on their length, the gene2cat = all_go, and
633 test.cats = "GO:BP", to specify the Biological Process GO term category to test for over
634 representation amongst the inquired genes) and corrected using the FDR method **[139]**. A GO term
635 was considered enriched in a module when FDR adjusted *P* value < 0.05 **(S6 Text)**. All figures shown
636 for enriched GO terms in the modules were produced using RAWGraphs software **[140]**.
637
638 **Orthologs of MGs in wheat**
639 A comprehensive literature search was performed for MGs in model plant species (mainly *A. thaliana*
640 and rice; **S17 Table**), identifying gene IDs based on the "Os-Nipponbare-Reference-IRGSP-1.0" for
641 rice (*Oryza sativa* Japonica Group) and "TAIR10" for *A. thaliana*. Wheat orthologs of MGs were then
642 retrieved from *Ensembl* Plants Genes 43 database through BioMart (in;
643 http://plants.ensembl.org/biomart) where orthologs calculated according to Vilella et al. **[141]** using
644 the following gene datasets: "Triticum aestivum genes (IWGSC)", "Oryza sativa Japonica Group
645 genes (IRGSP-1.0)" and "Arabidopsis thaliana genes (TAIR10)" for wheat **[37]**, rice **[142,143]** and
646 Arabidopsis **[144]**, respectively. For genes with no orthologs identified using this method, potential
647 wheat orthologs were identified by searching for amino acid sequence similarity using BLASTP **[145]**
648 in *Ensembl*Plants according to the following criteria: E-value < 1e-10; ID% > 25% with Arabidopsis
649 and > 70% with rice; then only wheat genes for which *Ensembl*Plants did not list any orthologs in rice
650 and Arabidopsis were considered as orthologs of MGs. Finally, 407 wheat gene IDs were identified as
651 orthologs of 103 plant MGs (listed in **S5 Table**). This group of genes was referred to in this study as
652 "Orthologs".
653
654 **Wheat genes with MG ontology (GO)**
655 A total number of 46,909 GO terms used by Ramírez-González et al. **[36]** to calculate GO term
656 accessions for wheat genes (IWGSC v1.0 gene annotation) were filtered for meiosis-related GO
657 terms, using 15 meiosis-specific keywords ("meiosis", "meiotic", "synapsis", "synaptonemal",
658 "prophase I", "metaphase I", "anaphase I", "telophase I", "leptotene", "zygotene", "pachytene",

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

659 "diplotene", "chiasma", "crossover" and "homologous chromosome segregation"). A total of 284
660 meiosis GO accessions were identified and used to retrieve 927 wheat genes with potential roles
661 during meiosis (**S7 Table**). All genes identified by gene orthologs and gene ontology methods were
662 then filtered to retain only genes had expression > 0.5 TPM in at least one meiosis sample. This group
663 of genes was referred to in this paper as "Meiotic GO". Enrichment analysis for the genes from
664 "Orthologs" and "Meiotic GO" groups in all module was conducted (**S7 Text**). The number of genes
665 from each group was assessed in all modules and compared with the expected number based on the
666 module size. There was a set of genes overlapping between "Orthologs" and "Meiotic GO" groups,
667 which was considered in the "Orthologs" group when undertaking gene enrichment analysis. Fisher's
668 exact test was used to calculate significant enrichment in the modules. Gene group considered over-
669 or under-represented in a module when $P < 0.5$.
670
671 **Identifying highly connected hub genes**
672 Hub genes within each module were identified using the WGCNA R package function signedKME()
673 to calculate the correlation between expression patterns of each gene and the module eigengene. Hub
674 genes were considered those more highly correlated to the eigengene **(S8 Text)**.
675
676 **Assessment of TF families in modules**
677 A total of 4889 wheat HC genes (IWGSC RefSeq Annotation v1.1; **[37]**) belonging to 58 TF families
678 were predicted from the annotation of the wheat genome sequence (https://github.com/Borrill-
679 Lab/WheatFlagLeafSenescence/blob/master/data/TFs_v1.1.csv). The number of TFs from each family
680 was assessed in all modules and compared with the expected number based on the module size.
681 Fisher's exact test was used to calculate significant enrichment of TFs in the modules. TF family
682 considered over- or under-represented in a module when $P < 0.5$ **(S9 Text)**.
683
684 **Defining gene categories based on number of homeologs**
685 A list of homeologs for all HC hexaploid wheat genes (IWGSC v1.1 gene annotation; **[37]**) was
686 retrieved from *Ensembl* Plants Genes 43 database through BioMart (in;

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

687 http://plants.ensembl.org/biomart). Based on number of homeologs from each of the A-, B- and D-
688 sub-genomes, genes were assigned to four groups: triads that refer to 1:1:1 triads (with a single copy
689 from each of the A-, B- and D-sub-genomes); duplets referring to 1:1:0, 1:0:1 and 0:1:1 duplets;
690 monads group containing genes with no homeologs (e.g. 0:0:1); and "others" containing genes with
691 more than two homeologs, in conjunction with genes from the homeologous groups 0:1:2, 0:2:1,
692 1:0:2, 2:0:1, 1:2:0, 2:1:0, 2:0:0, 0:2:0 and 0:0:2. Accordingly, 19801 triads (59403 genes), 7565
693 duplets (15130 genes), 15109 monads (single-copy genes) and 18250 genes from the "Others" group
694 were identified (**S1 Table**).
695
696 **Defining gene categories based on homeolog expression patterns in triads**
697 Homeolog expression pattern in triads was determined for each of the eight tissue types (**S10 Text**,
698 first part). For triads it was calculated according to Ramírez-González *et al*. **[36]** where a triad can be
699 described as balanced, A dominant, A suppressed, B dominant, B suppressed, D dominant or D
700 suppressed, based on the relative expression contribution of its A, B and D homeologs. Triads were
701 defined as expressed when one of its homeologs was expressed according to the criterion used in our
702 WGCNA analysis (**S18 Table**). This insured that all triads contain genes from modules were included
703 in the homeolog expression bias analysis (**S10 Text**, second part). Genes from a triad might not
704 belong to the same module due to dissimilarity of their expression patterns. Thus, to allow the
705 assessment of the expression pattern of genes in each module, each homeolog (A, B and D
706 homeologs) in a triad was assigned to one of the three categories "Balanced", "Dominant" and
707 "Suppressed" based on the homeolog origin (A, B and D sub-genome) and the triad description
708 (balanced, A dominant, A suppressed, B dominant, B suppressed, D dominant or D suppressed) as
709 shown in **S19 Table**. The values of the relative contributions of each homeolog per triad were used to
710 plot the ternary diagrams using the R package ggtern **[146]**.
711
712 **Co-expression gene network visualisation**
713 Cytoscape software (version 3.7.1; **[119]**) was used to visualise the network described in this study.
714 Firstly, the "exportNetworkToCytoscape" function was used to create edge files which could be used

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to visualize the network, then depending on network complexity, different weight value thresholds were used to filter genes to be visualised (**S11 Text**,). The term 'weight value' in the input files for Cytoscape refers to the connection strength between two nodes (genes) in terms of correlation value obtained from the topological overlap matrices (TOM). The co-expression network data has also been integrated with the wheat knowledge network **[132]** to make it publicly accessible through the KnetMiner web application (http://knetminer.rothamsted.ac.uk; **[133]**. The data was semantically modelled as nodes of type Gene, Co-Expression-Module, Co-Expression-Study, GOterm; connected by relations of type part-of and enriched. Each module was given a unique identifier composed of the module number and the prefix "AKA". KnetMiner can be searched with keywords (incl. module ID and GO terms) and wheat gene identifiers.

Data availability

All data files used in this manuscript are deposited in the Earlham Institute Open Data Platform (https://opendata.earlham.ac.uk/wheat/under_license/toronto/Martin_etal_2018_Alabdullah_etal_2019_wheat_meiosis_transcriptome_and_co-expression_network/). All R scripts are provided as .text files in the Supporting Information. The Gene network data will become accessible and searchable through the public plant website, KnetMiner (http://knetminer.rothamsted.ac.uk) with its next update this August.

Acknowledgements

This work was supported by the UK biotechnology and Biological Sciences Research Council (BBSRC) through a grant as part of Designing Future Wheat (DFW) Institute Strategic Programme (BB/P016855/1) and response mode grant (BB/R007233/1).

Author Contributions

Conceptualization: Abdul Kader Alabdullah, Philippa Borrill, Azahara C. Martin, Cristobal Uauy, Graham Moore.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

742 **Data curation:** Abdul Kader Alabdullah, Azahara C. Martin, Ricardo H. Ramirez-Gonzalez, Keywan
743 Hassani-Pak.
744 **Formal analysis:** Abdul Kader Alabdullah
745 **Funding acquisition:** Graham Moore, Peter Shaw.
746 **Methodology:** Abdul Kader Alabdullah, Philippa Borrill, Cristobal Uauy, Graham Moore.
747 **Software:** Abdul Kader Alabdullah, Philippa Borrill, Ricardo H. Ramirez-Gonzalez, Keywan
748 Hassani-Pak.
749 **Supervision:** Graham Moore, Peter Shaw.
750 **Writing – original draft:** Abdul Kader Alabdullah
751 **Writing – review & editing:** Abdul Kader Alabdullah, Philippa Borrill, Azahara C. Martin, Ricardo
752 H. Ramirez-Gonzalez, Keywan Hassani-Pak, Cristobal Uauy, Peter Shaw and Graham Moore.
753
754

755 ## References

756 1. Kleckner N. Meiosis: how could it work?. Proc Natl Acad Sci USA. 1996;93: 8167–8174.
757 2. Mercier R, Mézard C, Jenczewski E, Macaisne N, Grelon M. The Molecular Biology of
758 Meiosis in Plants. Annu Rev Plant Biol. 2015;66: 297-327. doi: 10.1146/annurev-arplant-
759 050213-035923.
760 3. Zickler D. and Kleckner N. Recombination, Pairing, and Synapsis of Homologs during
761 Meiosis. Cold Spring Harb Perspect Biol 2015;7:a016626. doi: 10.1101/cshperspect.a016626
762 4. Capilla L, Garcia Caldés M, Ruiz-Herrera A. Mammalian Meiotic Recombination: A Toolbox
763 for Genome Evolution. Cytogenet Genome Res. 2016;150: 1-16.
764 5. Melamed-Bessudo C, Shilo S, Levy AA. Meiotic recombination and genome evolution in
765 plants. Curr Opin Plant Biol. 2016;30: 82–87.
766 6. Wijnker E, de Jong H. Managing meiotic recombination in plant breeding. Trends Plant
767 Sci. 2008;12: 640-6. doi: 10.1016/j.tplants.2008.09.004.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

