**Methods**

**RNA-seq data analysis**

***Library preparation****:* A panel of 400 diverse wheat lines was selected based on the geographic distribution and genetic diversity (Supplementary Table 1), and available from the USDA National Small Grains Collection. Total RNA was isolated from 2-week old seedlings of 204 lines, with each line grown in three biological replicates. Ground tissues from three biological replicates were combined in equal amounts before RNA isolation using the RNeasy Plant mini kit. RNA-seq libraries were prepared with TruSeq™ RNA Sample Prep Kit (Illumina) using the Beckman’s Biomek® FXP Laboratory Automation Workstation. Up to eight barcoded RNA-seq libraries were pooled per lane of NextSeq2000 flow cell to generate 2 x 100 bp reads.

***RNA-seq read mapping:*** A total of 13,415,679,980 paired-end 2 x 100 bp reads were generated for 204 wheat accessions from the wheat diversity panel, with a mean of 65,763,137 reads per accession. On average, 81% of all reads were mapped to the genome uniquely, with an average of 7% reads failing to map (Supplementary Table 1).

In addition, we have analyzed previously published RNA-seq data generated for 90 wheat lines from spikes at the ﻿double-ridge development stage28. Fastq files were downloaded from NCBI BioProject PRJNA348655 using ‘fastq-dump’ from the SRA Toolkit (v. 2.9.6). The spike RNA-seq dataset contained 46,394,170 paired-end 2 x 125 bp reads, of which 86% reads could be mapped to the reference genome uniquely, and 5% of reads failed to map.

***Processing RNA-seq data from 2-week old seedlings:***We generated RNA-seq data for 2-week old seedlings from 204 wheat lines. We removed samples with a substantial amount of rRNA contamination and samples with less than 40% uniquely mapped reads. The resulting set of 198 RNA-seq samples was used for further analysis (Supplementary Table 1, PRJNA670223). We used the Kallisto program that uses pseudoalignment of RNA-seq reads to reference gene models to assess the transcript abundance66. Its performance has previously been evaluated in the polyploid wheat genome16.

All high confidence (HC) and low confidence (LC) gene models from the IWGSC RefSeq v. 1.05, as well as *de novo,* assembled transcripts were combined for estimating the TPM values using Kallisto (v. 0.4.6.0)66. Gene models with expression standard deviation > 0.5 and expressed (TPM > 0.5) in at least three samples have been used in our analyses. This set of genes included 52,511 HC gene models, 29,226 LC gene models, and 13,861 *de novo* assembled transcripts. The expression data were log2-transformed followed by robust quantile normalization in R. The ﻿probabilistic estimation of expression residuals (PEER) was used to remove hidden confounding factors in the expression data67, and residuals were used for studying the genetic effects on expression levels in the population.

***Assessing the accuracy of transcript abundance estimation:*** To assess the accuracy of transcript abundance estimation by mapping RNA-seq reads to the polyploid wheat genome, we have applied several approaches. The RNA-seq data was simulated using gene models of cultivar Chinese Spring using Flux Simulator (http://confluence.sammeth.net/display/SIM/Home). Comparison of transcript abundance estimated for simulated data using Kallisto with actual transcript abundance levels showed a high level of correlation (SCC = 0.98).

We also evaluated the accuracy of Kallisto-based transcript abundance estimates for duplicated homoeologous genes. For this purpose, we simulated RNA-seq datasets using gene models only from one of the wheat genomes (for example, the A genome) and then used all gene models from the wheat reference genome to calculate TPM values. Simulation performed for the A genome showed a high level of correlation (SCC=0.92, N=91,437) between the real values and those estimated using Kallisto. Only 0.1% of reads simulated using the A genome gene models were mapped to the B and D genomes, indicating high accuracy of transcript abundance estimates for the homeologous gene sets.

The same RNA-seq simulated dataset was used to estimate the accuracy of read mapping to the correct location in the wheat reference genome using HISAT268. We found that 98% of simulated reads could be unambiguously mapped by HISAT2 to the correct location in the wheat genome.

**SNP genotyping of diverse wheat accessions used for eQTL mapping**

We used a combination of different approaches to obtain genotyping data for the wheat diversity panel: 1) targeted re-sequencing of the regulatory regions of the wheat genome using a Nimblegen capture assay69, 2) wheat 90K SNP iSelect assay70, 3) complexity-reduced genome sequencing71, and 4) RNA-seq transcriptome dataset. SNPs discovered using the RNA-seq, and regulatory sequence capture datasets for 203 wheat accessions were combined, and missing genotype calls were imputed using Beagle72. This dataset was combined with the SNPs identified in the entire panel of 400 wheat accessions using the 90K iSelect assay70 and complexity-reduced genome sequencing71. Further, a common set of SNPs shared between our panel of 400 wheat accessions and 1,000 wheat exome dataset8, were used for genotype imputation (see details below).

***90K SNP iSelect assay:*** Flanking sequences of agenetically mapped set of ﻿46,977 SNPs from the 90K SNP iSelect assay70 were aligned to the IWGSC RefSeq v.1.0 using the BLAT program followed by filtering alignments using the following parameters: alignment coverage > 95%, sequence identity > 97%, e-value < 1e-10. We identified genomic coordinates for 23,577 uniquely aligned SNPs, which also showed consistency with the marker order in the previously created genetic maps70. For these SNP sites, we identified 16,037 SNPs segregating in our wheat panel of 400 wheat accessions (Supplementary Table 1).

