Genome Biology

**SHORT REPORT**

# Enhancer transcription detected in the nascent transcriptomic landscape of bread wheat

Yilin Xie[1,2,3†], Yan Chen[1†], Zijuan Li[2†], Jiafu Zhu[1], Min Liu[1], Yijing Zhang[4*] and Zhicheng Dong[1*]

*Correspondence:
zhangyijing@fudan.edu.cn;
zc_dong@gzhu.edu.cn
†Yilin Xie, Yan Chen and
Zijuan Li contributed equally
to this work.
[1] Guangdong Provincial
Key Laboratory of Plant
Adaptation and Molecular
Design, Guangzhou
Key Laboratory of Crop
Gene Editing, Innovative
Center of Molecular
Genetics and Evolution,
School of Life Sciences,
Guangzhou University,
Guangzhou 510006, China
[4] State Key Laboratory
of Genetic Engineering,
Collaborative Innovation
Center of Genetics
and Development,
Department of Biochemistry,
Institute of Plant Biology,
School of Life Sciences,
Fudan University,
Shanghai 200438, China
Full list of author information
is available at the end of the
article

## Abstract

The precise spatiotemporal gene expression is orchestrated by enhancers that lack general sequence features and thus are difficult to be computationally identified. By nascent RNA sequencing combined with epigenome profiling, we detect active transcription of enhancers from the complex bread wheat genome. We find that genes associated with transcriptional enhancers are expressed at significantly higher levels, and enhancer RNA is more precise and robust in predicting enhancer activity compared to chromatin features. We demonstrate that sub-genome-biased enhancer transcription could drive sub-genome-biased gene expression. This study highlights enhancer transcription as a hallmark in regulating gene expression in wheat.

**Keywords:** Bread wheat, Nascent RNA, Enhancer, eRNA, Subgenome-bias

## Background

Widely cultivated wheat has an extremely large genome, which harbors abundant regulatory elements (REs) in noncoding regions [1, 2]. REs, including gene-proximal promoters and gene-distal enhancers, are key to the precise spatiotemporal regulation of gene expression [3, 4]. Enhancers affect the transcription of cognate genes independent of the relative distance, location or orientation [5–11]. Epigenomic studies revealed that REs often harbor particular chromatin features, including chromatin accessibility and certain histone marks, which have been widely used to characterize enhancers in both mammals and plants [1, 12–18]. Our labs and others recently reported that more than half of the regulatory hallmarks, including chromatin accessibility, histone modifications, and DNA methylation, are located in the intergenic regions of the wheat genome, suggesting the potential role of these epigenomic features in remote control of gene activity [1, 2]. However, an active chromatin environment does not necessarily mean an active enhancer [1, 19]. The transcription of enhancers has been widely detected by nascent RNA sequencing methods in animals [19–22], and enhancer RNAs (eRNAs) show more predictive

Xie *et al. Genome Biology*     (2022) 23:109

Page 2 of 15

power of enhancer activity than chromatin features [19]. Whether a plant enhancer can produce eRNA and its relevance to enhancer activity remain unknown.

## Results and discussion

To answer these questions, we modified two nascent RNA sequencing methods for wheat seedlings: global nuclear run-on sequencing (GRO-seq) and plant native elongating transcript sequencing (pNET-seq) [23, 24]. Both methods capture nascent transcripts regardless of their stability, thus enabling the localization of the exact position, amount and orientation of transcriptionally engaged RNA polymerase II (Pol II) (Additional file 1: Figure S1). Overall, 13% of the bread wheat genome is revealed to be transcribed by nascent RNA-seq in contrast to 8% by conventional rRNA depleted strand specific RNA-seq (ssRNA-seq) in terms of uniquely mapped reads. Approximately 5–65 million (GRO-seq) and 26–40 million (pNET-seq) uniquely mapped reads were obtained from each biological replicate, with better reproducibility and more valid data for pNET-seq and rRNA removal GRO-seq than traditional GRO-seq (see Methods; Additional file 2: Supplementary Table 1). In the genome, genic regions are considered to be transcribed. Pol II-mediated transcription is slower in exons than in introns to facilitate splicing recognition. Therefore, it is likely that more nascent RNAs are detected in exons than in introns [25]. As expected, one of the highest fractions of the aligned reads was derived from exons (Fig. 1a). Notably, the other highest proportion is transcribed from intergenic regions. In contrast, the Arabidopsis genome is less than 1% of the wheat genome, and only 18% is intergenic. Transcription was detected in 19% of Arabidopsis intergenic regions, accounting for 1% of nascent RNAs detected by GRO-seq (Additional file 1: Figure S2a). Compared with protein-coding transcripts, intergenic transcripts in the wheat genome are more likely to be involved in quick turnover events and therefore cannot be comprehensively detected by ssRNA-seq. Figure 1b shows a specific example of adjacent intergenic and protein-coding transcription within a 23 kb region, where strong GRO-seq and pNET-seq signals are found at both regions, while ssRNA-seq signals are confined to the protein-coding region. Moreover, the genome-wide read coverage of GRO-seq or pNET-seq was significantly higher than that of ssRNA-seq, and the gap widened with increasing sequencing depth (Additional file 1: Figure S2b). Not surprisingly, this gap was also observed in other species including human [26, 27] and

(See figure on next page.)

**Fig. 1** Unstable intergenic transcripts captured by GRO-seq and pNET-seq and eRNA characterization. **a** GRO-seq and pNET-seq reads aligned to different genomic regions. TSS1K indicates the 1 kb upstream region of the gene transcription start site. TES1K indicates the 1 kb downstream region of the transcription end site/ polyadenylation site. **b** Browser shot of a protein-coding transcript and an intergenic transcript, the number of reads aligned to forward (+) and reverse (−) genomic strands are separately displayed. **c** Expression levels of nascent/steady-state transcripts from protein-coding and intergenic regions. **d** Read densities of pNET-seq, DHS and ChIP-seq of Pol II, H3K9ac, H3K4me3, H3K36me3, H3K27ac, and H3K4me1 around the intergenic and genic TCs (± 3 kb). All intergenic and genic TCs were ranked in a descending order of pNET-seq signals (±250 bp around the 5′ end), and chromatin features were plotted around the 5′ end of each intergenic TC. **e** pNET-seq signals around the intergenic TCs. Intergenic TCs were divided into ten equal parts based on the decreasing level of pNET-seq signals (± 250 bp). **f** Read densities of DHS and ChIP-seq of H3K9ac, H3K4me3, H3K36me3, H3K27ac, and H3K4me1 around each of the ten parts of intergenic TCs. **g** Intergenic TCs counts within different chromatin states defined in wheat genome [1]. **h** Chromatin signatures of transcribed and untranscribed enhancer examples. **i** Boxplots showing the expression levels of unidirectional and bidirectional enhancers determined by pNET-seq. **j** A heatmap displays the expression levels of unidirectional and bidirectional enhancers in primary (sense) and/or secondary (antisense) orientations (± 3 kb)