768 7. Choi K. Advances towards Controlling Meiotic Recombination for Plant Breeding. Mol
769 Cells. 2017;40(11): 814-822. doi: 10.14348/molcells.2017.0171.
770 8. Fernandes JB, Séguéla-Arnauda M, Larchevêquea C, Lloyda AH, Merciera R. Unleashing
771 meiotic crossovers in hybrid plants. PNAS. 2018;115: 10.2431-2436.
772 9. Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JS. Polyploidy and interspecific
773 hybridization: partners for adaptation, speciation and evolution in plants. Ann Bot. 2017;120:
774 183-194. doi:10.1093/aob/mcx079.
775 10. Comai L. The advantages and disadvantages of being polyploid. Nat Rev Genet. 2005;6: 836–
776 846. doi:10.1038/nrg1711.
777 11. Ramsey J, Schemske D. Neopolyploidy in flowering plants. Annu Rev Ecol Syst. 2002;33:
778 589-639.
779 12. Stenberg P, Saura A. Meiosis and its deviations in polyploid animals. Cytogenet Genome Res.
780 2013;140: 185–203. doi:10.1159/000351731.
781 13. Kobayashi T, Hotta Y. Tabata S. Isolation and characterization of a yeast gene that is
782 homologous with a meiosis-specific cDNA from a plant. Mol. Gen. Genet. 1993;237: 225-
783 232.
784 14. Kobayashi T, Kobayashi E, Sato S, Hotta Y, Miyajima N, Tanaka A, Tabata S.
785 Characterization of cDNAs induced in meiotic prophase in lily microsporocytes. DNA Res.
786 1994;1: 15-26.
787 15. Mercier R, Grelon M. Meiosis in plants: ten years of gene discovery. Cytogenet Genome Res.
788 2008;120(3-4): 281-90. doi: 10.1159/000121077.
789 16. Luo Q, Li Y, Shen Y, Cheng Z. Ten Years of Gene Discovery for Meiotic Event Control in
790 Rice. 2014;41(3): 125-137.
791 17. Blary A Jenczewski E. Manipulation of crossover frequency and distribution for plant
792 breeding. Theor Appl Genet. 2019;132(3): 575-592. doi: 10.1007/s00122-018-3240-1.
793 18. Riley R, Chapman V. Genetic control of the cytologically diploid behaviour of hexaploid
794 wheat. Nature. 1958;182: 713–715.
795 19. Riley R. The diploidisation of polyploid wheat. Ibid. 1960;15: 407-429.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

796 20. Holm PB. Chromosome pairing and chiasma formation in allohexaploid wheat, triticum
797 aestivum analyzed by spreading of meiotic nuclei. Carlsberg Res.Commun. 1986;51: 239-
798 294.
799 21. Martinez-Perez E, Shaw P, Moore G. Polyploidy induces centromere association. J Cell Biol
800 2000;148: 233-238.
801 22. Martinez-Perez E, Shaw P, Aragon-Alcaide L, Moore G. Chromosomes form into seven
802 groups in hexaploid and tetraploid wheat as a prelude to meiosis. Plant J. 2003;36: 21–29.
803 23. Riley R, Chapman V, Johnson R. The incorporation of alien disease resistance in wheat by
804 genetic interference with the regulation of meiotic chromosomal synapsis. Genet Res.
805 1968;12: 199-219.
806 24. Holm PB. Chromosome pairing and synaptonemal complex formation in hexaploid wheat,
807 monosomic for chromosome 5B. Carlsberg Res Commun. 1988;53: 57-89.
808 25. Holm PB, Wang X. The effect of chromosome 5B on synapsis and chiasma formation in
809 wheat, Triticum aestivum cv. Chinese Spring. Carlsberg Res Commun. 1988;53: 191-208.
810 26. Feldman M. Cytogenetic Activity and Mode of Action of the Pairing Homoeologous (Ph1)
811 Gene of Wheat. Crop Sci. 1993;33: 894-897.
812 27. Aragon-Alcaide L, Reader S, Miller T, Moore G. Centromeric behaviour in wheat with high
813 and low homoeologous chromosomal pairing. Chromosoma. 1997;106: 327-33.
814 28. Martinez-Perez E, Shaw P, Moore G. The Ph1 locus is needed to ensure specific somatic and
815 meiotic centromere pairing. Nature. 2001;411: 204-208.
816 29. Prieto P, Shaw P. and Moore G. Homologue recognition during meiosis is associated with a
817 change in chromatin conformation. Nature Cell Biol. 2004;6: 906-908.
818 30. Colas I, Shaw P, Prieto P, Wanous M, Spielmeyer W, Mago R, Moore G. Effective
819 chromosome pairing requires chromatin remodeling at the onset of meiosis. Proc Natl Acad
820 Sci USA. 2008;105(16): 6075-80. doi: 10.1073/pnas.0801521105.
821 31. Martín AC, Rey MD, Shaw P, Moore G. Dual effect of the wheat Ph1 locus on chromosome
822 synapsis and crossover. Chromosoma. 2017; doi:10.1007/s00412-017-0630-0.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

823 32. Rey MD, Martín AC, Smedley M, Hayta S, Harwood W, Shaw P, et al. Magnesium increases
824 homoeologous crossover frequency during meiosis in ZIP4 (Ph1 Gene) mutant wheat-wild
825 relative hybrids. Front Plant Sci. 2018;9: 509. doi:10.3389/fpls.2018.00509.
826 33. Martín AC, Borrill P, Higgins J, Alabdullah A, Ramírez-González, RH, Swarbreck D, et al.
827 Genome-Wide Transcription During Early Wheat Meiosis Is Independent of Synapsis, Ploidy
828 Level, and the Ph1 Locus. Front Plant Sci. 2018;9:1 791. doi: 10.3389/fpls.2018.01791.
829 34. Jiao Y, Wickett NJ, Ayyampalayam S., Chanderbali, A.S., Landherr, L., Ralph, E., et al.
830 Ancestral polyploidy in seed plants and angiosperms. Nature. 2011;473: 97–100.
831 doi:10.1038/nature09916.
832 35. Lloyd AH, Ranoux M, Vautrin S, Glover N, Fourment J, Charif D, et al. Meiotic gene
833 evolution: Can you teach a new dog new tricks? Mol Biol Evol. 2014;31: 1724-1727.
834 36. Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, et al. The
835 transcriptional landscape of polyploid wheat. Science. 2018;361: eaar6089. doi:
836 10.1126/science.aar6089.
837 37. International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in wheat
838 research and breeding using a fully annotated reference genome. Science. 2018;361:
839 eaar7191. doi: 10.1126/science.aar7191.
840 38. Edger PP, Smith R, McKain MR, Cooley AM, Vallejo-Marin M, Yuan Y, et al. Subgenome
841 dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally
842 established neo-allopolyploid monkeyflower. Plant Cell. 2017;29: 2150-2167.
843 39. Freeling M. Bias in plant gene content following different sorts of duplication: Tandem,
844 whole-genome, segmental, or by transposition. Annu. Rev. Plant Biol. 2009;60: 433-453.
845 40. Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, et al. Coexpression
846 tools for plant biology: opportunities for hypothesis generation and caveats. Plant Cell
847 Environ. 2009;32: 1633–1651.
848 41. He F, Maslov S. Pan- and core- network analysis of co-expression genes in a model plant. Sci
849 Rep. 2016;6: 38956. DOI: 10.1038/srep38956.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

42. Krishnan A, Gupta C, Ambavaram MMR, Pereira A. RECoN: Rice Environment Coexpression Network for Systems Level Analysis of Abiotic-Stress Response. Front. Plant Sci. 2017;8: 1640. doi: 10.3389/fpls.2017.01640.

43. Ma X, Zhao H, Xu W, You Q, Yan H, Gao Z, Su Z. Co-expression Gene Network Analysis and Functional Module Identification in Bamboo Growth and Development. Front. Genet. 2018;9: 574. doi: 10.3389/fgene.2018.00574.

44. Liu W, Lin L, Zhang Z, Liu S, Gao K, Lv Y, Tao H, He H. Gene co-expression network analysis identifies trait-related modules in Arabidopsis thaliana. Planta. 2019;249: 1487. https://doi.org/10.1007/s00425-019-03102-9.

45. Yu H, Jiao B, Lu L, Wang P, Chen S, Liang C, et al. NetMiner-an ensemble pipeline for building genome-wide and high-quality gene coexpression network using massive-scale RNA-seq. PLoS ONE. 2018;13(2): e0192613. https://doi.org/10.1371/journal.pone.0192613.

46. Tan M, Cheng D, Yang Y, Zhang G, Qin M, Chen J, et al. Co-expression network analysis of the transcriptomes of rice roots exposed to various cadmium stresses reveals universal cadmium-responsive genes. BMC Plant Biol. 2017;**17**: 194. https://doi.org/10.1186/s12870-017-1143-y

47. Aya K, Suzuki G, Suwabe K, Hobo T, Takahashi H, Shiono K, et al. Comprehensive Network Analysis of Anther-Expressed Genes in Rice by the Combination of 33 Laser Microdissection and 143 Spatiotemporal Microarrays. PLoS ONE. 2011;6(10): e26162. doi: 10.1371/journal.pone.0026162.

48. Silva AT, Ribone PA, Chan RL, Ligterink W, Hilhorst HW. A predictive co-expression network identifies novel genes controlling the seed-to-seedling phase transition in Arabidopsis thaliana. Plant Physiol. 2016;170: 2218–2231. doi: 10.1104/pp.15.01704.

49. Costa MC, Righetti K, Nijveen H, Yazdanpanah F, Ligterink W, Buitink J, et al. A gene co-expression network predicts functional genes controlling the re-establishment of desiccation tolerance in germinated Arabidopsis thaliana seeds. Planta. 2015;242: 435-449. doi: 10.1007/s00425-015-2283-7

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

50. Borrill P, Ramirez-Gonzalez R, Uauy C. expVIP: A customisable RNA-seq data analysis and visualization platform. Plant Physiol. 2016;170: 2172-2186. doi: 10.1104/ pp.15.01667.

51. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9: 559. doi: 10.1186/1471-2105-9-559.

52. Langfelder P, Horvath S. Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw. 2012;46: 11, 1-17. URL http://www.jstatsoft.org/v46/i11/.

53. Ji LH, Langridge P. An early meiosis cDNA clone from wheat. Mol. Gen. Genet. 1994;243: 17-23.

54. Boden SA, Shadiac N, Tucker EJ, Langridge P, Able JA. Expression and functional analysis of TaASY1 during meiosis of bread wheat (Triticum aestivum). BMC Mol. Biol. 2007;8: 65. http://www.biomedcentral.com/1471-2199/8/65.

55. Boden SA, Langridge P, Spangenberg G, Able JA. TaASY1 promotes homologous chromosome interactions and is affected by deletion of Ph1. Plant J. 2009;57: 487-497. doi: 10.1111/j.1365-313X.2008.03701.x

56. Lloyd AH, Milligan SA, Langridge P, Able JA. TaMSH7: A cereal mismatch repair gene that affects fertility in transgenic barley (Hordeum vulgare L.). BMC Plant Biol. 2007;7: 67. doi:10.1186/1471-2229-7-67

57. Khoo KHP, Jolly HR, Able JA. The RAD51 gene family in bread wheat is highly conserved across eukaryotes, with RAD51A1 upregulated during early meiosis. Funct Plant Biol 2008;35: 1267-1277.

58. Devisetty UK, Mayes K, Mayes S. The RAD51 and DMC1 homoeologous genes of bread wheat: cloning, molecular characterization and expression analysis. BMC Res Notes 2010;3: 245.

59. Khoo KHP, Able AJ, Able JA. Poor Homologous Synapsis 1 Interacts with Chromatin but Does Not Colocalise with ASYnapsis 1 during Early Meiosis in Bread Wheat. Int J Plant Genomics. 2012;23: 514398. DOI: 10.1155/2012/514398.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

903 60. Gardiner L-J, Wingen LU, Bailey P, Joynson R, Brabbs T, Wright J, et al. Analysis of the
904 recombination landscape of hexaploid bread wheat reveals genes controlling recombination
905 and gene conversion frequency. Genome Biol. 2019;20: 6. https://doi.org/10.1186/s13059-
906 019-1675-6
907 61. Caryl AP, Armstrong SJ, Jones GH, Franklin FC. A homologue of the yeast HOP1 gene is
908 inactivated in the Arabidopsis meiotic mutant asy1. Chromosoma. 2000;109: 62-71. doi:
909 10.1007/s004120050413.
910 62. Armstrong SJ, Caryl AP, Jones GH, Franklin FCH. Asy1, a protein required for meiotic
911 chromosome synapsis, localizes to axisassociated chromatin in Arabidopsis and Brassica. J.
912 Cell Sci. 2002;115: 3645-3655. doi: 10.1242/jcs.00048.
913 63. Klimyuk VI, Jones JDG. AtDMC1, the Arabidopsis homologue of the yeast DMC1 gene:
914 characterization, transposon-induced allelic variation and meiosis-associated expression. Plant
915 J. 1997;11: 1-14. doi: 10.1046/j.1365-313X.1997.11010001.x.
916 64. Trapp O, Seeliger K, Puchta H. Homologs of breast cancer genes in plants. Front. Plant Sci.
917 2011;2: 19. doi: 10.3389/ fpls.2011.00019.
918 65. Lam WS, Yang X, Makaroff CA. Characterization of Arabidopsis thaliana SMC1 and SMC3:
919 evidence that AtSMC3 may function beyond chromosome cohesion. J Cell Sci 2005;118:
920 3037-3048. doi:10.1242/jcs.02443.
921 66. Gonzalo A, Lucas M-O, Charpentier C, Sandmann G, Lloyd A, Jenczewski E. Reducing
922 MSH4 copy number prevents meiotic crossovers between non-homologous chromosomes in
923 Brassica napus. Nature commun. 2019;10:2354. https://doi.org/10.1038/s41467-019-10010-9.
924 67. Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, et al. Draft genome of the wheat A-
925 genome progenitor Triticum urartu. Nature. 2013; 496(7443): 87-90. doi:
926 10.1038/nature11997
927 68. Luo MC, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, et al. Genome sequence of the
928 progenitor of the wheat D genome Aegilops tauschii. Nature. 2017;551: 498-502. doi:
929 10.1038/nature24486.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

930 69. Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, et al. Wild emmer genome
931 architecture and diversity elucidate wheat evolution and domestication. Science.
932 2017;357(6346): 93-97. DOI: 10.1126/science.aan0032.
933 70. Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, The International Wheat Genome
934 Sequencing Consortium, et al. Ancient hybridizations among the ancestral genomes of bread
935 wheat. Science. 2014;345(6194): 1250092. DOI: 10.1126/science.1250092.
936 71. Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. Multiple rounds of speciation
937 associated with reciprocal gene loss in polyploid yeasts. Nature. 2006;440: 341-345.
938 72. Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, et al. Structural and functional
939 partitioning of bread wheat chromosome 3B. Science. 2014;345: 1249721. doi:
940 10.1126/science.1249721
941 73. Akhunov ED, Akhunova AR, Linkiewicz AM, Dubcovsky J, Hummel D, Lazo G, et
942 al. Synteny perturbations between wheat homoeologous chromosomes caused by locus
943 duplications and deletions correlate with recombination rates. Proc Natl Acad Sci
944 USA. 2003;100: 10836-10841. doi:10.1073/pnas.1934431100.
945 74. Schreiber AW, Hayden MJ, Kerrie L Forrest KL, Kong SL, Langridge P, Baumann U.
946 Transcriptome-scale homoeolog-specific transcript assemblies of bread wheat. BMC
947 Genomics. 2012;13: 492. doi: 10.1186/1471-2164-13-492.
948 75. Chelysheva L, Gendrot G, Vezon D, Doutriaux MP, Mercier R, Grelon M. Zip4/Spo22 is
949 required for class I CO formation but not for synapsis completion in Arabidopsis thaliana.
950 PLoS Genet. 2007;3(5): e83. doi:10.1371/journal.pgen.0030083.
951 76. Shen Y, Tang D, Wang K, Wang M, Huang J, Luo W, et al. ZIP4 in homologous
952 chromosome synapsis and crossover formation in rice meiosis. J of Cell Sci. 2012;125: 2581-
953 2591.
954 77. De Muyt A, Pyatnitskaya A, Andréani J, Ranjha L, Ramus C, Laureau R, et al. A meiotic
955 XPF-ERCC1-like complex recognizes joint molecule recombination intermediates to promote
956 crossover formation. Genes Dev. 2018; DOI: 10.1101/gad.308510.117

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

957 78. Wang L, Xu Z, Khawar MB, Chao Liu C, Li W. The histone codes for meiosis.
958 Reproduction. 2017;154: R65–R79.
959 79. Oliver C, Pradillo M, Corredor E, Cuñado N. The dynamics of histone H3
960 modifications is species-specific in plant meiosis. Planta. 2013;238(1): 23-33. doi:
961 10.1007/s00425-013-1885-1.
962 80. Maleki S & Keeney S 2004 Modifying histones and initiating meiotic recombination. Cell
963 118 404–406. (doi:10.1016/j.cell.2004.08.008)
964 81. Hu JL, Donahue G, Dorsey J, Govin J, Yuan ZF, Garcia BA, et al. H4K44 acetylation
965 facilitates chromatin accessibility during meiosis. Cell Reports. 2015;13: 1772–1780.
966 doi:10.1016/j. celrep.2015.10.070.
967 82. Luense LJ, Wang XS, Schon SB, Weller AH, Shiao EL, Bryant JM, et al. Comprehensive
968 analysis of histone post-translational modifications in mouse and human male germ cells.
969 Epigenetics and Chromatin. 2016;9. doi:10.1186/s13072-016- 0072-6.
970 83. Xu Z, Song Z, Li G, Tu H, Liu W, Liu Y, et al. H2B ubiquitination regulates meiotic
971 recombination by promoting chromatin relaxation. Nucleic Acids Res. 2016;44: 9681–9697.
972 doi:10.1093/nar/gkw652.
973 84. Pawlowski WP, Golubovskaya IN, Timofejeva L, Robert B, Meeley RB, William F, et al.
974 Coordination of Meiotic Recombination, Pairing, and Synapsis by PHS1. Science.
975 2004;303(5654): 89-92. DOI: 10.1126/science.1091110.
976 85. Ronceret A, Doutriaux MP, Golubovskaya IN, Pawlowski WP. PHS1 regulates meiotic
977 recombination and homologous chromosome pairing by controlling the transport of RAD50
978 to the nucleus. Proc Natl Acad Sci USA. 2009;106(47): 20121-20126.
979 86. Alonso-Martin, S, Rochat A, Mademtzoglou D, Morais J, de Reynies A, Aurade F, et al. Gene
980 expression profiling of muscle stem cells identifies novel regulators of postnatal myogenesis.
981 Front. Cell Dev. Biol. 2016;4: 58-58.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

87. Antosch M, Schubert V, Holzinger P, Houben A, Grasser KD. Mitotic lifecycle of chromosomal 3xHMG-box proteins and the role of their N-terminal domain in the association with rDNA loci and proteolysis. New Phytologist. 2015;208: 1067-1077.

88. Pedersen DS, Coppens F, Ma L, Antosch M, Marktl B, Merkle T, et al. The plant-specific family of DNA-binding proteins containing three HMG-box domains interacts with mitotic and meiotic chromosomes. New Phytologist. 2011;192: 577-589.

89. Di Agostino S, Fedele M, Chieffi P, Fusco A, Rossi P, Geremia R, Sette C. Phosphorylation of high-mobility group protein A2 by Nek2 kinase during the first meiotic division in mouse spermatocytes. Mol Biol Cell. 2004;15: 1224-1232.

90. Tessari MA, Gostissa M, Altamura S, Sgarra R, Rustighi A, Salvagno C, et al. Transcriptional activation of the cyclin A gene by the architectural transcription factor HMGA2. Mol Cell Biol. 2003;23: 9104-9116.

91. Akiyama T, Kim JM, Nagata M, Aoki F. Regulation of histone acetylation during meiotic maturation in mouse oocytes. Mol Reprod Dev. 2004;69: 222-227.

