nature portfolio

Peer Review File

Population-scale gene expression analysis reveals the contribution of expression diversity to the modern wheat improvement

Corresponding Author: Dr Fei He

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This study performs population RNA-seq on a diverse wheat panel with existing genome sequences. The authors present complex datasets effectively, though some sections are overly verbose. Key advances include using pan-genome resources for RNA-seq alignment (identifying >20,000 novel transcripts and improving alignment by 20%) and integrating GWAS, SMR, and TWAS to prioritize pan-genome candidates. However, critical concerns regarding developmental-stage mismatches and overstated breeding implications must be addressed.

Major Concerns
1. Temporal disconnect between RNA-seq and trait data
The 34 field-measured agronomic traits (e.g., grain weight/number) do not align temporally with RNA-seq data from 2-week seedlings. Gene networks governing reproductive traits are unlikely to be established at this early stage. The authors must:
- Provide biological justification for linking seedling transcriptomes to adult phenotypes
- Acknowledge limitations of this approach or present evidence for conserved regulatory signatures across development

2. Overextended breeding claims
Statements about direct breeding applications (e.g., "optimizing regulatory networks") lack empirical support given the study's focus on transcriptional mechanisms. While the pan-genome resource has breeding relevance, the authors should:
- Substantiate breeding claims with specific testable hypotheses
- Reframe discussions to distinguish demonstrated findings from speculative applications

Minor Concerns

Data & Methods
3. Clearly cite sources of reused agronomic trait data. For novel traits, detail field trials (locations, replicates, phenotyping protocols).
4. Clarify disease resistance scoring:
- Number of plants assessed per accession
- Handling of intra-accession variability
- Validation of the 0-4 scale (published standard or study-specific?)

Analytical Rigor
5. Reduce excessive abbreviations; define all non-standard acronyms at first use.
6. Specify whether the SNP matrix used in PEER analysis matches the eQTL detection matrix.
7. Justify expression thresholds (TPM > 0.5, SCC > 0.3) with empirical validation or field-standard references.

Presentation
8. Figures:
- Fig 2E: Replace overlapping lines with histograms/marginal plots.

- Fig 4G: Annotate data point counts.
- Fig 5F: State randomization iterations in figure/legend.
9. Text:
- Fix section header formatting in "Detection of eQTL".
- Tighten verbose phrasing (e.g., Introduction: "SNP diversity has been extensively characterized" vs original).

Reviewer #2

(Remarks to the Author)
This study by Zhang et al describes a large-scale gene expression analysis in bread wheat, utilizing diverse material, mainly landraces and cultivars from China and US.
The work presented contributes to a highly relevant topic, which is to better understand plasticity in gene expression patterns and their influence on important traits. Another aspect is the investigation of reference genome biases, and especially the influence of introgressions and alien material, on gene expression patterns.
While the authors generated an impressive set and amount of data, I have a few major concerns especially related to data analysis, and consequently interpretation of the results.

1.) I feel the chapter “The impact of single reference in RNA-seq reads mapping leads to underestimate expression for introgression” does not really add much novelty but mainly confirms what has been shown before by Coombes et al.

2.) The construction of the pan-gene atlas has major flaws:
The authors combine data (gene predictions) from highly heterogenous resources, genome assemblies (long and short read) and annotation pipelines with no attempt of harmonization or making the resources comparable. As a consequence, the reported variable and private gene portions will without doubt at least in part be a result of technical differences rather than biological differences. This also renders the downstream finding based on the pan-gene atlas at least in part unreliable. The simple selection of the longest gene per OG ignores substantial variation possible within some of the OGs.

3.) Selection of the wheat reference material: Is there a reason why the recently published genotypes from this wheat pan-genome where not used: https://doi.org/10.1038/s41586-024-08277-0? These data add valuable reference genome resources especially for the so-far under-represented material from Asia, which is an important part of the population genomics study in the second half of this manuscript.

4.) eQTL map: While the newly constructed pan-gene atlas is being used to determine gene expression levels, all the SNPs are based on the single reference Chinese Spring. While I expect a certain portion of the association to be robust nevertheless this approach contradicts somewhat the concept of reference aware resources propagated here. This should especially be true for genes and loci not being present in CS, either for technical (v1.1 predictions refer to the short read assembly version…especially the intergenic regions may be problematic) or biological reasons.
I do not really understand how “private genes” not present in CS could be associated with eQTLs based on CS SNPs… including this statement (line 301): “So, the syntenic chromosome location of the non-CS genes can be anchored by the strongest eQTL signal located at the Chinese Spring genome”.