Xie *et al. Genome Biology*     (2022) 23:109

Page 3 of 15



**Fig. 1** (See legend on previous page.)
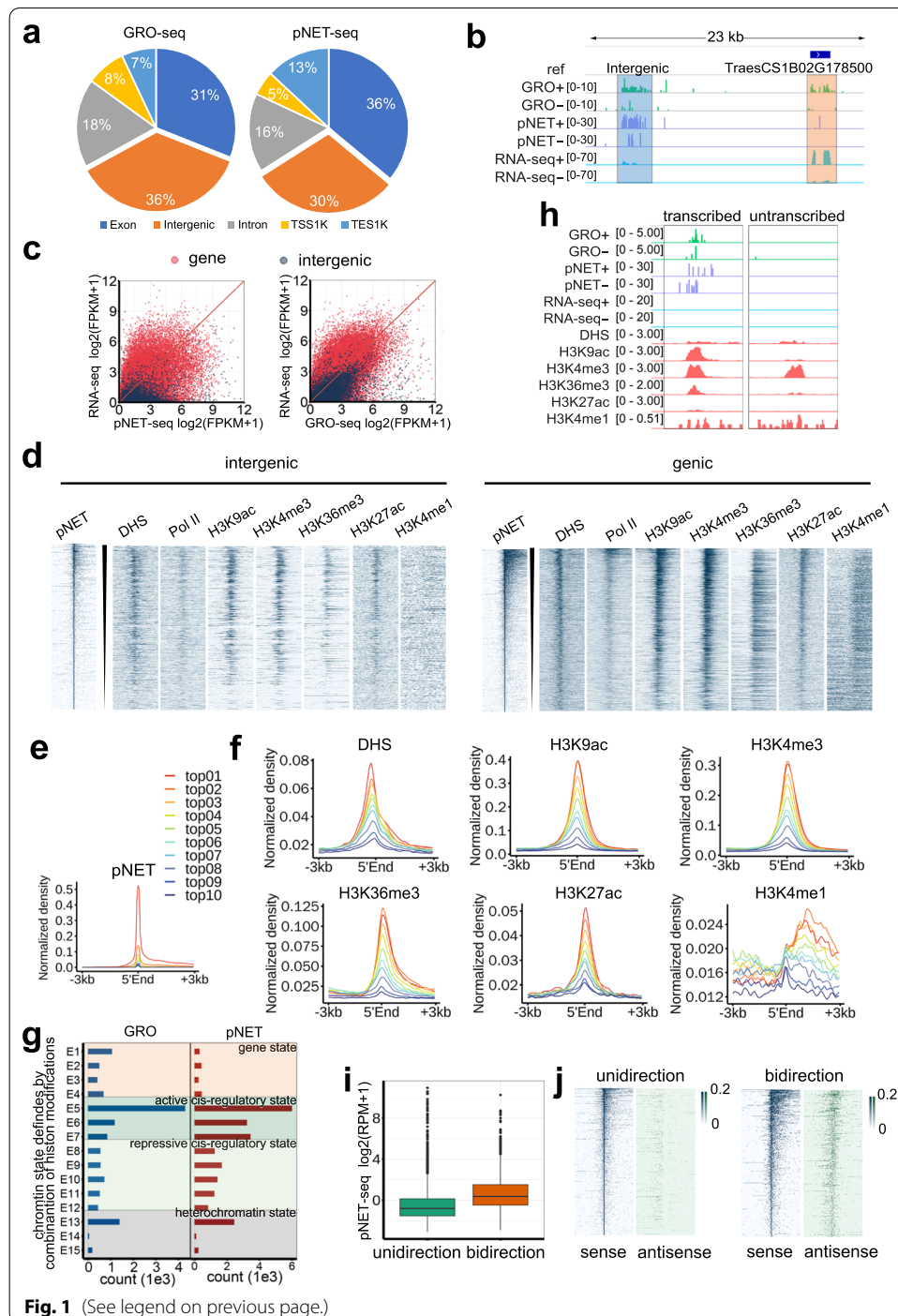
maize [28, 29], but not in Arabidopsis [23], which has a compact genome and much fewer intergenic regions.

Consistent with previous studies, apparent transcriptional pausing was observed in close proximity to active transcription start sites (TSSs) of annotated genes [23, 24] (Additional file 1: Figure S1, S2c). In total, we obtained 1,044,770 (GRO-seq) or 88,830 (pNET-seq) TSS clusters (TCs) through TSScall [20], of which a significant number were

Xie *et al. Genome Biology*     (2022) 23:109

Page 4 of 15

found to be within 1 kb upstream or downstream of the annotated genes (see Methods; Additional file 1: Figure S1, S2c; Additional file 3: Supplementary Table 2). As shown in Figure S3, the gene-proximal TCs identified above are indeed enriched around annotated TSSs, suggesting that TSScall is a robust approach for global TSS identification. Among the detected TCs, 45,872 (GRO-seq, 43.8%) and 33,355 (pNET-seq, 37.5%) were localized to intergenic regions, and intergenic TCs displayed higher nascent RNA levels than mature RNA levels, whereas no such bias was observed for protein coding transcripts (Fig. 1c). The overall correlations of transcription clusters between GRO-seq and pNET-seq were 0.57 and 0.60 at genic and intergenic regions, respectively (Additional file 1: Figure S4). Collectively, these results demonstrate that intergenic transcripts are indeed much less stable than gene transcripts and that the GRO-seq and pNET-seq combination is more sensitive in detecting intergenic transcripts than mature RNA sequencing.