92. Wang Q, Yin S, Ai JS, Liang CG, Hou Y, Chen DY, et al. Histone deacetylation is required for orderly meiosis. Cell Cycle. 2006;5: 766-74.

93. Magnaghi-Jaulin L, Jaulin C. Histone deacetylase activity is necessary for chromosome condensation during meiotic maturation in Xenopus laevis. Chromosome Res. 2006;14: 319-332. DOI: 10.1007/s10577-006-1049-2.

94. Getun IV, Wu Z, Fallahi M, Ouizem S, Liu Q, Li W, et al. Functional roles of acetylated histone marks at mouse meiotic recombination hot spots. Mol Cell Biol. 2017;37: e00942-15. https:// doi.org/10.1128/MCB.00942-15.

95. Zheng N, Wang Z, Wei W. Ubiquitination-mediated degradation of cell cycle-related proteins by F-box proteins. Int J Biochem Cell Biol. 2016;73: 99-110.

96. Bolcun-Filas E, Bannister LA, Barash A, Schimenti KJ, Suzanne A. Hartford SA, et al. A-MYB (MYBL1) transcription factor is a master regulator of male meiosis. Development. 2011;138: 3319-3330. doi: 10.1242/dev.067645.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

97. Yan Z, Fan D, Meng Q, Yang J, Zhao W, Guo F, et al. Transcription factor ZFP38 is essential for meiosis prophase I in male mice. Reproduction. 2016;152(5): 431-7. doi: 10.1530/REP-16-0225.

98. Xu L, Ajimura M, Padmore R, Klein C, Kleckner N. NDT80, a meiosis-specific gene required for exit from pachytene in Saccharomyces cerevisiae. Mol Cell Biol. 1995;15: 6572-6581.

99. Horie S, Watanabe Y, Tanaka K, Nishiwaki S, Fujioka H, Abe H, et al. The Schizosaccharomyces pombe mei4 + Gene Encodes a Meiosis-Specific Transcription Factor Containing a forkhead DNA-Binding Domain. Mol Cell Biol. 1998;18(4): 2118-29.

100. Pierce M, Benjamin KR, Montano SP, Georgiadis MM, Winter E, Vershon AK. Sum1 and Ndt80 proteins compete for binding to middle sporulation element sequences that control meiotic gene expression. Mol Cell Biol. 2003;23: 4814-4825.

101. Beaudoin J, Ioannoni R, Normant V, Labbé S. A role for the transcription factor Mca1 in activating the meiosis-specific copper transporter Mfc1. PLoS ONE. 2018;13(8): e0201861. https://doi.org/10.1371/journal.pone.0201861

102. Zhang J, Yan G, Tian M, Ma Y, Xiong J, Miao W. A DP-like transcription factor protein interacts with E2fl1 to regulate meiosis in Tetrahymena thermophila. Cell Cycle. 2018;17(5): 634-642. doi: 10.1080/15384101.2018.1431595.

103. Zwicker J, Liu N, Engeland K, Lucibello FC, Müller R. Cell cycle regulation of E2F site occupation in vivo. Science. 1996;271(5255): 1595-7.

104. Zheng N, Fraenkel E, Pabo CO, Pavletich NP. Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. Genes Dev. 1999;13(6): 666-74.

105. Sozzani R, Maggio C. Varotto S, Canova S, Bergounioux C, Albani D, Cella F. Interplay between Arabidopsis Activating Factors E2Fb and E2Fa in Cell Cycle Progression and Development. Plant Physiol. 2006;140(4): 1355-1366.

106. Hollenhorst PC, Bose ME, Mielke MR, Müller U, Fox CA. Forkhead Genes in Transcriptional Silencing, Cell Morphology and the Cell Cycle: Overlapping and Distinct

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Functions for FKH1 and FKH2 in Saccharomyces cerevisiae. Genetics. 2000;154(4): 1533-1548

107. Zhu G, Spellman PT, Volpe T, Brown PO, Botstein D, Davis TN, Futcher B. Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. Nature. 2000;406(6791): 90-4.

108. Kim M, Ahn JW, Song K, Paek KH, Pai HS. Forkhead-associated Domains of the Tobacco NtFHA1 Transcription Activator and the Yeast Fhl1 Forkhead Transcription Factor Are Functionally Conserved. J Biol Chem. 2002;277: 38781-38790.

109. Sun Z, Hsiao J, Fay DS, Stern DF. Rad53 FHA domain associated with phosphorylated Rad9 in the DNA damage checkpoint. Science. 1998;281: 272-274.

110. Bashkirov VI, Bashkirova EV, Haghnazari E, Heyer WD. Direct kinase-to-kinase signaling mediated by the FHA phosphoprotein recognition domain of the Dun1 DNA damage checkpoint kinase. Mol Cell Biol. 2003;23: 1441-1452.

111. Palmbos PL, Wu D, Daley JM, Wilson TE. Recruitment of Saccharomyces cerevisiae Dnl4-Lif1 complex to a double-strand break requires interactions with Yku80 and the Xrs2 FHA domain. Genetics. 2008;180: 1809-1819.

112. Liang J, Suhandynata RT, Zhou H. Phosphorylation of Sae2 Mediates Forkhead-associated (FHA) Domain-specific Interaction and Regulates Its DNA Repair Function. J Biol Chem. 2015;290: 10751-10763.

113. Pérez-Hidalgo L, Moreno S, San-Segundo PA. Regulation of meiotic progression by the meiosis-specific checkpoint kinase Mek1 in fission yeast. J of Cell Sci. 2003;116: 259-271.

114. Cigliano RA, Sanseverino W, Cremona G, Consiglio FM, Conicella C. Evolution of parallel spindles like genes in plants and highlight of unique domain architecture. BMC Evol. Biol. 2011;11: 78. doi: 10.1186/1471-2148-11-78.

115. Crown KN, Savytskyy OP, Malik S, Logsdon J, Williams RS, Tainer JA, Zolan MEA. Mutation in the FHA Domain of Coprinus cinereus Nbs1 Leads to Spo11-Independent

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Meiotic Recombination and Chromosome Segregation. G3 (Bethesda). 2013;3(11): 1927-1943.

116. Dukowic-Schulze S, Harris A, Li J, Sundararajan A, Mudge J, Retzel EF, et al. Comparative transcriptomics of early meiosis in *Arabidopsis* and maize. J. Genet. Genomics. 2014;41: 139–152 10.1016/j.jgg.2013.11.007

117. Ma J, Skibbe DS, Fernandes J, Walbot V. Male reproductive development: gene expression profiling of maize anther and pollen ontogeny. Genome Biol. 2008;9: R18110.1186/gb-2008-9-12-r181

118. Krasileva KV, Vasquez-Gross HA, Howell T, Bailey P, Paraiso F, Clissold L, et al. Uncovering hidden variation in polyploid wheat. PNAS, 2017;114: 6. E913-E921.

119. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11): 2498-504.

120. Singh M, Goel S, Meeley RB, Dantec C, Parrinello H, Michaud C, et al. Production of viable gametes without meiosis in maize deficient for an argonaute protein. Plant Cell. 2011;23(2): 443-58.

121. Durán-Figueroa N, Vielle-Calzada JP. Argonaute9-dependent silencing of transposable elements in pericentromeric regions of arabidopsis. Plant Signal. Behav. 2010;5(11): 1476-79.

122. Li X, Chang Y, Xin X, Zhu C, Li X, Higgins JD, Wu C. Replication protein a2c coupled with replication protein a1c regulates crossover formation during meiosis in rice. Plant Cell. 2013;25(10): 3885-99.

123. Domenichini S, Raynaud C, Ni DA, Henry Y, Bergounioux C. Atmnd1-delta1 is sensitive to gamma-irradiation and defective in meiotic dna repair. DNA Repair (Amst). 2006;5(4): 455-64.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1088 124. Kerzendorfer C, Vignard J, Pedrosa-Harand A, Siwiec T, Akimcheva S, Jolivet S, et
1089 al. The arabidopsis thaliana mnd1 homologue plays a key role in meiotic homologous pairing,
1090 synapsis and recombination. J. Cell Sci. 2006;119(Pt 12): 2486-96.
1091 125. Hartung F, Suer S, Bergmann T, Puchta H. The role of atmus81 in dna repair and its
1092 genetic interaction with the helicase atrecq4a. Nucleic Acids Res. 2006;34(16): 4438-48.
1093 126. Higgins JD, Buckling EF, Franklin FCH, Jones GH. Expression and functional
1094 analysis of atmus81 in arabidopsis meiosis reveals a role in the second pathway of crossing-
1095 over. Plant J. 2008;54(1): 152-62.
1096 127. Lu P, Wijeratne AJ, Wang Z, Copenhaver GP, Ma H. Arabidopsis ptd is required for
1097 type i crossover formation and affects recombination frequency in two different chromosomal
1098 regions. J. Genet. Genomics. 2014;41(3): 165-75.
1099 128. Wijeratne AJ, Chen C, Zhang W, Timofejeva L, Ma H. The arabidopsis thaliana
1100 parting dancers gene encoding a novel protein is required for normal meiotic homologous
1101 recombination. Mol. Biol. Cell. 2006;17(3): 1331-43.
1102 129. Watanabe Y, Yamamoto M. S. pombe mei2+ encodes an RNA-binding protein
1103 essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA
1104 species meiRNA. Cell. 1994;78(3): 487-98.
1105 130. Watanabe Y, Shinozaki-Yabana S, Chikashige Y, Hiraoka Y, Yamamoto M.
1106 Phosphorylation of RNA-binding protein controls cell cycle switch from mitotic to meiotic in
1107 fission yeast. Nature. 1997;386: 187-190.
1108 131. Reddy TV, Kaur J, Agashe B, Sundaresan V, Siddiqi I. The duet gene is necessary for
1109 chromosome organization and progression during male meiosis in arabidopsis and encodes a
1110 phd finger protein. Development. 2003;130(24): 5975-87.
1111 132. Hassani-Pak K, Castellote M, Esch M, Hindle M, Lysenko A, Taubert J, Rawlings C.
1112 Developing integrated crop knowledge networks to advance candidate gene discovery.
1113 Applied & Translational Genomics. 2016;11: 18-26.
1114 https://doi.org/10.1016/j.atg.2016.10.003.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

133. Hassani-Pak K, KnetMiner - An integrated data platform for gene mining and biological knowledge discovery [dissertation]. Bielefeld: Universität Bielefeld; 2017.

134. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 2016;34: 525-527. doi: 10.1038/nbt.3519.

135. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550. doi: 10.1186/s13059-014-0550-8

136. Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis. Stat Appl Genet Mol Biol. 2005;4: Article 17. doi: 10.2202/1544-6115.1128.

137. Benjamini Y, Yekutieli D. The Control of the False Discovery Rate in Multiple Testing under Dependency. Ann. Stat. 2001;29: 1165-1188.

138. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11: R14.

139. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological), 1995:57(1): 289-300.

140. Mauri M, Elli T, Caviglia G, Uboldi G, Azzi M. RAWGraphs: A Visualisation Platform to Create Open Outputs. CHItaly '17. 2017; https://doi.org/10.1145/3125571.3125585. [Software] Available at: http://app.rawgraphs.io/.

141. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. EnsemblComparaGeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Research. Cold Spring Harbor Laboratory Press. 2009;19: 327-335.

142. .Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, et al. Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. Plant Cell Physiol. 2013;54(2): e6. doi: 10.1093/pcp/pcs183.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

143. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, et al. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. Rice (N Y). 2013;6(1): 4. doi: 10.1186/1939-8433-6-4.

144. Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, et al. ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. Plant Cell. 2017;29: 1806-1821. DOI: https://doi.org/10.1105/tpc.17.00073.

145. Korf I, Yandell M, Bedell J. BLAST: An Essential Guide to the Basic Local Alignment Search Tool. O'Reilly. 2003;360. ISBN:0-596-00299-8.

146. Hamilton N. ggtern: An extension to 'ggplot2', for the creation of ternary diagrams. R package version 2.2.1 [software]. 2016. Available from: https://CRAN.R-project.org/package=ggtern.

Supporting information

S1 Fig. Homeolog expression patterns of expressed triads in hexaploid wheat. Homeolog expression pattern was calculated for 19,801 triads (59,403 genes) across 8 tissue types according to published criteria **[36]**, where triad defined as expressed when the sum of the A, B, and D subgenome homoeologs was > 0.5 TPM. **(A)** Proportion of triads in each homoeolog expression pattern across the 8 tissues. *n* is number of expressed triads. **(B)** Ternary plot showing relative expression abundance of 14,837 expressed triads (44,511 genes) in the meiotic anther tissue. Each circle represents a gene triad with an A, B, and D coordinate consisting of the relative contribution of each homeolog to the overall triad expression. Triads in vertices correspond to single-subgenome–dominant categories, whereas triads close to edges and between vertices correspond to suppressed categories. Box plots indicate the relative contribution of each subgenome based on triad assignment to the seven categories. Percentages between brackets indicate the percentage of triad number in each category to the total number of triads.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

S2 Fig. Proportion of genes in each homeologs number category. Expressed genes across 8 tissues were assigned to four categories (triads, duplets, monads and others). *n* indicates number of expressed genes.

S3 Fig. Module-tissue relationship. Each row corresponds to a module; each column corresponds to a tissue type; Each cell contains the correlation value (*r*) and, in brackets, its corresponding FDR adjusted *P* value. *n* indicates number of samples. Only modules that have correlation value > 0.5 are shown.

S4 Fig. Enriched GO terms in the meiosis-related and other tissue-related modules. Top 5 GO terms are shown for each module. Black bars indicate the number of genes in the GO term.

S1 Table. Assignment of hexaploid wheat HC genes to four groups based on their copy number. *n* is number of samples per tissue type.

S2 Table. The 130 RNA-Seq samples included in the co-expression analysis. The samples belong to eight different types of tissues (Intermed.tissue).

S3 Table. Number of genes and gene IDs in the modules identified using WGCNA.

S4 Table. Enriched GO terms in all modules including the three meiosis-related modules.

S5 Table. Gene IDs of wheat orthologs of known meiotic genes in other plants species.

S6 Table. Meiotic genes and their orthologs in wheat according to EnsemblPlants prediction.

S7 Table. List of wheat genes with meiotic GO term(s)

S8 Table. Fishers exact test and FDR adjusted p-value for "Orthologs" and "Meiotic GO" genes enrichment in modules.

S9 Table. Assignment of wheat orthologs of known meiotic genes to the meiosis-related modules.

S10 Table. Expression pattern of wheat orthologs of meiotic genes that were not assigned to the meiosis-related modules. Average expression values (TPMs) is shown for the 8 tissue types.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

S11 Table. Copy number of orthologs of meiotic gene in wheat and its ancestors. MR indicates wheat orthologs of known meiotic recombination genes. ZMM indicates orthologs of ZMM pathway genes.

S12 Table. Assignment of the genes in triads to the co-expression network modules. Genes belong to Meiotic, Non-meiotic, Leaves, Grain and Roots modules and "Orthologs & Meiotic GO" gene sets are indicated in the table. Balance as a gene category refer to the genes belong to balanced triads.

S13 Table. HC genes distribution across wheat genomic compartment.

S14 Table. List of the core histone and histone modification genes identified in the wheat genome. All wheat genes annotated as core histones (Fun.annotation) or having GO terms related to histone modification (GO) were selected.

S15 Table. List of transcription factors in the meiosis-related modules 2, 28 and 4.

S16 Table. Fishers exact test for TF families enrichment in modules.

S17 Table. Literature review of the functionally characterized meiotic genes in model plant species.

S18 Table. Defining the expressed triads in meiotic anther tissue. Triad is considered as expressed when one of its homeologs is expressed according to the criterion used in our WGCNA analysis.

S19 Table. Gene category assignment based on the sub-genome of origin and triad description.

S1 Text. R script used to summarize expression values (counts) from transcript to gene level.

S2 Text. R script used to combine samples from all studies and normalize counts for WGCNA analysis.

S3 Text. R script used to calculate the soft power threshold for the WGCNA analysis.

S4. Text. R script used to run the WGCNA analysis and MEs calculation and plotting.

bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

S5. Text. R scripts used to calculate module-tissue correlation.

S6. Text. R scripts used for enrichment analysis of GO and GO slim terms in the modules.

S7. Text. R scripts used for enrichment analysis of "Orthologs" and "Meiotic GO" genes in the modules.

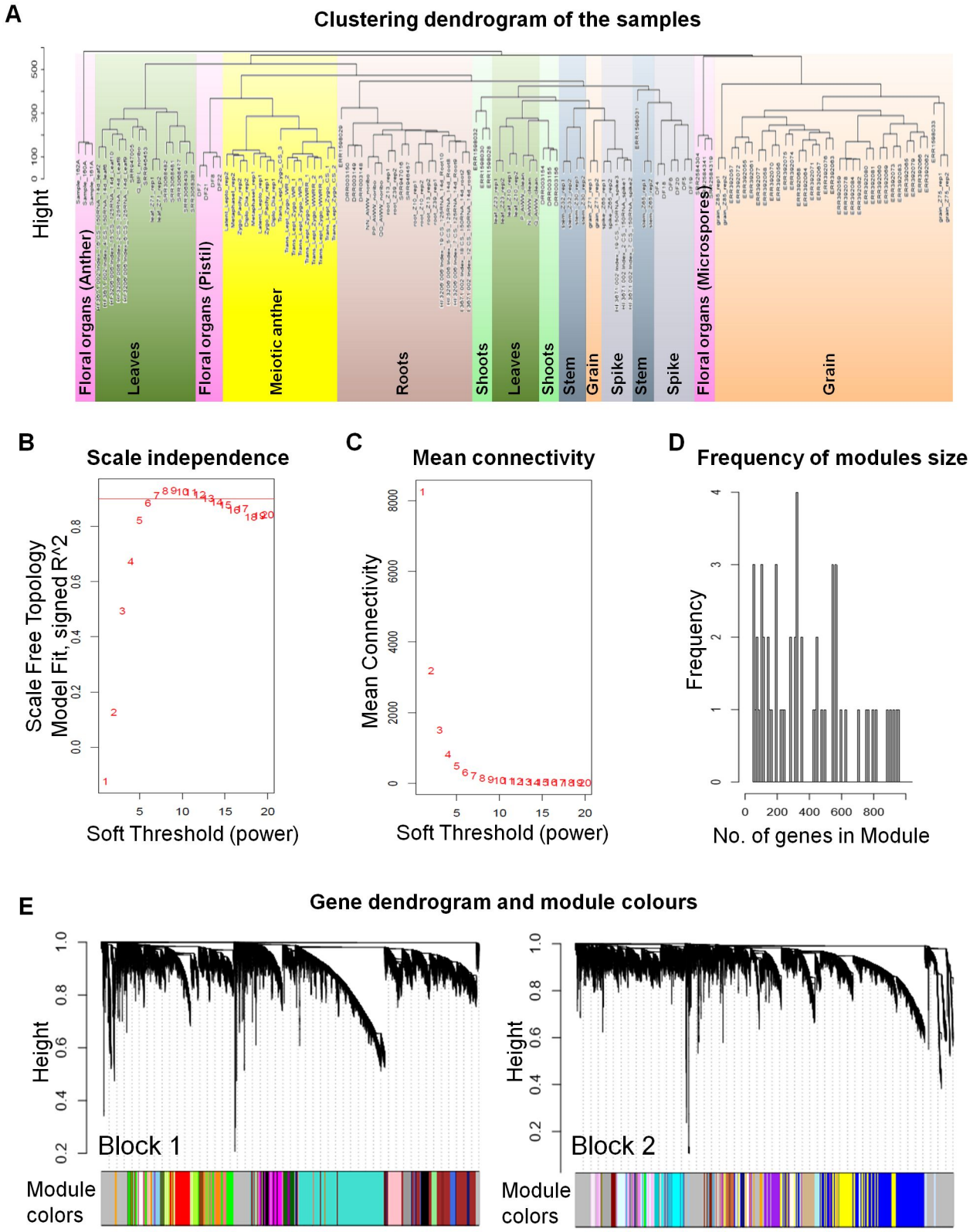
S8 Text: R script used to calculate hub genes for each module. Top 10 hub genes were selected for each module.

S9. Text. R script used for the assessment of TF families in modules.