5.) Line 347: I would term this analysis an integrative approach rather than “Multi-omics modeling”. No additional true omics data is introduced here to my understanding.

6.) This manuscript needs some substantial editing for language and also accuracy. Text for example in Line 659 and Line 712-713 is broken completely and that makes it hard to read and understand. There may even be something missing here.

7.) Figure 3a-c: how can the authors be sure that observed differences in % of expressed genes are not originating from different experimental setups and/or deviations in sampling time etc.? If I understand correctly, publicly available Rye expression data was used for this comparison from 10.1371/journal.pone.0288520. While as a genomic reference Rye Lo7 is being used, the Rye expression data in Krepski et al. was sampled from three Rye inbred lines. The results may be the same, but this approach is somewhat contradicting the call for suitable reference genomes made in this same study.

8.) Line 243: the authors write about “suppressed expression” in different places throughout that chapter, actually referring to a lower number of expressed orthologous genes, not reduced or elevated expression strength (if I understand correctly… how would you compare expression quantities across different experiments other than the experimental approach using qRT-PCR?). I think this is somewhat confusing and authors could make clear this is about number of expressed genes (where the issues of the pan-gene atlas described above may have an influence on the results). Later qRT-PCR results are being presented for selected example genes.

Reviewer #3

(Remarks to the Author)
This is an excellent piece of work that will have a significant impact on the field. My comments are worth considering, but they can largely be addressed by discussing limitations rather than requiring extensive re-analysis or refocusing of sections.

General Comments
The first three sections largely reiterate the arguments made by Coombes et al. (2024), demonstrating the impact of using a

single reference and the advantages of a pan-transcriptome reference. Since this argument has already been made, the authors should focus on why their new Panreference atlas is an improvement over existing approaches.

A key limitation of Coombes et al. (2024) is that they only use three out of the seven ancestral groups. This study extends the approach by incorporating progenitor species into an existing 10-plus pan-transcriptome reference. However, it is crucial to assess whether the addition of 24 references adequately represents the seven ancestral groups identified in Wat-seq. Simply increasing the number of references may raise computational costs without substantially reducing bias. An iterative approach to panel design, including rounds of design testing, would be beneficial. Additionally, benchmarking should include comparisons with the existing 10-plus pan-transcriptome reference, as benchmarking solely against Chinese Spring is no longer sufficient.

One indication of reference bias is the down-regulation of multiple genes within a locus. In the analysis of the diversity set, are any of these signatures still present?

Reference Bias in eQTL Mapping
In terms of eQTL mapping, reference bias affects two key areas:
1. RNA-seq reference bias, which this paper effectively addresses.
2. Reference bias in WGS read mapping, where WGS reads are mapped to the Chinese Spring reference to generate SNPs. This issue should be explicitly discussed.

TWAS Study
The TWAS approach helps mitigate some of the challenges associated with reference bias in eQTL mapping, as it enables the identification of non-Chinese Spring genes. This is perhaps the most innovative aspect of the paper. However, the integration of GWAS and TWAS is not entirely clear, particularly given the limitations of mapping WGS reads to a Chinese Spring reference. Further clarification on this integration would improve the manuscript.

Differential Gene Expression Analysis
The research question is well-defined: the authors aim to explore how different breeding programs and landraces have influenced differential gene regulation and networks. However, the results, methods, and figures are difficult to follow. Providing a clearer explanation of the differential expression analysis would enhance the readability of the subsequent network analysis. Using example genes—such as Vern 2 as a figure panel and its correlation between winter and spring lines—could improve clarity.

Data Availability
The current Panreference atlas is not readily usable. It is critical that the authors provide a FASTA file for the pan-atlas rather than just a list of gene IDs.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have addressed all my concerns. I have no further comments.

Reviewer #2

(Remarks to the Author)
The revised manuscript by Zhang et al. addresses many of my previous concerns. In particular, the generation and (re-) analysis of native rye expression data strengthens the author's claims, and I also appreciate the inclusion of material from the latest pan-genome and following re-analysis. This also further improves the value of the data generated here as a critical resource for wheat improvement.
I have a few points of concern following the author's responses:

1.) However, this section remains important because we go beyond confirming this bias by analyzing introgressed segments specific to our wheat panel, identifying expression underestimated in approximately 124 accessions with large introgressions. Additionally, we propose potential donor genome models from existing literature that could mitigate the bias. This section thus not only validates previous findings but extends them by offering practical solutions tailored to the unique structure of our panel.