We next characterized the epigenomic features surrounding intergenic and genic transcribed regions. Overall, both intergenic and genic transcription are strongly associated with open chromatin regions, RNA Pol II engagement and transcriptionally active histone marks including H3K9ac, H3K4me3, H3K36me3, and H3K27ac (Fig. 1d; Additional file 1: Figure S5). To investigate the relationship between transcription and chromatin features, we divided the intergenic TCs into ten groups according to their transcriptional activity determined by pNET-seq and plotted the epigenetic profiles surrounding the 5′ end of intergenic TCs (Fig. 1e, f). We observed that intergenic transcription was positively correlated with chromatin accessibility and the active histone marks H3K9ac and H3K4me3 as mentioned above (Fig. 1f). Within the genic region, the peaks of H3K9ac, H3K4me3, H3K36me3, and H3K27ac were biased to the transcriptional direction, while the peaks in the intergenic regions were symmetric (Fig. 1f; Additional file 1: Figure S6). These patterns suggest that the expression directions of genic and intergenic regions are different. Consistent with previous plant epigenome studies [1, 30], more H3K4me1 was detected in gene bodies than in intergenic regions (Fig. 1d; Additional file 1: Figure S5). However, there was a weak correlation between nascent RNA abundance and H3K4me1 levels in intergenic regions, whereas there was no clear relationship between H3K4me1 and nascent RNA levels in genes (Fig. 1f; Additional file 1: Figure S6). Analyses based on GRO-seq data led to the same results (Additional file 1: Figure S7). Together, transcribed intergenic regions were essentially marked by active chromatin states.

Given that intergenic transcribed regions in wheat hold a variety of active regulatory marks, we speculated that a fraction of intergenic TCs represents eRNAs. As indicated in Figure S8a, an intergenic transcribed region displayed high levels of enhancer hallmarks [1], including open chromatin, H3K9ac, H3K4me3, H3K36me3, and H3K27ac. Previously, we categorized the wheat genome into 15 chromatin states based on combinatorial histone modification marks using a multivariate hidden Markov model (HMM) which has been applied in detection of REs related to individual marks. Chromatin states 5–7 were marked by high DNA accessibility, histone acetylation, CpG islands, and relatively conserved sequences. Therefore, we defined the chromatin state 5–7 regions, which are 3 kb distant from the nearest gene, as enhancer-like elements. In a tobacco transient system, approximately half of the putative enhancer-like elements increase the expression of the reporter gene [1]. A total of 25% (11,484 out of 45,872, GRO-seq) and 38% (12,687 out of 33,355, pNET-seq) of intergenic TCs were from states 5–7 (Fig. 1g).

In fact, the proportions of intergenic enhancer-like elements being transcribed in states 5–7 were significantly higher than those in other states (Additional file 1: Figure S9). When compared to those that were untranscribed, the transcribed enhancer-like elements showed higher accessibility, more active histone marks, fewer repressive histone marks, and lower levels of CpG methylation (Additional file 1: Figure S8b, c; Fig. 1 h). No observable difference was uncovered between these two types of enhancers with respect to H3K4me1. Although the high H3K4me1 to H3K4me3 ratio is supposed to be a typical enhancer feature in mammals [1, 31, 32], emerging evidence in both *Drosophila* and mammals indicates that highly active enhancers are generally marked by H3K4me3 [20, 33–36], which is consistent with our results that transcribed enhancers are significantly enriched with H3K4me3 but not H3K4me1 (Additional file 1: Figure S8c). We next wondered whether there are sequence signatures within enhancers and found that WRKY, GeBP, zfGRF, and MYB-related transcription factor binding motifs were specifically overrepresented in enhancers rather than promoters (Additional file 1: Figure S10a). Moreover, these motifs are significantly enriched in transcribed enhancers compared with untranscribed enhancers (Additional file 1: Figure S10b).

Bidirectional transcription at enhancers and promoters has been widely discovered in mammals [37]. However, from Arabidopsis [23] and the wheat genome, no significant bidirectional promoter was observed (Additional file 1: Figure S2c). We then ask whether this is the case for plant enhancers. A total of 4407/2144 (pNET-seq/GRO-seq) bidirectional transcribed enhancers (BEs) and 8280/9340 (pNET-seq/GRO-seq) unidirectional transcribed enhancers (UEs) were revealed. Explicitly, here, for enhancers, the strand more actively transcribed is defined as the primary strand and the other strand is secondary. Corresponding transcripts are considered to be in primary or in secondary direction. From the pNET-seq assay, the transcriptional levels of BEs were found to be significantly higher than those of UEs in both the primary and secondary directions ($p < 2.2E-16$, Welch's two-sample *t*-test) (Fig. 1i), as evidenced by the stronger pNET-seq signals around the 5′ end of BEs (Fig. 1j). However, the expression levels of BEs were lower than UEs according to GRO-seq (Additional file 1: Figure S11a, b).

To further elucidate the biological significance of enhancer transcription, we linked these transcribed enhancers to putative gene targets. Herein, we employed the information from both epigenetic correlation and direct chromatin interaction between transcribed enhancers and gene promoters to define the targets of enhancers (see Methods; Additional file 1: Figure S1), resulting in 9925/9299 (pNET-seq/GRO-seq) transcribed enhancers connected with at least one target and 26,712/27,092 (pNET-seq/GRO-seq) genes targeted by at least one transcribed enhancer (Additional file 4: Table S3 and Additional file 5: Table S4). As indicated in Fig. 2a and Fig. S11c, genes associated with transcribed enhancer(s) displayed significantly higher expression levels than those without a transcribed enhancer. Further quantitative analysis revealed a positive correlation between enhancer and target transcription (Fig. 2b, Additional file 1: Figure S11d, S12). It is worth noting that none of the epigenetic markers except H3K9ac enriched in enhancers showed a quantitative association with target gene expression (Additional file 1: Figure S12). Moreover, the greater the number of associated transcribed enhancers, the higher the expression level of the target genes (Fig. 2c, Additional file 1: Figure S11e). To determine whether the direction of enhancer transcription would affect

Xie *et al. Genome Biology*     (2022) 23:109

Page 6 of 15

target gene expression, we calculated the expression levels of genes associated with unidirectional or bidirectional enhancers. The results showed that the target gene expression level is independent of the transcriptional direction of the associated enhancers (Fig. 2d). Likewise, enhancers in wheat regulate target gene expression regardless of the distance, position, and orientation (Additional file 1: Figure S13).

We then asked whether the enhancer transcription level was related to enhancer activity. We tested the "enhancer activity" of 36 putative intergenic enhancers with or without eRNA, which was determined using *gfp* or dual *luciferases* as reporters in wheat protoplasts (see Methods; Additional file 1: Figure S14, S1). The putative enhancer region was fused with a 35S minimal promoter (mini 35S pro) controlling the transcription of reporter genes, and the "enhancer activity" was defined as its effect on the reporter gene expression in comparison to the blank and negative controls, measured qualitatively by fluorescence of GFP or quantitatively by luciferase activity. The transcribed candidate enhancers were more active than untranscribed ones in wheat protoplasts (Fig. 2e; Additional file 1: Figure S15a). Furthermore, the positive candidate enhancers validated in the wheat protoplasts had higher transcriptional activity in wheat seedlings (Additional file 1: Figure S15b). Pearson correlation coefficients were calculated between the relative luminescence intensities (enhancer activity) and the read densities of nascent RNA (pNET-seq and GRO-seq) and different chromatin marks. Notably, enhancer activity was the most strongly correlated with enhancer transcription (GRO-seq and pNET-seq, $r = 0.55$), followed by chromatin accessibility (DHS, $r = 0.43$) and active histone modification (H3K9ac, $r = 0.17$; and H3K27ac, $r = 0.16$), which indicated that eRNA is more robust in predicting enhancer activity than canonical chromatin features (Fig. 2f).