***Targeted re-sequencing of regulatory regions by Nimblegen sequence capture:*** We have used previously developed regulatory capture assay69 to re-sequence 203 wheat accessions used for the transcriptome analysis. A total of 9,418,016,463 paired-end 2 x 150 bp reads were generated for 203 accessions, with the mean of 46,394,170 reads per accession. Reads were aligned using HISAT2 (v. 2.1.0) with the following parameters: --max-intronlen 70000, --no-spliced-alignment. On average, 87% of all reads were mapped to the genome uniquely, with an average of 8% reads failed to map (Supplementary Fig. 1). The recommended best practices were followed to call SNPs using GATK73. Base quality recalibration was performed using genotyping data generated for the same set of lines using the 90K iSelect assay70. The genotype calls for sites with less than 3 reads depth of coverage were set as missing data. SNPs were filtered to remove sites with more than two alleles, MAF<0.05, more than 50% genotype calls missing, and more than 3% heterozygote genotypes. In total, we have identified 3,320,006 SNPs segregating in the putative regulatory regions (Supplementary Table 1).

***SNP calling in RNA-seq data:*** The raw RNA-seq fastq files were processed using the NGSQC Toolkit (v2.3.3) with default parameters. We used HISAT2 (v. 2.1.0) to align reads to the IWGSC RefSeq v.1.0 with the default parameters, except for parameter --max-intronlen set to 70,000. We filtered out reads that are not uniquely mapped to avoid detecting variable sites due to misalignment to the homoeologous genomes. The GATK’s ’HaplotypeCaller’ was used to generate a *gvcf* file for each sample with the following parameters, ‘-dontUseSoftClippedBases -stand\_call\_conf 20.0 ‘. ‘GenotypeGVCFs’ was used to generate a multiple-sample VCF file for all variants. Only biallelic sites were used in our analysis. Genotype calls generated for sites with the depth of read coverage less than three or more than 50% genotype missing were set as missing data. Sites with more than 3% heterozygote genotype calls were removed. A total of 2.4 million SNPs were detected in the dataset, of which 138,481 SNPs with MAF > 0.05 were used for analyses.

***Complexity-reduced genome sequencing:*** Construction of complexity-reduced genomic libraries was performed as previously described71. Variant calling was accomplished using Tassel 5 GBS pipeline74. A total of 49,150 SNPs with MAF >0.01 were identified in the population of 400 wheat accessions (Supplementary Table 1).

***Imputation to fill un-genotyped sites:*** Genotype data from the 1,000 wheat exome project8 was used as a reference panel for imputation. An integrated VCF file was created, including all samples from 90K iSelect, complexity-reduced sequencing, RNA-seq, and 1000 exome capture panel. Beagle v. 4.172 (beagle.21Jan17.6cc.jar) was then used to impute missing genotype calls with the following settings: ‘overlap=500 window=5000  ne=12000’. The genotype calls with probability (GP) less than 0.8 were considered as missing. Sites with > 3% heterozygous genotype calls or > 75% missing data were removed, resulting in a set of about 195,000 SNPs.

***Imputation to predict missing genotypes:*** The VCF files from RNA-seq and regulatory sequence capture datasets were combined into a single VCF file. Imputation was used to fill in missing genotype calls using the same Beagle settings. After imputation, we set genotype calls with GP < 0.8 as missing data. All SNP sites with missing rate > 75% or heterozygosity rate < 3% were removed, resulting in a set of 4,453,487 SNPs. These SNPs were then merged with the variants identified using the 90K iSelect array and complexity-reduced sequencing, resulting in a set of 4,449,989 SNPs. A total of 2,021,936 SNPs with MAF > 0.05 in a panel of 198 wheat lines were used for eQTL mapping.

***Testing genotyping concordance:*** To assess the accuracy of genotype calling, weused genotyping data obtained for our panel using 90K iSelect array70. The genotype concordance rate for different SNP datasets was ~0.98 before imputation and 0.93 after imputation. We removed four wheat lines (GF54, GF30, GF38, GF312) (Supplementary Table 1) from the regulatory capture dataset that showed genotype data concordance below 0.8.

***SNP calling using RNA-seq from wheat spikes:*** Fastq files of 90 RNA-seq samples were downloaded from NCBI (BioProject PRJNA348655) using the ‘fastq-dump’ tool from SRA Toolkit (version 2.9.6). A total of 1.7 million SNPs were identified using the GATK pipeline. The same settings used for calling variants in the RNA-seq data generated for wheat seedlings were applied to RNA-seq from wheat spikes, except that 1) no imputation was performed, and 2) genotype calls supported by < 2 reads were set as missing. After filtering sites with more than 75% missing, 227,922 SNPs with MAF > 0.05 were used for eQTL mapping. The PEER residuals67 were calculated using the same method used for the seedling RNA-seq dataset. A total of 50,367 HC gene models from the IWGSC RefSeq v.1 were used for eQTL analysis.

**Detection of eQTL**

The association between SNPs and gene expression PEER residuals was performed by Matrix eQTL (v. 2.1.0)75 with the setting ‘useModel = modelLINEAR’. The first three principal components of the SNP matrix were used as covariates. All associations with FDR < 1e-5 were considered as significant. For each transcript, significantly associated SNPs were merged based on LD (*r*2>0.2) and distance (less than 100 kbp) into genomic intervals. SNP with the strongest association signal within an interval was defined as an eQTL of the transcript. If an eQTL was located within 2 Mb from the eGene, it was defined as *cis*-eQTL. eQTL located more than 2 Mbp away from an eGene were used to define distant *cis*-eQTL. In our analyses we used a conservative definition of *trans*-eQTL, which was an eQTL significantly associated with eGene located on a different chromosome.