I see and appreciate this aspect, but I do feel that the paragraph needs to be modified accordingly. As it stands now, it still mainly reads as a confirmation of existing knowledge, and does not transport the message of a "practical solution tailored to the panel" well.

2.) However, our work does not represent a conventional pan-genome construction. Instead, we aimed to build a pan-gene collection by directly utilizing available gene models from wheat genomes of different ploidy levels….The definitions of variable and private genes apply only within the context of our constructed atlas and are not intended as genome-wide

classifications.

I understand that, and appreciate that the authors are not making claims about (pan-) genome-wide gene classifications here. However, the problem remains that the genes classified as variable and private in this non-harmonized analysis are being used with these labels in downstream analyses, and conclusions are being drawn (also) based on those unreliable classifications. Or in other words, if a substantial portion of genes is classified as “private” genes as a result of gene prediction differences only, this would still have an effect on their use and interpretation in the pan-gene atlas presented here.

3.) To avoid redundancy in quantification, we selected the longest representative transcript from each orthogroup (OG), as gene expression quantification software often cannot resolve reads mapped to homologous regions with similar sequences. Retaining all homologs may lead to artificial read averaging, thereby compromising quantification accuracy. This strategy ensures more precise read assignment. For candidate gene analysis, we revisited all homologs within each OG to investigate sequence variation in detail.

I do understand the technical rational behind this approach with respect to read assignment. However, my point is that presumably in many cases, the longest transcript may not be a good representative of the orthologous group per se. This is especially true in a non-harmonized gene prediction dataset which is used in this study.
Additionally I am wondering about this: the pan-gene atlas consists of genes from a broad and diverse panel of wild, domesticated wheats, landraces, and even other triticeae genomes (? Line 205), and I am not clear how that affects the mapping accuracy of the also diverse transcriptome samples. The authors evaluated this on the basis of numbers of expressed genes, but this is not really indicative of mapping quality. For example, what happens if expression data from modern cultivars just does not map to the wild wheat representative gene model which was chosen from an OG just because it was the longest transcript? I do think that this should be evaluated more deeply, for example by looking at and comparing the mapping quality of all alignments within an orthologous group.

4.) We also fully acknowledge the limitation that SNPs called solely on the CS reference may not capture regulatory variants for genes absent in CS. In such cases, the corresponding eQTLs are necessarily categorized as trans-acting, although their true physical proximity may be closer in alternative genomes.

I fully support the use of CSv1.1 as back bone for SNP detection here, for the same reasons as the authors. My main concern is also not the assembly quality here, rather the cases termed as “trans-acting” here. So it can be expected that a portion of those in fact are false positively classified eQTLs? If that is the case, this limitation should be estimated, described and discussed, as many of those potential eQTLs may be the ones most interesting.

5.) in the context of TWAS and SMR analyses, we approximated the syntenic location of non-Chinese Spring genes by using their most significant eQTL peak mapped to the Chinese Spring genome. As shown in Fig. 4e, the strongest eQTL signals tend to cluster near the transcription start site, allowing this approach to serve as an approximate of the physical position.

Is there any reference for this approach demonstrating the overall validity? The fact that the signals cluster near transcription start sites does not appear to be a good indicator that indeed syntenic locations are being hit. I would expect that in many cases syntenic relationships are much more complex to resolve and the most significant eQTL peak hit may be just the next best random match.

Reviewer #3

(Remarks to the Author)
After reviewing this new version of the manuscript, I have no further comments to make. I am satisfied the authors have addressed our initial comments as far as could reasonably be expected.

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
This revised version of the manuscript by Zhang et al. addresses most of my previous concerns, although mostly not by new or updated analyses. In particular I appreciate that the authors now include both evaluation data/background information for critical analyses steps, and discuss obvious limitations with some of the approaches. These include:

Point 2: “Therefore, based on the OrthoFinder classification results, we have renamed
variable genes as assigned genes and private genes as unassigned genes.”