(See figure on next page.)

**Fig. 2** A putative role of enhancer transcription in regulating subgenome-divergent gene expression and conservation of eRNA regions. **a**–**d** Boxplots showing transcript levels of target genes. **a** Genes associated with transcribed enhancer(s) display higher expression levels. **b** The enhancer transcriptional level positively correlates with its target gene expression. **c** The number of associated transcribed enhancers positively correlates with its target gene expression. **d** Genes targeted by unidirectional or bidirectional enhancers show no significant differences in expression levels. **e** The enhancer activity (relative intensity) determined in wheat protoplasts of transcribed and untranscribed enhancer candidates. **f** Correlations between the enhancer activity and the nascent RNA-seq, DHS, and ChIP-seq read density in predicted enhancer regions. **g** Schematic illustration of the biased expression homeolog genes targeted by enhancers with different transcription levels. Histone modification correlations and physical interactions between enhancers and putative target genes were counted. **h** Snapshot of a triad (*PR10*) with subgenome-biased expression and eRNAs. **i** Subgenome biased transcribed enhancers are enriched with unbalanced expressed homoeologs between each two subgenome pairs (from pNET-seq data). Each dot represents a gene pair in which one homeolog is associated with at least one transcribed enhancer in the same chromosome, while the other homeolog is not. The *X*-axis and *Y*-axis coordinates represent the expression levels of the two homeolog genes respectively. Red dots represent gene pairs in which the expression level of a homeolog gene associated with transcribed enhancer(s) was significantly higher than that of the other one not associated with transcribed enhancer; while the grey dots represent a homeolog gene associated with transcribed enhancer(s) was equally or lower expressed than the other one. "A w/ transcribed enhancer, B w/o transcribed enhancer" means homeolog A is associated with transcribed enhancer(s), but homeolog B is not. The odds ratio (OR) is a ratio of two sets of odds to measure the degree to which homeolog gene expression is correlated with its association with enhancer(s). See detailed analyses and interpretation in Additional file 1: Figure S15. **j** Merge all homoeolog pairs in **i** showed that the dominantly expressed homoeologs essentially associated with transcribed enhancers. **k**, **l** Nucleotide diversity distribution of enhancers with eRNA and random intergenic regions across the genome (**k**) and within A, B, and D subgenomes (**l**)
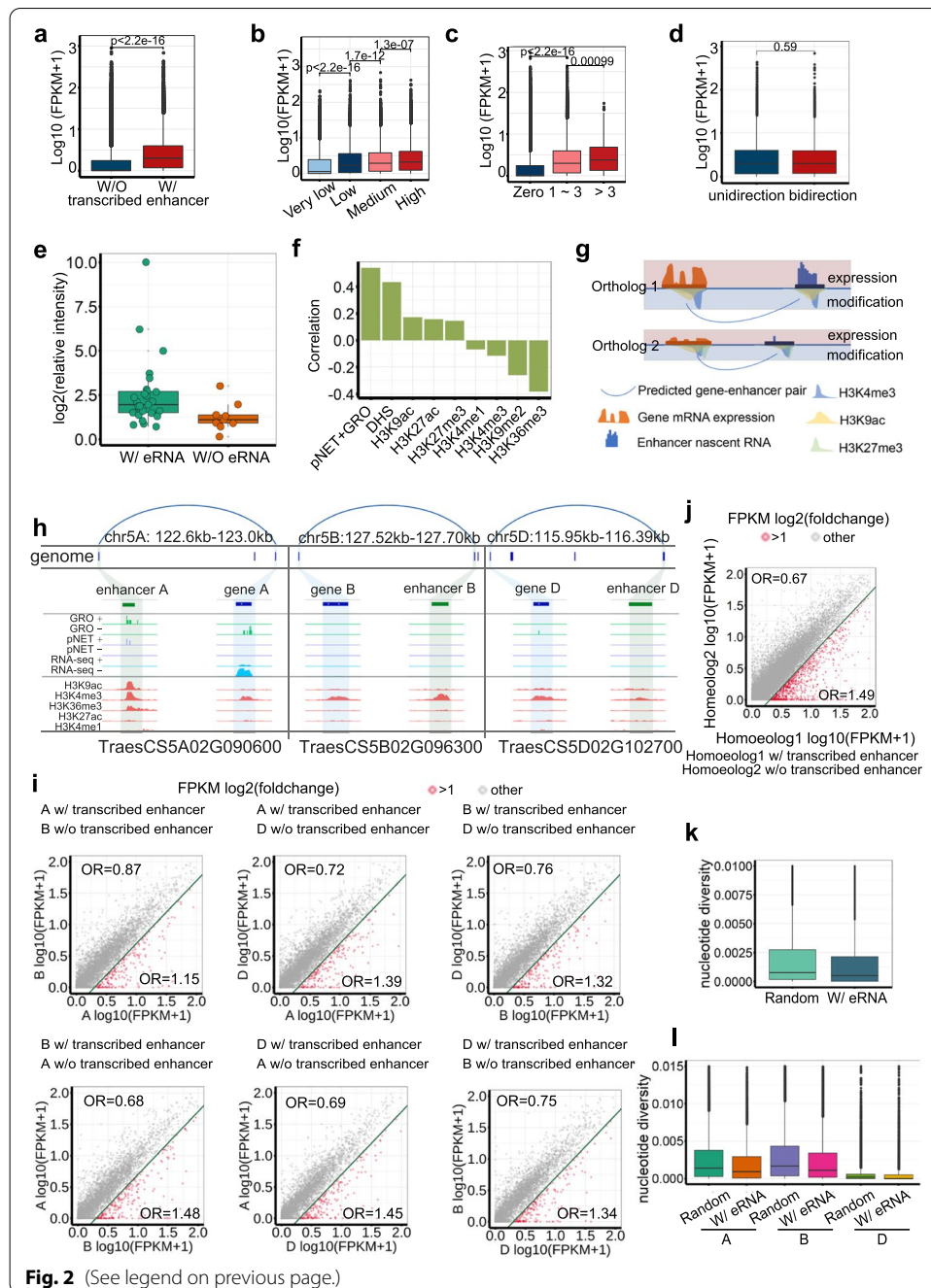
**Fig. 2** (See legend on previous page.)