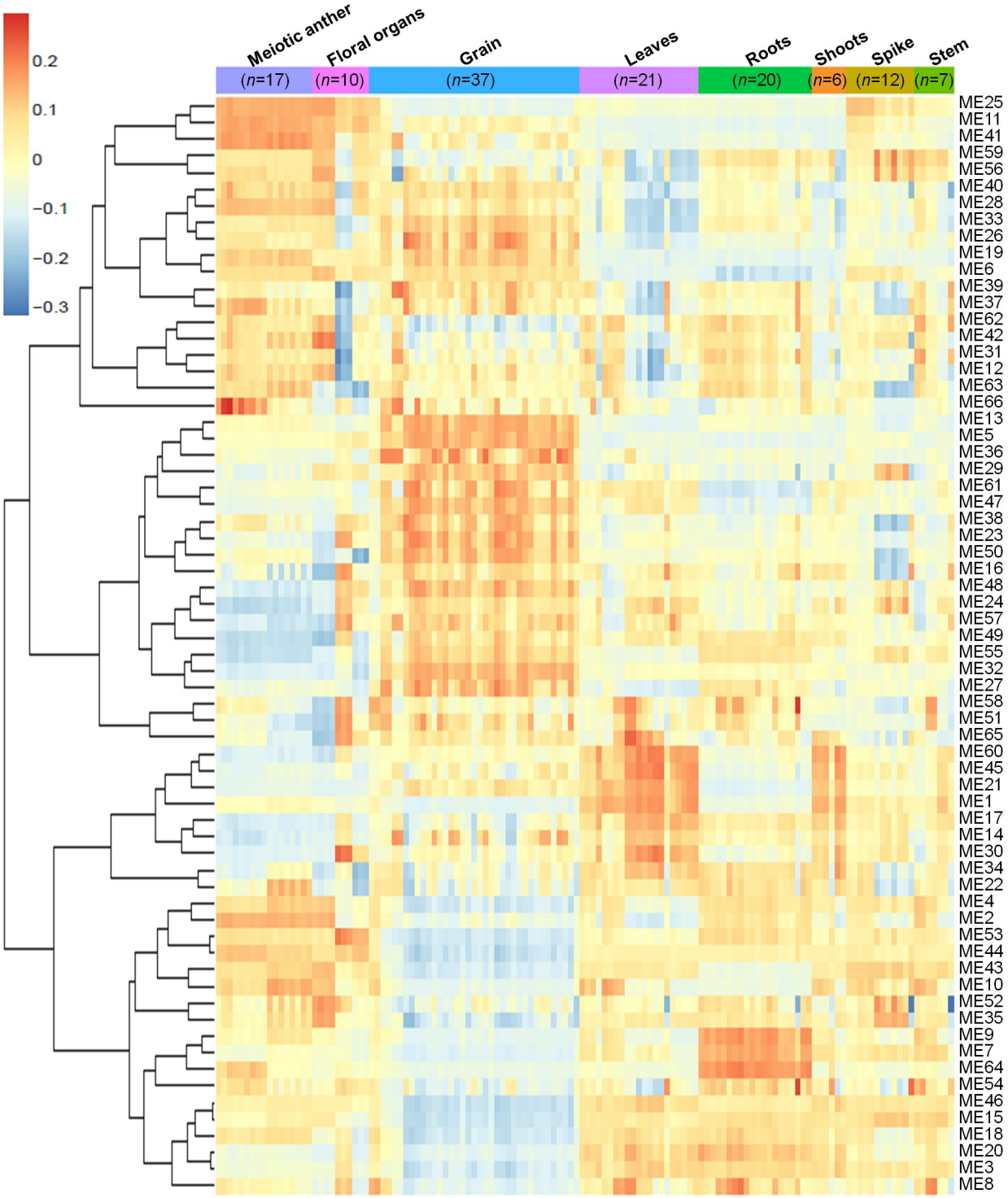
S10 Text. R script used to calculate homeolog expression patterns in triads.

S11. Text. R script used to export meiosis-related modules data for visualisation.

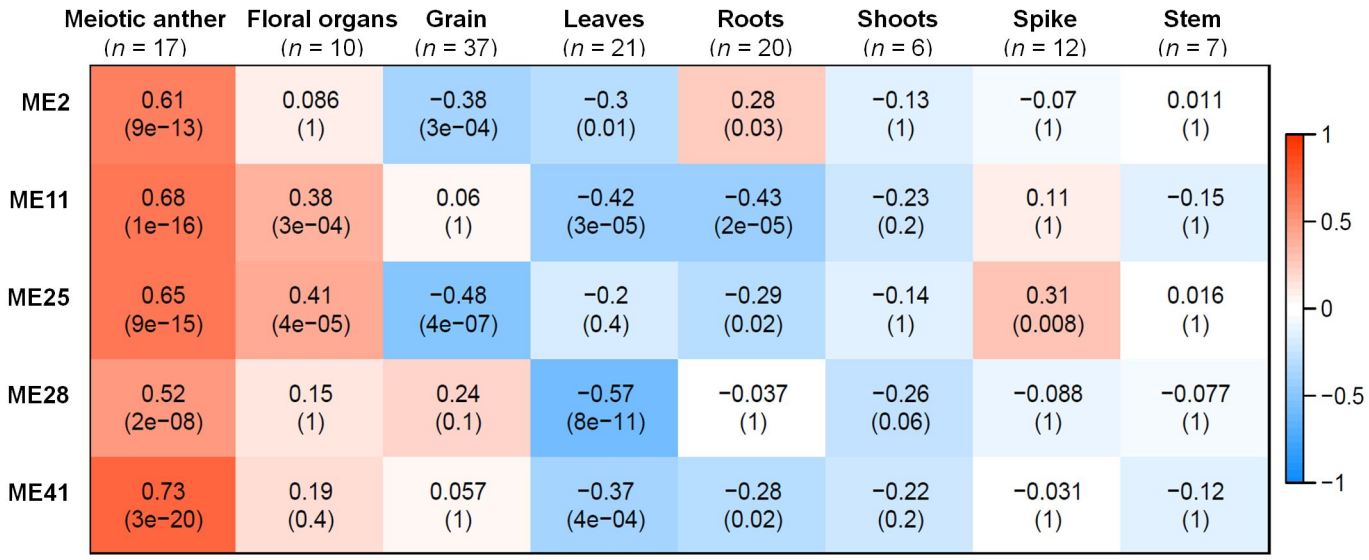
bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



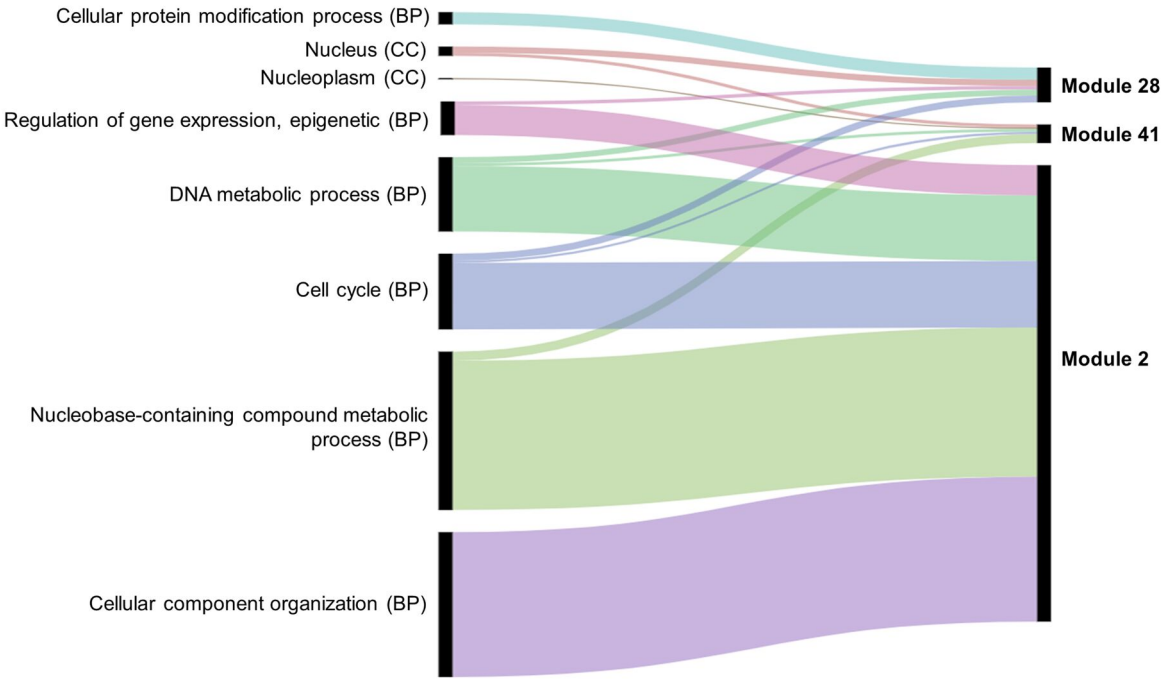
bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



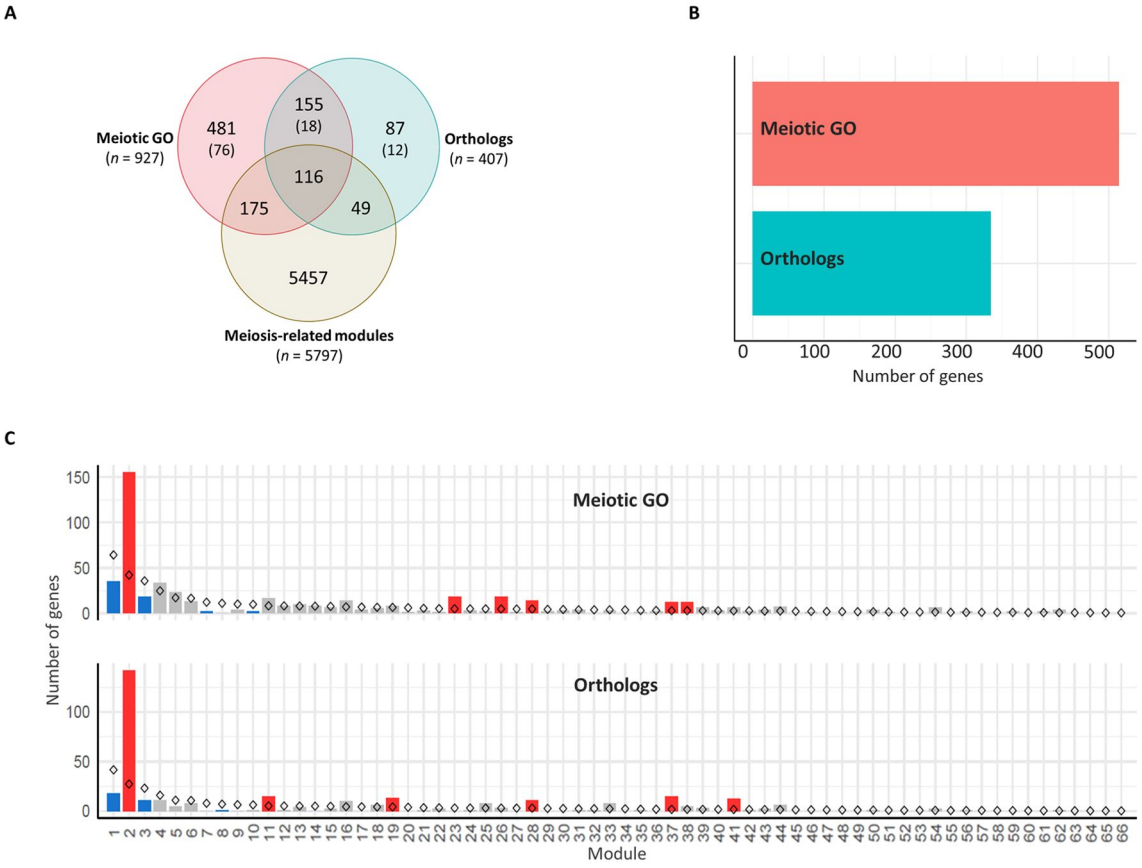
bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