My intention here was not primarily to suggest changes in terminology (although these updated terms help to navigate the potential problems) but to either take measures to harmonise gene predictions used here (which I acknowledge to be a load of work) or to estimate and discuss potential errors introduced by the current strategy. As the authors lean towards option 2, I

do think this is a welcomed improvement, although the extend of downstream consequences of miss-classifications remain unclear to me (it is not about the effect of change of terminology on downstream analyses, but of genes wrongly labeled as e.g unassigned/private).

Point 3: “The reviewer's concern about alignment reliability is therefore intrinsically addressed by Salmon’s statistical framework, which has demonstrated superior accuracy over comparable tools in both simulated and real datasets.”

My concerns were not about the alignment reliability and the quantification tool used, but rather on what gene model (from what wheat line) was selected as the (main and only) template to map against. I do think this is a very critical step and the choice matters for both overall quantification and interpretation of the results obtained. The authors now present an evaluation demonstrating some of the effect of choosing the longest isoform per OG, and found “Among orthogroups whose longest transcript had TPM > 1, the longest transcript consistently showed significantly higher expression than its homologs within the same orthogroup”.
To me, this demonstrates that indeed the longest transcript captures a greater proportion of expression, but I do not agree that this necessarily means it serves as a better representative of the orthogroup (thinking of gene prediction artefacts etc… one could also claim that an OG representative close to the OG mean expression could serve as the "best"). However, since the authors now include information and discussion about this potential concern, readers will have all required information.

Point 4 and 5: I was indeed concerned that those signals may not represent true eQTLs at all when they are lacking a physical position in the reference genome (and not so much about a misclassification as trans-QTL). This is mainly because of the introgressed regions and translocation as also pointed out by the authors. I may not have understood the full methodology here, so I appreciate the added discussion especially on the limitations of placing the non-reference eQTLs on the genome.

Version 3:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
I appreciate the detailed responses and the inclusion of a discussion of the limitations in the latest revised manuscript version.
I have no further comments.

Point-by-point response to Reviewers

Zhang et al., Population-scale gene expression analysis reveals the contribution of expression diversity to the modern wheat improvement

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This study performs population RNA-seq on a diverse wheat panel with existing genome sequences. The authors present complex datasets effectively, though some sections are overly verbose. Key advances include using pan-genome resources for RNA-seq alignment (identifying >20,000 novel transcripts and improving alignment by 20%) and integrating GWAS, SMR, and TWAS to prioritize pan-genome candidates. However, critical concerns regarding developmental-stage mismatches and overstated breeding implications must be addressed.

Response: We sincerely thank the reviewer for the insightful comments and constructive suggestions. Below we address each concern in detail.

Major Concerns

1. Temporal disconnect between RNA-seq and trait data

The 34 field-measured agronomic traits (e.g., grain weight/number) do not align temporally with RNA-seq data from 2-week seedlings. Gene networks governing reproductive traits are unlikely to be established at this early stage. The authors must:

- Provide biological justification for linking seedling transcriptomes to adult phenotypes

- Acknowledge limitations of this approach or present evidence for conserved regulatory signatures across development

Response: We thank the reviewer for pointing out this important limitation. We have addressed it with the following clarifications: 1) We performed RNA-seq on 2-week-old seedlings from 328 wheat accessions, and integrated these transcriptomes with the published whole-genome resequencing data[1] to infer regulatory relationships. 2) Gene expression is inherently spatiotemporally specific, and transcriptomes from the seedling stage may not fully capture expression dynamics relevant to the 34 field traits, which are mostly measured during reproductive stages. However, modern wheat lines carry numerous introgressed segments (e.g., 1RS from rye, chromosomal fragments from *T. ponticum* and *T. timopheevii*, as well as from diploid and tetraploid relatives) that create presence/absence variation (PAV) at the expression level. Genes located in

these introgressions often show binary expression patterns, being present in carriers and absent in non-carriers, even in seedling-stage tissue, especially for highly specific genes. Thus, our expression-based associations largely reflect PAV-driven genetic regulation. 3) Our regulatory network analysis is based on seedling-stage expression data. We selected candidate genes involved in reproductive traits to illustrate how co-expression networks have changed from landraces to modern cultivars. 4) We now explicitly acknowledge the limitation in the manuscript (lines 582-585): “As the RNA-seq data used in this study were collected at the seedling stage, the detection of expression changes associated with developmental traits was limited. However, introgressed genes often exhibit presence/absence variation in expression, which was effectively captured.” Future studies will aim to profile multiple tissues and developmental stages to complement and extend these findings.