For hexaploid, subgenome-divergent transcription was proposed as a major factor contributing to the superior adaptability and agronomic traits of bread wheat [38, 39]. Approximately 30% of wheat homoeolog triads are unevenly expressed, with one homoeolog being dominant or recessive relative to the other two [40]. Previous studies have demonstrated that subgenome-biased expression among homoeolog triads is closely associated with both epigenetic divergence and variation in transposable elements within promoter regions across subgenomes (Fig. 2g) [1, 40–42]. Here, we found a homoeolog triad of the *Pathogenesis-Related 10* (*PR10*) gene, which is differentially

Xie *et al. Genome Biology*    (2022) 23:109

Page 8 of 15

expressed in various tissues [43] and is associated with differently expressed enhancers. *PR10* from subgenomes A was targeted by a transcribed enhancer and displayed higher expression levels than the other two homoeologs from subgenomes B and D, which were associated with untranscribed enhancers (Fig. 2h). To verify whether subgenome-divergent gene expression is related to subgenome-biased enhancer transcription in addition to epigenetic modifications (Fig. 2g), we focused on 67,108 pairs of homoeolog genes that have 1:1 correspondence between any two of the three subgenomes, using the odds ratio (OR) to measure how strongly "one homoeolog gene being expressed higher than the other gene" is related to "one is associated with transcribed enhancer(s) but the other is not" (Additional file 1: Figure S16). We found that subgenome-biased gene expression was significantly enriched with subgenome-biased enhancer transcription (Fig. 2i, Additional file 1: Figure S17a). Specifically, dominantly expressed homoeologs were essentially associated with transcribed enhancers, at an OR of 1.49 (pNET-seq) or 1.64 (GRO-seq) (Fig. 2j, Additional file 1: Figure S17b).

Finally, we addressed whether transcribed enhancer regions were conserved in the natural wheat population. The nucleotide diversity was lower in the enhancer regions with eRNA than in the randomly chosen intergenic regions both across the genome and within each subgenome (Fig. 2k, l, Additional file 1: Figure S11f, g), which suggested that eRNA regions were selected for in the genetic improvement of wheat. We further examined the wheat eRNA homoeologous regions in the grass family; however, eRNA regions were no more conserved than the randomly picked intergenic regions (Additional file 1: Figure S18, S19). Taken together, eRNA coding regions are conserved among different wheat cultivars but not different grass species.

## Conclusions

The key insights into enhancer transcription, a defining nature for active enhancers, have thus far been provided by studies in *Drosophila* and mammals. However, the first nascent RNA study in plants demonstrated that few transcripts are produced from intergenic enhancers in Arabidopsis [24]. This is probably due to the relatively compact genome that contains fewer distal regulatory elements. Here, we surveyed the 16 Gb bread wheat genome comprising large intergenic regions and a wide variety of REs, including enhancers. By nascent RNA profiling, we detected actively transcribed enhancers in wheat. The overall chromatin signature of transcribed enhancers is quite similar to that of actively transcribed genes, marked with open chromatin, H3K9ac, H3K4me3, H3K36me3, and H3K27ac. However, bidirectional transcription was often observed at enhancers rather than at genic regions. The transcription levels of enhancers effectively predict the transcriptional status of their target genes in wheat. Furthermore, the subgenome-biased target gene expression was significantly enriched with subgenome-biased enhancer transcription, indicating that enhancer transcription should be one of the reasons for the asymmetric expression patterns between subgenomes and could be a reliable indicator of enhancer activity. As further genetic and biochemical evidences is required to dissect their functional mechanisms in plants, enhancers/eRNAs could be breeding targets either by traditional crossing or advanced genome editing in future crop improvement.

## Methods

### Plant materials and growth conditions

The bread wheat (*Triticum aestivum*) cultivar "Chinese Spring" was grown under long-day conditions. Leaf tissue of 12-day-old seedlings was collected by flash freezing with liquid nitrogen followed by nuclei isolation.

### GRO-seq experiment

Nuclei were isolated by the Percoll gradient procedure as previously described except that the concentration of Triton X-100 was reduced to 0.5% [23]. In vitro run-on and nascent RNA purification were performed as previously reported [23]. The obtained nascent RNA was treated with T4 PNK (NEB), followed by TRIzol purification and subject to small RNA library construction using NEXTflex™ Small RNA-Seq Kit v3 (PerkinElmer) according to the manufacturer's manual. We performed two rounds of GRO-seq. For the first time (rep1 and rep2), the ratio of unique mapping reads (we used for further transcription analyses) to total raw reads was 1% (Additional file 2: Supplementary Table 1). For the second time (v2_rep1, v2_rep2, and v2_rep3), to obtain a deeper sequencing coverage at a reasonable sequencing cost, we added an rRNA removal step after nuclear RNA isolation and before the affinity purification of nascent RNA using riboPOOL kit (siTOOLs BIOTECH). Then, the unique mapping ratio was increased to 16% (Additional file 2: Supplementary Table 1). The cDNA libraries were sequenced on an Illumina NovaSeq platform by Shanghai Hanyu Biotech. The two rounds of GRO-seq data were then merged for further analyses.

### pNET-seq experiment

Nuclei were isolated by a lysis-and-wash procedure as previously described except that the concentration of Triton X-100 was reduced to 0.5% [23]. The RNA Pol II-associated RNA was immunoprecipitated using Pol II antibody (Abcam, ab817), followed by small RNA cDNA library construction [23]. Ab817 was shown to recognize both unphosphorylated and phosphorylated wheat RNA Polymerase II by immunoblotting with total protein from control and flavopiridol-treated seedling tissue (Additional file 1: Figure S20). The size selection of nascent RNA was omitted. The cDNA libraries were sequenced on an Illumina NovaSeq platform by Shanghai Hanyu Biotech.