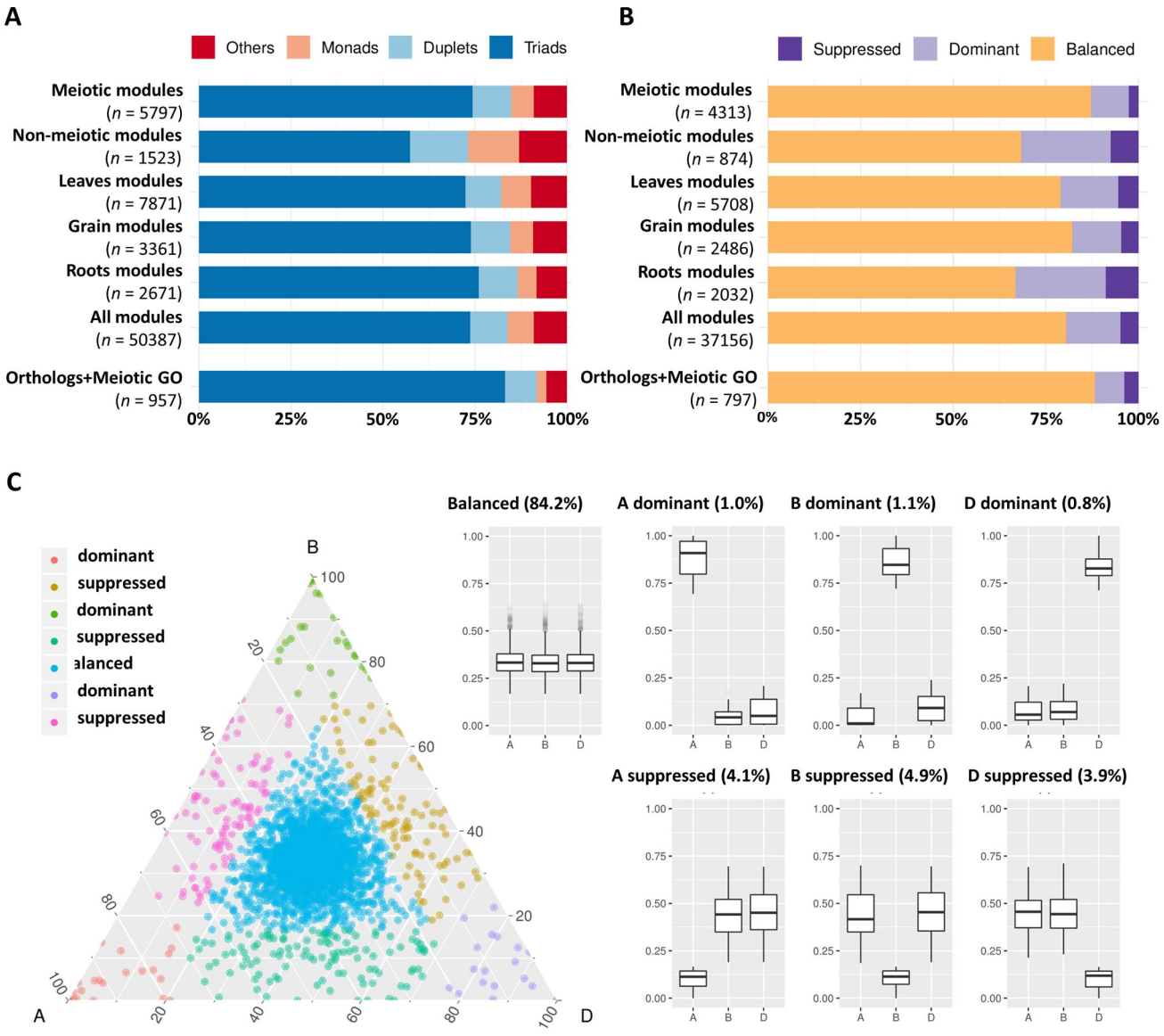
bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



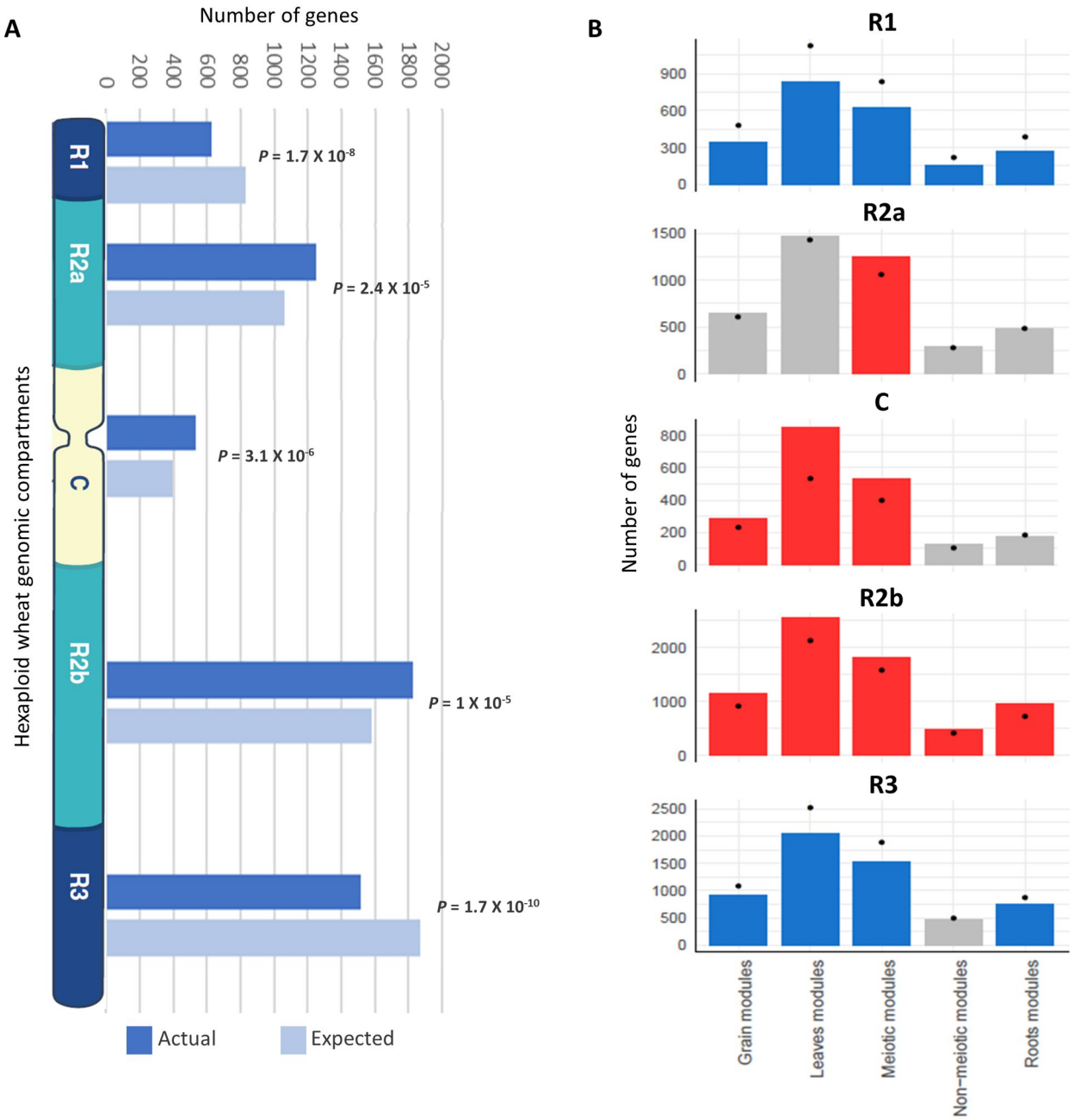
bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