2. Overextended breeding claims

Statements about direct breeding applications (e.g., "optimizing regulatory networks") lack empirical support given the study's focus on transcriptional mechanisms. While the pan-genome resource has breeding relevance, the authors should:

- Substantiate breeding claims with specific testable hypotheses

- Reframe discussions to distinguish demonstrated findings from speculative applications

Response: We thank the reviewer for the valuable comment regarding the interpretation of regulatory network changes. Our investigation of differences between landraces and cultivars in gene regulatory networks was based on the number of co-expression pairs between eGenes and active-eQTGs. We found that both modern Chinese cultivars and U.S. cultivars exhibited significantly more co-expression pairs ($|SCC| > 0.3$, p-value < 0.05) involving the same sets of eGenes and active-eQTGs compared to landraces. This suggests that the extent of co-expression relationships is reduced in landraces. However, we agree that this observation alone cannot confirm whether regulatory networks have been optimized during breeding. It is more accurate to state that the co-expression network structure has changed.

To address this concern, we revised two relevant statements in the manuscript: 1) The sentence: “Therefore, future breeding strategies should not only focus on the selection of specific genes but also on optimizing and utilizing the overall regulatory networks.” has been revised to (lines 617-621): “In addition to individual expression changes, breeding has altered gene co-expression and regulatory network architecture. The number of co-expression pairs involving active-eQTGs and eGenes was significantly lower in landraces than in cultivars, suggesting that modern breeding has increased network complexity. Future breeding strategies may benefit from considering not only gene-specific selection but also optimization of broader regulatory networks.” 2) The sentence: “These findings also indicate that modern breeding has reshaped co-expression regulatory networks,

potentially enhancing the ability to cope with adverse environmental conditions and improving population stability." has been revised to (lines 513-514): "Taken together, these findings suggest that modern breeding has substantially rewired co-expression and regulatory networks." These changes aim to avoid overinterpretation and better reflect the evidence provided in our study.

Minor Concerns

Data & Methods

3. Clearly cite sources of reused agronomic trait data. For novel traits, detail field trials (locations, replicates, phenotyping protocols).

Response: We thank the reviewer for the helpful suggestion. We now clearly distinguish between 20 previously published agronomic traits and 14 newly collected ones in the Methods section (lines 628-649), and provide comprehensive information on field locations, years, planting methods, and phenotyping protocols.

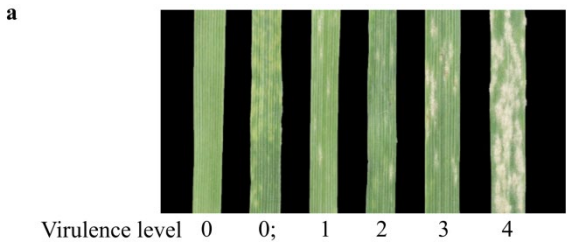
4. Clarify disease resistance scoring:

- Number of plants assessed per accession

- Handling of intra-accession variability

- Validation of the 0-4 scale (published standard or study-specific?)

Response: We appreciate the reviewer's comment. For powdery mildew scoring, 3-4 seedlings were evaluated per accession, and the average value was used. The 0-4 scoring standard was based on a published study[2], where infection types 0-2 were considered as resistant and 3-4 as susceptible. Representative images of the phenotypic standard are provided in Supplementary Fig. 11a.



Analytical Rigor

5. Reduce excessive abbreviations; define all non-standard acronyms at first use.

Response: Thank you for this valuable suggestion. We have replaced nearly all abbreviations with their full forms—for example, "CS" is now written as "Chinese Spring", "KN9204" as "Kenong 9204", "AK58" as "Aikang 58", and "HC" as "high-confidence". We retain only the most frequently

used terms such as “landraces (LRs)”, “modern Chinese cultivars (MCCs)”, and “modern U.S. cultivars (USMCs)”, which are defined at first use.

6. Specify whether the SNP matrix used in PEER analysis matches the eQTL detection matrix.

Response: We thank the reviewer for the suggestion. This is now clarified in the Methods section (lines 757-758), where we specify that the SNP matrix used in PEER analysis is consistent with that used for eQTL detection.

7. Justify expression thresholds (TPM > 0.5, SCC > 0.3) with empirical validation or field-standard references.

Response: Thank you for the reviewer’s insightful comments. The threshold of TPM > 0.5 for defining gene expression was adopted in our study based on two previous transcriptomic studies in wheat: *Ramírez-González RH et al.* (Science, 2018) and *He et al.* (Nature Communications, 2022), both of which employed this cutoff[3, 4]. We have now cited these references accordingly in the revised manuscript.