### GRO-seq data mapping and processing

Sequencing reads were cleaned with Trim Galore (version 0.6.4) and cutadapt (version 1.16) programs [44] to remove sequencing adapters, low quality bases ($< 20$), short reads and the 4 bases in the 5′ end and 3′ end of R1 and R2 reads, which were randomly introduced. Clean reads were aligned to the International Wheat Genome Sequencing Consortium (IWGSC) reference sequence (version 1.0) with the bowtie2 (version 2.3.5.1) program [45], and only the reads with one best unique alignment were retained. We also used the SortMeRNA (version 2.1b) program [46] to remove the reads originating from chloroplasts, mitochondria, and rRNA. The 5′ coordinate of R1 reads was considered the position of the engaged polymerase for future analyses. Each site density of GRO-seq was then calculated in plus and minus strands and normalized by the count of unique mapped reads. The correlations of the biological replicates were calculated by the read

density in 500-bp bins, and the alignments were then combined to obtain a higher depth for subsequent analysis.

### pNET-seq data mapping and processing

Sequencing reads were cleaned and mapped as described above except the 5′ coordinate of R2 reads with the directionality indicated by R1 reads was considered the position of the engaged polymerase. The pNET-seq signal at each site and the correlations of the 4 biological replicates were calculated as described above. The alignments of 4 biological replicates were combined for subsequent analysis.

### Intergenic transcript annotation

TSScall [20] with the default parameter was applied to define the GRO-seq and pNET-seq transcripts. Although this program was designed to identify TSSs from Start-seq data, visual inspection found that it performs well in calling short transcript start sites of GRO-seq and pNET-seq. For longer transcripts, it may call multiple peaks along with one transcript. Therefore, the nearby TSSs within 3 kb of one another were merged into TSS clusters (TCs). The 5′ end of the TC were selected as the TSS of this transcript. To avoid random sites and noise obtained by the experimental technique, only the TSSs found in all replicates were retained.

With the aim of accurately achieving the intergenic transcript, we removed the TCs that overlapped with the annotated high confidence and low confidence genes (IWGSC v1.1). To avoid false-positives due to incomplete annotations, we also removed the TCs that overlapped with the upstream and downstream 1 kb of annotated genes. In addition, we applied StringTie [47] and CPC2 [48] to assemble and predict the potential protein coding transcripts using multitissue ssRNA-seq alignments. TCs with overlap with these transcripts were also removed.

### Comparison of genomic coverage in bread wheat and other species

Human data were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2400210 for RNA-seq and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM340901 for GRO-seq). Maize data were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSM2041246 for RNA-seq and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1309038 for GRO-seq). Arabidopsis data were download from NCBI (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2974957 for RNA-seq and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2974947 for GRO-seq).

For each dataset, we randomly selected 0.1–80 million alignments from the bam files, calculated the coverage in the 10-kb bins of the corresponding genome and set the cutoff to 5 to determine whether the bin was covered.

### RNA-seq data processing

RNA-seq data used for comparison with nascent RNA-seq data were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3449733). The RNA-seq dataset used to assemble transcripts with coding potential was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139019). The

sequencing reads were trimmed with Trim Galore (version 0.6.4), and clean reads were mapped to the IWGSC reference sequence (version 1.0) with the HISAT2 program (version 2.1.0) [49]. The featureCount program of the Subread package (version 1.5.3) [50] was used to determine the RNA-seq read density of the high-confidence genes in the IWGSC RefSeq genome assembly (version 1.1). The RNA-seq read density of each gene was normalized based on the exon length in the gene and the sequencing depth [i.e., fragments per kilobase of exon model per million mapped reads (FPKM)]. Sixteen RNA-seq samples were used to assemble the potential transcripts using the StringTie program (2.1.4) [47]. The coding potential was predicted by the CPC2 program [48].

### Bisulfite-seq, ChIP-seq, and DNase-seq data processing

Bisulfite-seq, ChIP-seq, and DNase-seq data of Chinese spring seedling tissue were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121903). The reads were cleaned and mapped as described previously [1]. To compare the differences in the methylation modification levels in different intervals of the genome, the average methylation ratios in the 500-bp bins of the whole genome were calculated.

### Motif enrichment analysis

The motifs were analyzed with "findMotifs.pl" implemented in Homer software [51]. Taking random genomic regions as the background, the 500-bp region flanking the promoters or the predicted enhancers was chosen as the primary sequence input, and then, we calculated the enrichment fold change of each known motif in plants.

Taking promoter sequence as the background and the predicted enhancer sequence as the input and vice versa, we found motifs that were differentially enriched in the promoter and enhancer. Then, with random genomic regions as background, we calculated the enrichment fold change of these motifs in transcribed and untranscribed enhancers separately.

### Assignment of transcribed enhancers to target genes

We searched the transcribed enhancers and target gene pairs according to the distance, the correlation of histone modification, and the 3D interaction intensity between them. First, only genes within 500 kb of the enhancer are considered. The recently published modification peaks were used [52], and we chose the peaks that overlapped with the enhancer and promoter pairs to calculate the correlation. The cutoff was set to 0.7. Hi-C data were used to count the interaction intensity between the candidate pairs [53]. The filtered result must be connected by at least one Hi-C read pair. If no eligible target gene was found within 500 kb, then the search range was expanded to 2 Mb.

### Verification of enhancer activity by a luciferase reporter assay in wheat protoplasts

Sequences of the predicted intergenic enhancers were amplified from Chinese Spring genomic DNA (primers are listed in Additional file 6: Table S5) and fused with a minimal 35S promoter, which drives the transcription of a *Fluc* (*firefly luciferase*) or *gfp* (*green fluorescent protein*) gene [1] (Additional file 1: Figure S14). Wheat protoplast preparation and transformation were performed according to a previously reported protocol [54]. Briefly, wheat leaves were cut into strips with a razor blade and then incubated with cellulase and macerozyme to release protoplasts. A total of 15 μg plasmids

were transformed into $1 \times 10^5$ protoplasts using the PEG-mediated transfection method. After 24 h of incubation under dark conditions at room temperature, protoplasts were collected for visualization of fluorescence or luciferase assays. The protoplasts transformed with *gfp* reporter constructs were observed with a confocal microscope (Nikon C2 plus) (Additional file 1: Figure S15a). For the dual luciferase reporter system, *Rluc* (*Renilla luciferase*) driven by a *UBQ10* promoter was used to monitor transfection efficiency in the same vector (Additional file 1: Figure S14). The luciferase assay was completed using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions on a BioTek Synergy 2 microplate reader. The relative expression level of each construct was defined as the ratio of Fluc to Rluc activity. To calculate the enhancer activity (relative intensity), the relative expression level of each enhancer candidate construct was normalized to that of the blank control, in which *Fluc* was only driven by the minimal 35S promoter. An intergenic region where there were no DHS, pNET-seq, and GRO-seq signals was used as a negative control. The average of three independent biological replicates was recorded for each construct. The enhancer activity (relative intensity) of the negative control was 1.6; thus, we set 2.0 as the cutoff of the positive enhancer. The activity for each enhancer candidate is listed in Additional file 7: Table S6. For correlation coefficient calculation between enhancer activity and eRNA/histone modification/DHS levels, extreme out group points were omitted.