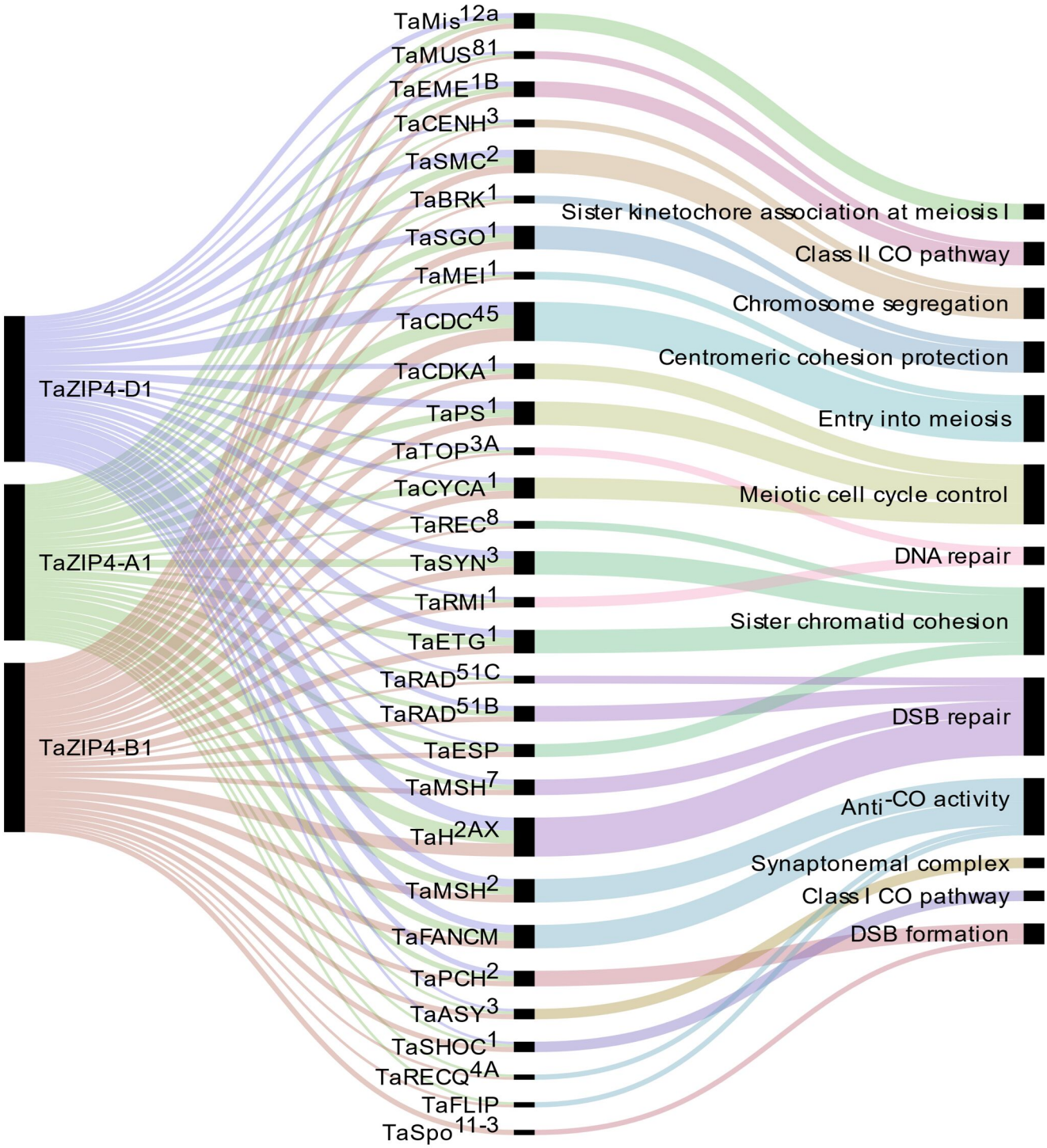
bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



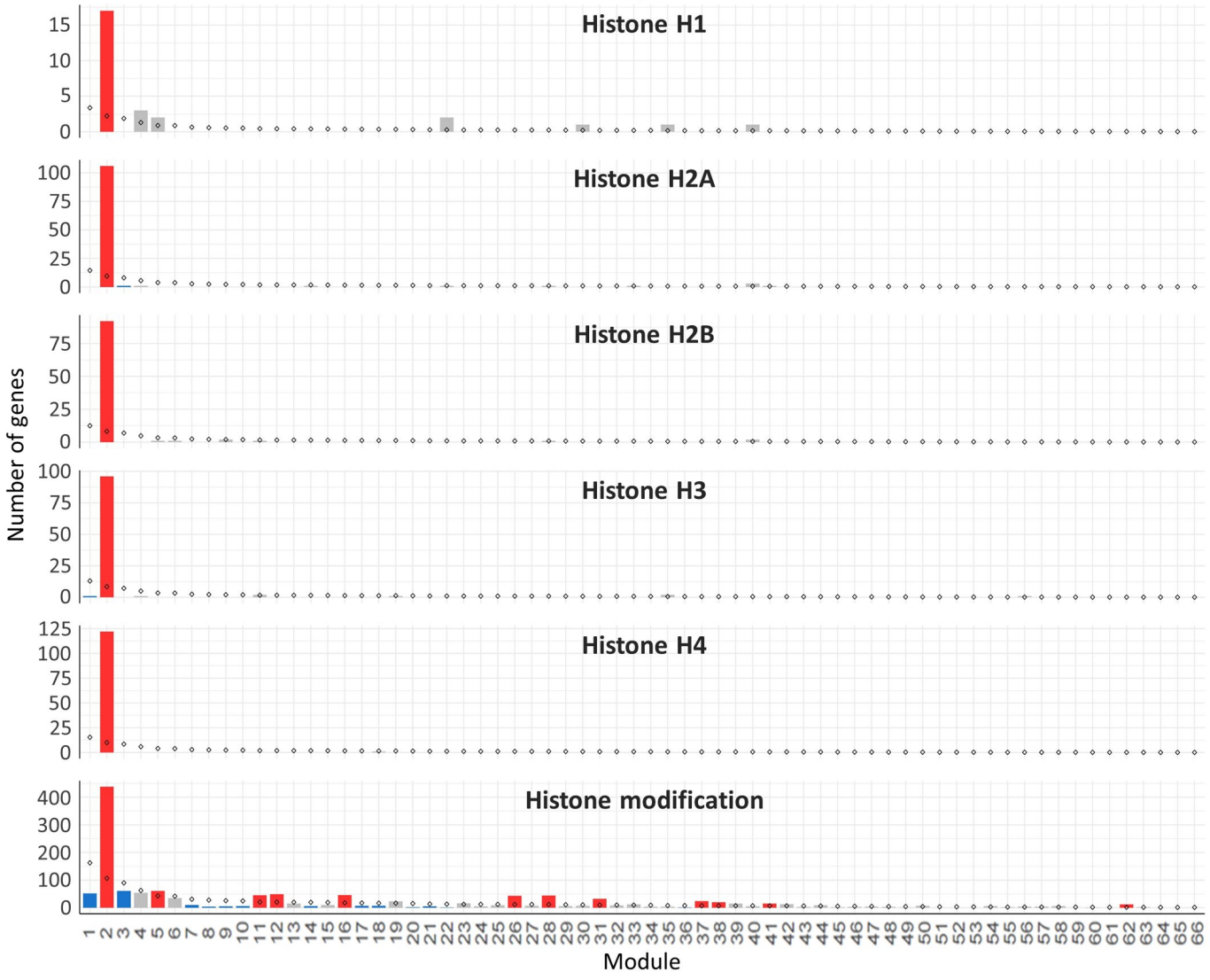
bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



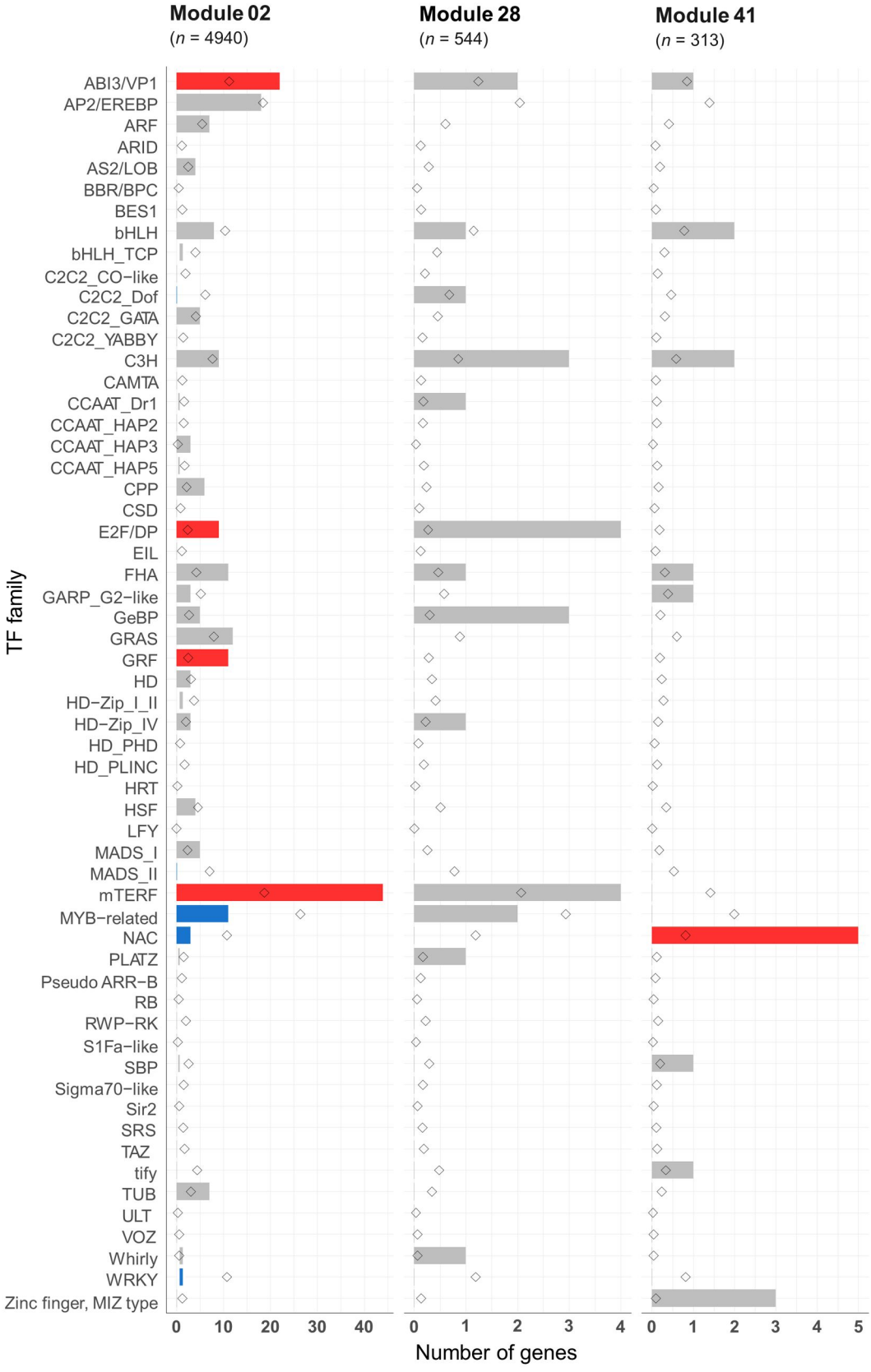
bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/695759; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.