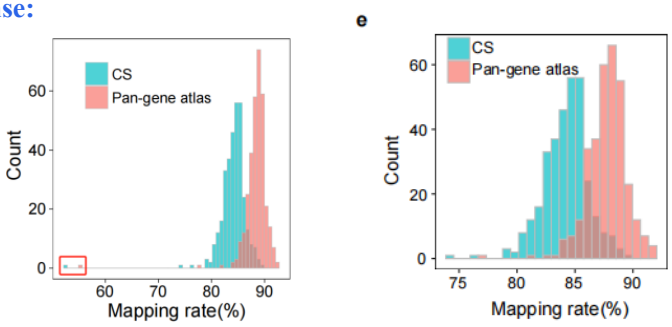
To define expression correlation, we considered gene pairs with |SCC| > 0.3 and $p < 0.05$ as significantly correlated. This threshold is supported by *Mukaka MM* (Malawi Medical Journal, 2012), who suggested that SCC values below 0.2 indicate negligible correlation, while values above 0.3 indicate positive correlation[5]. We have now cited these references accordingly in the revised manuscript.

Presentation

8. Figures:

- Fig 2E: Replace overlapping lines with histograms/marginal plots.

Response:



We thank the reviewer for this detailed suggestion. We have replaced the overlapping line plot with a frequency histogram, which better visualizes the distribution of alignment rates. In doing so, we noticed that one accession (1699B) had an unusually low mapping rate. Analysis of unmapped reads revealed contamination by Barley Yellow Mosaic Virus, so this sample was removed. Fig. 2e now

shows the alignment rates for the remaining 327 samples used in downstream analysis.

- Fig 4G: Annotate data point counts.

Response: Thank you. The figure has been revised to include the number of data points.

- Fig 5F: State randomization iterations in figure/legend.

Response: Thank you. We have added the number of randomization iterations to the legend of Fig. 5f.

9. Text:

- Fix section header formatting in "Detection of eQTL".

Response: We appreciate the reviewer's careful reading. The header formatting has been corrected.

- Tighten verbose phrasing (e.g., Introduction: "SNP diversity has been extensively characterized" vs original).

Response: Thank you for this writing suggestion. We have carefully revised the manuscript to eliminate verbose phrasing and improve clarity and logical flow throughout.

Reviewer #2 (Remarks to the Author):

This study by Zhang et al describes a large-scale gene expression analysis in bread wheat, utilizing diverse material, mainly landraces and cultivars from China and US.

The work presented contributes to a highly relevant topic, which is to better understand plasticity in gene expression patterns and their influence on important traits. Another aspect is the investigation of reference genome biases, and especially the influence of introgressions and alien material, on gene expression patterns.

While the authors generated an impressive set and amount of data, I have a few major concerns especially related to data analysis, and consequently interpretation of the results.

Response: We sincerely thank the reviewer for the thoughtful comments and constructive suggestions, which have helped us to improve the overall quality and clarity of our manuscript. Below, we provide detailed responses to each point.

1.) I feel the chapter "The impact of single reference in RNA-seq reads mapping leads to underestimate expression for introgression" does not really add much novelty but mainly confirms what has been shown before by Coombes et al.

Response: We thank the reviewer for raising this point. We agree that the observation of expression bias caused by using a single reference genome is consistent with the findings of *Coombes et al.* (2021), as shown in Figs. 1a, b and Supplementary Fig. 2. However, this section remains important because we go beyond confirming this bias by analyzing introgressed segments specific to our wheat panel, identifying expression underestimated in approximately 124 accessions with large introgressions. Additionally, we propose potential donor genome models from existing literature that could mitigate the bias. This section thus not only validates previous findings but extends them by offering practical solutions tailored to the unique structure of our panel.

2.) The construction of the pan-gene atlas has major flaws:

The authors combine data (gene predictions) from highly heterogenous resources, genome assemblies (long and short read) and annotation pipelines with no attempt of harmonization or making the resources comparable. As a consequence, the reported variable and private gene portions will without doubt at least in part be a result of technical differences rather than biological differences. This also renders the downstream finding based on the pan-gene atlas at least in part unreliable.