### Nucleotide diversity calculation

Twenty whole genome resequencing data were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA476679) [55]. The reads were cleaned and mapped as described previously [1]. The low-quality and duplicate reads were removed. SNP/indel detection was performed using the GATK HaplotypeCaller GenotypeGVCFs (version 4.1.9.0) set for diploids with default filtering settings [56]. Then, SNPs were further excluded by Plink (v1.90b6.18) [57] with the parameter "-geno 0.05 -mind 0.05 -hwe 0.0001 -maf 0.05". The nucleotide diversity was calculated using vcftools [58] in 500 bp windows across the genome.

### Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-022-02675-1.

**Additional file 1: Figures S1**. A diagram of the experimental approach. **Figures S2-S4**. GRO-seq and pNET-seq data quality. **Figures S5-S10**. Features of genic and intergenic transcription loci. **Figures S11-S13**. Characterization of eRNAs. **Figures S14-S15**. Validation of enhancer activity in wheat protoplasts. **Figures S16-S17**. eRNA and sub-genome-biased gene expression. **Figures S18-S19**. Conservation of eRNA regions [67]. **Figure S20**. Validation of anti-Pol II antibodies. **Figure S21**. Uncropped images for Figure S20.

**Additional file 2: Supplementary Table S1**. Statistics of nascent RNA sequencing data quality.

**Additional file 3: Supplementary Table S2**. TSS and TC counts identified by TSScall.

**Additional file 4: Supplementary Table S3**. Transcribed enhancer counts linked to homeolog pairs (pNET).

**Additional file 5: Supplementary Table S4**. Transcribed enhancer counts linked to homeolog pairs (GRO).

**Additional file 6: Supplementary Table S5**. Genomic positions and primers used for the reporter assay.

**Additional file 7: Supplementary Table S6**. Location, eRNA level, and enhancer activity of 36 putative enhancers tested in wheat protoplasts.

**Additional file 8.** Review history.

Xie *et al. Genome Biology*     (2022) 23:109

Page 13 of 15

## Review history
The review history is available as Additional file 8.

## Peer review information
Wenjing She was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

## Authors' contributions
YZ and ZD designed the research. YC, ZL, and JZ performed the research. YX and ML analyzed the data. YC, YX, YZ, and ZD wrote the manuscript. All authors contributed to the article and approved the submitted version.

## Availability of data and materials
The GRO-seq and pNET-seq data of bread wheat in this study are available in the Gene Expression Omnibus (GEO) database under accession number GSE178276 [59]. RNA-seq, BS-seq, ChIP-seq, and DNase-seq data of bread wheat were published previously [60, 61]. For comparison of genomic coverage, both RNA-seq and nascent RNA-seq data of bread wheat were analyzed in parallel with the data of other species including human [62, 63], maize [64, 65], and *Arabidopsis* [66].

# Declarations

## Ethics approval and consent to participate
Not applicable.

## Competing interests
The authors declare that they have no competing interests.

## Author details
[1]Guangdong Provincial Key Laboratory of Plant Adaptation and Molecular Design, Guangzhou Key Laboratory of Crop Gene Editing, Innovative Center of Molecular Genetics and Evolution, School of Life Sciences, Guangzhou University, Guangzhou 510006, China. [2]National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. [3]University of the Chinese Academy of Sciences, Beijing 100049, China. [4]State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, Department of Biochemistry, Institute of Plant Biology, School of Life Sciences, Fudan University, Shanghai 200438, China.

## References
1. Li Z, Wang M, Lin K, Xie Y, Guo J, Ye L, et al. The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. Genome Biol. 2019;20:139.
2. International Wheat Genome Sequencing C, investigators IRp, Appels R, Eversole K, Feuillet C, Keller B, Rogers J, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science. 2018;361:eaar7191.
3. Weber B, Zicola J, Oka R, Stam M. Plant enhancers: a call for discovery. Trends Plant Sci. 2016;21:974–87.
4. Sheffield NC, Furey TS. Identifying and characterizing regulatory sequences in the human genome with chromatin accessibility assays. Genes (Basel). 2012;3:651–70.
5. Spilianakis CG, Lalioti MD, Terrence T, Gap R. Interchromosomal associations between alternatively expressed loci. Nature. 2005;435:637–45.
6. Benko S, Fantes JA, Amiel J, Kleinjan D-J, Thomas S, Ramsay J, et al. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nat Genet. 2009;41:359–64.
7. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell. 2011;144:327–39.
8. Zhang Y, Wong C-H, Birnbaum RY, Li G, Favaro R, Ngan CY, et al. Chromatin connectivity maps reveal dynamic promoter–enhancer long-range associations. Nature. 2013;504:306–10.
9. Maniatis T, Goodbourn S, Fischer JA. Regulation of inducible and tissue-specific gene expression. Science. 1987;236:1237–45.
10. Kim T-K, Shiekhattar R. Architectural and functional commonalities between enhancers and promoters. Cell. 2015;162:948–59.
11. Tsai P-F, Dell'Orso S, Rodriguez J, Vivanco KO, Ko K-D, Jiang K, et al. A muscle-specific enhancer RNA mediates cohesin recruitment and regulates transcription in trans. Mol Cell. 2018;71:129–141.e128.
12. Wright JB, Sanjana NE. CRISPR screens to discover functional noncoding elements. Trends Genet. 2016;32:526–9.
13. Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schernhuber K, et al. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature. 2014;512:91–5.
14. Firpi HA, Ucar D, Tan K. Discover regulatory DNA elements using chromatin signatures and artificial neural network. Bioinformatics. 2010;26:1579–86.