The simple selection of the longest gene per OG ignores substantial variation possible within some of the OGs.

Response: We sincerely thank the reviewer for the constructive suggestions regarding our methodology. We fully acknowledge that in most pan-genome studies, genome annotations are harmonized using a unified pipeline to minimize technical variation. Therefore, the observed proportions of variable and private genes in our study may partly result from differences in annotation strategies, a concern we greatly appreciate. However, our work does not represent a conventional pan-genome construction. Instead, we aimed to build a pan-gene collection by directly utilizing available gene models from wheat genomes of different ploidy levels.

Because Chinese Spring is the most widely used reference genome in the community, all genes from this genome were retained. To address the limitations of relying on a single reference, additional genes from other genomes that are absent from or lack homologs in Chinese Spring were also included. The definitions of variable and private genes apply only within the context of our constructed atlas and are not intended as genome-wide classifications. We clarified it in lines 530-536 of the revised version: “In our case, the detection of private and variable genes is constrained by the completeness and consistency of available gene annotations, which may vary due to differences in annotation pipelines across studies. Nevertheless, the goal of this work was to establish a usable reference model. Importantly, the definitions of variable and private genes in this work are specific to the gene models within our pan-gene atlas, rather than the entire genome space, aiming to classify the atlas itself rather than to comprehensively categorize all genome-encoded genes.”

Our objectives were threefold: 1) to provide a lightweight, easily reproducible method to construct a pan-gene set from published models; 2) to maximize the number of genes whose expression could be quantified, thereby improving read alignment rates compared to the traditional use of a single genome as reference; and 3) to identify more candidate genes absent from the Chinese Spring reference thus provide an evaluation of the contribution of expression diversity to the agronomic improvement.

To avoid redundancy in quantification, we selected the longest representative transcript from each orthogroup (OG), as gene expression quantification software often cannot resolve reads mapped to homologous regions with similar sequences. Retaining all homologs may lead to artificial read averaging, thereby compromising quantification accuracy. This strategy ensures more precise read assignment. For candidate gene analysis, we revisited all homologs within each OG to investigate sequence variation in detail.

3.) Selection of the wheat reference material: Is there a reason why the recently published genotypes from this wheat pan-genome where not used: https://doi.org/10.1038/s41586-024-08277-0? These data add valuable reference genome resources especially for the so-far under-represented material from Asia, which is an important part of the population genomics study in the second half of this manuscript.

Response: We thank the reviewer for this excellent suggestion. The newly published wheat pan-genome[6] indeed provides valuable resources, particularly for underrepresented Asian accessions. At the time this manuscript was initially completed, the pan-genome had not yet been published, and therefore was not included in our original analyses.

To improve the quality and scope of our work, we have now incorporated the gene models from this recent release, along with additional annotations from other wheat genomes with different ploidy levels. Specifically, we expanded our reference set from 24 to 44 genomes - we used only genome assemblies with the gene models supported by transcriptomic evidence. Because the population RNA-seq were processed using this updated reference, all analyses following Fig. 2 were re-done. All corresponding figures and results were re-generated in our revised version. While the core conclusions of the study remain unchanged, the updated version provides a more complete and informative perspective. All modifications have been clearly track-marked in the revised manuscript. Here we list some of the important changes:

1) The number of genes in the pan-gene atlas increased from 201,223 to 232,736.
2) The number of genes expressed at the population level increased from 77,447 to 80,128.
3) The number of candidate genes identified increased from 231 (including 54 non-Chinese Spring genes) to 299 (including 74 non-Chinese Spring genes).
4) The number of candidate genes validated using the Kenong 9204 mutant library increased from 58 to 86.

Therefore, expanding the reference set to include genomes suggested by this reviewer enabled the detection of more expressed genes and better evaluate the contribution of introgression to the agronomic traits.

4.) eQTL map: While the newly constructed pan-gene atlas is being used to determine gene expression levels, all the SNPs are based on the single reference Chinese Spring. While I expect a certain portion of the association to be robust nevertheless this approach contradicts somewhat the concept of reference aware resources propagated here. This should especially be true for genes and loci not being present in CS, either for technical (v1.1 predictions refer to the short read assembly version…especially the intergenic regions may be problematic) or biological reasons.

I do not really understand how "private genes" not present in CS could be associated with eQTLs based on CS SNPs…including this statement (line 301): "So, the syntenic chromosome location of the non-CS genes can be anchored by the strongest eQTL signal located at the Chinese Spring genome".