Xie *et al. Genome Biology* (2022) 23:109

Page 14 of 15

15. Zhu B, Zhang W, Zhang T, Liu B, Jiang J. Genome-wide prediction and validation of intergenic enhancers in Arabidopsis using open chromatin signatures. Plant Cell. 2015;27:2415–26.
16. Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, et al. High-resolution mapping of open chromatin in the rice genome. Genome Res. 2012;22:151–62.
17. Meng F, Zhao H, Zhu B, Zhang T, Yang M, Li Y, et al. Genomic editing of intronic enhancers unveils their role in fine-tuning tissue-specific gene expression in Arabidopsis thaliana. Plant Cell. 2021;33:1997–2014.
18. Oka R, Zicola J, Weber B, Anderson SN, Hodgman C, Gent JI, et al. Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. Genome Biol. 2017;18:137.
19. Tippens ND, Liang J, Leung AK, Wierbowski SD, Ozer A, Booth JG, et al. Transcription imparts architecture, function and logic to enhancer units. Nat Genet. 2020;52:1067–75.
20. Henriques T, Scruggs BS, Inouye MO, Muse GW, Williams LH, Burkholder AB, et al. Widespread transcriptional pausing and elongation control at enhancers. Genes Dev. 2018;32:26–41.
21. Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature. 2013;498:516–20.
22. Hah N, Murakami S, Nagari A, Danko CG, Kraus WL. Enhancer transcripts mark active estrogen receptor binding sites. Genome Res. 2013;23:1210–23.
23. Zhu J, Liu M, Liu X, Dong Z. RNA polymerase II activity revealed by GRO-seq and pNET-seq in Arabidopsis. Nat Plants. 2018;4:1112–23.
24. Hetzel J, Duttke SH, Benner C, Chory J. Nascent RNA sequencing reveals distinct features in plant transcription. Proc Natl Acad Sci U S A. 2016;113:12316–21.
25. Leng X, Ivanov M, Kindgren P, Malik I, Thieffry A, Brodersen P, et al. Organismal benefits of transcription speed control at gene boundaries. EMBO Rep. 2020;21:e49315.
26. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.
27. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science. 2008;322:1845–8.
28. Lu X, Zhou X, Cao Y, Zhou M, McNeil D, Liang S, et al. RNA-seq analysis of cold and drought responsive transcriptomes of Zea mays ssp. mexicana L. Front Plant Sci. 2017;8:136.
29. Erhard KF Jr, Talbot JE, Deans NC, McClish AE, Hollick JB. Nascent transcription affected by RNA polymerase IV in Zea mays. Genetics. 2015;199:1107–25.
30. Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol. 2009;10:R62.
31. Robertson AG, Bilenky M, Tam A, Zhao Y, Zeng T, Thiessen N, et al. Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. Genome Res. 2008;18:1906–17.
32. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007;39:311–8.
33. Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, et al. H3K4 tri-methylation provides an epigenetic signature of active enhancers. EMBO J. 2011;30:4198–210.
34. Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, et al. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nat Struct Mol Biol. 2011;18:956–63.
35. Andersson R, Sandelin A, Danko CG. A unified architecture of transcriptional regulatory elements. Trends Genet. 2015;31:426–33.
36. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473:43–9.
37. Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat Genet. 2014;46:1311–20.
38. Bird KA, VanBuren R, Puzey JR, Edger PP. The causes and consequences of subgenome dominance in hybrids and recent polyploids. New Phytol. 2018;220:87–93.
39. Powell JJ, Fitzgerald TL, Stiller J, Berkman PJ, Gardiner DM, Manners JM, et al. The defence-associated transcriptome of hexaploid wheat displays homoeolog expression and induction bias. Plant Biotechnol J. 2017;15:533–43.
40. Ramirez-Gonzalez RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, et al. The transcriptional landscape of polyploid wheat. Science. 2018;361:eaar6089.
41. Gardiner LJ, Quinton-Tulloch M, Olohan L, Price J, Hall N, Hall A. A genome-wide survey of DNA methylation in hexaploid wheat. Genome Biol. 2015;16:273.
42. Guo X, Han F. Asymmetric epigenetic modification and elimination of rDNA sequences by polyploidization in wheat. Plant Cell. 2014;26:4311–27.
43. Mohammadi M, Srivastava S, Hall JC, Kav NN, Deyholos MK. Two wheat (Triticum aestivum) pathogenesis-related 10 (PR-10) transcripts with distinct patterns of abundance in different organs. Mol Biotechnol. 2012;51:103–8.
44. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–2.
45. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
46. Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28:3211–7.
47. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11:1650–67.
48. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res. 2017;45:W12–6.
49. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357–60.
50. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.

Xie *et al. Genome Biology*    (2022) 23:109

Page 15 of 15

51. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576–89.

52. Wang M, Li Z, Ye Z, Zhang Y, Xie Y, Ye L, et al. An atlas of wheat epigenetic regulatory elements reveals subgenome-divergence in the regulation of development and stress responses. Plant Cell. 2021;33:865–81.

53. Concia L, Veluchamy A, Ramirez-Prado JS, Martin-Ramirez A, Huang Y, Perez M, et al. Wheat chromatin architecture is organized in genome territories and transcription factories. Genome Biol. 2020;21:104.

54. Shan Q, Wang Y, Li J, Gao C. Genome editing in rice and wheat using the CRISPR/Cas system. Nat Protoc. 2014;9:2395–410.

55. Zhou Y, Zhao XB, Li YW, Xu J, Bi AY, Kang LP, et al. Triticum population sequencing provides insights into wheat adaptation. Nat Genet. 2020;52:1412–22.

56. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.

57. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

58. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8.

59. Xie Y, Chen Y, Li Z, Zhu J, Liu M, Zhang Y, Dong Z, Enhancer transcription detected in nascent transcriptomic landscape of bread wheat. Gene Expression Omnibus Database.2022; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE178276.

60. Li Z, Wang M, Lin K, Xie Y, Guo J, Ye L, Zhuang Y, Teng W, Ran X, Tong Y, et al: The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. Gene Expression Omnibus Database.2019; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121903.

61. Wang M, Li Z, Zhang Y, Zhang Y, Xie Y, Ye L, Zhuang Y, Lin K, Zhao F, Guo J, et al: An atlas of wheat epigenetic regulatory elements reveals subgenome divergence in the regulation of development and stress responses. Gene Expression Omnibus Database.2021; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139019.

62. ENCODE Project Consortium: An integrated encyclopedia of DNA elements in the human genome. Gene Expression Omnibus Database.2016; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE90257.

63. Core LJ, Waterfall JJ, Lis JT: Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Gene Expression Omnibus Database. 2008; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13518.

64. Lu X, Zhou X, Cao Y, Zhou M, McNeil D, Liang S, Yang C: RNA-seq analysis of cold and drought responsive transcriptomes of *Zea mays* ssp. *mexicana* L. Gene Expression Omnibus Database. 2016; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76939.

65. Erhard KF, Jr., Talbot JE, Deans NC, McClish AE, Hollick JB: Nascent transcription affected by RNA polymerase IV in Zea mays. Gene Expression Omnibus Database. 2015; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54166.

66. Zhu J, Liu M, Liu X, Dong Z. RNA polymerase II activity revealed by GRO-seq and pNET-seq in Arabidopsis. Gene Expression Omnibus Database. 2018; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE109974.

67. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.