Response: We sincerely thank the reviewer for this important and insightful comment. To investigate the regulatory mechanisms of gene expression across a genetically diverse wheat population, we constructed a pan-gene atlas using a set of published genome assemblies, enabling us to capture a broader range of gene expression, including genes absent from the Chinese Spring (CS) reference. For genotype data, we utilized whole-genome-sequenced genotype data based on the CS v1.1 reference genome, which was generated by *Niu et al.*[1].

We agree with the reviewer that CS v1.1, which is based on short-read sequencing, may contain assembly errors, particularly in intergenic regions. We chose this version because it remains the most widely used reference in the wheat research community, and it provides compatibility with existing large-scale genotyping resources such as the 1000 Wheat Exomes Project[7] and the Watkins panel[8]. Moreover, most merged eQTLs are located within genic regions, with 84.8% of genes having at least one eQTL detected within the gene (Figs. 4c, d), which suggests that the potential assembly errors located in highly repeated regions in Chinese Spring v1.1 do not likely affect the main features of our eQTL map.

We also fully acknowledge the limitation that SNPs called solely on the CS reference may not capture regulatory variants for genes absent in CS. In such cases, the corresponding eQTLs are necessarily categorized as *trans*-acting, although their true physical proximity may be closer in alternative genomes. Due to current technical and computational limitations, we have not yet developed a pan-genome-based variant map. In future work, we aim to construct a wheat tribe pan-genome and perform reference-free variant calling—including SNPs, structural variants, copy number variations, and mobile element insertions—to better capture the full regulatory landscape. We have now discussed this limitation explicitly in lines 596-604 of the revised manuscript.

Regarding the reviewer's question on the association between Chinese Spring-based SNPs and private genes not present in Chinese Spring: in the context of TWAS and SMR analyses, we approximated the syntenic location of non-Chinese Spring genes by using their most significant eQTL peak mapped to the Chinese Spring genome. As shown in Fig. 4e, the strongest eQTL signals tend to cluster near the transcription start site, allowing this approach to serve as an approximate of the physical position. Since this is a methodological assumption and not a main conclusion, the original sentence "So, the syntenic chromosome location of the non-Chinese Spring genes can be anchored by the strongest eQTL signal located at the Chinese Spring genome"' has been moved from the main text to the Methods section (lines 798-800):"For non-Chinese Spring genes, the genomic positions of their most significant eQTL signals in the Chinese Spring reference genome were used as proxies", to avoid confuse the readers.

5.) Line 347: I would term this analysis an integrative approach rather than "Multi-omics modeling". No additional true omics data is introduced here to my understanding.

Response: We thank the reviewer for this observation. Our use of the term "multi-omics modeling" refers to the integration of genome-wide SNP data (genomics), gene expression levels (transcriptomics), and phenotypic traits. According to accepted definitions[9], integration of two or more omics layers qualifies as a multi-omics approach. Therefore, we consider our usage of the term to be reasonable in the context of this study. However, given that our study primarily emphasizes population-scale transcriptomic data, we have revised the term "multi-omics modeling" to "integrative modeling" to improve clarity and better reflect the focus of the study.

6.) This manuscript needs some substantial editing for language and also accuracy. Text for example in Line 659 and Line 712-713 is broken completely and that makes it hard to read and understand. There may even be something missing here.

Response: We thank the reviewer for carefully pointing this out. We have thoroughly revised the manuscript to correct unclear phrasing and incomplete sentences, including those in lines 659 and 712-713. The revised version is now more concise and readable.

7.) Figure 3a-c: how can the authors be sure that observed differences in % of expressed genes are not originating from different experimental setups and/or deviations in sampling time etc.? If I understand correctly, publicly available Rye expression data was used for this comparison from 10.1371/journal.pone.0288520. While as a genomic reference Rye Lo7 is being used, the Rye expression data in Krepski et al. was sampled from three Rye inbred lines. The results may be the same, but this approach is somewhat contradicting the call for suitable reference genomes made in this same study.

Response: We sincerely thank the reviewer for this helpful suggestion. In response to the concern regarding the comparability of public rye expression data, we generated new RNA-seq data from

six rye varieties at the seedling stage. The sequencing data have been submitted to the National Genomics Data Center (GSA: CRA025944). We reanalyzed this section using the new dataset, and the results are presented in revised Fig. 3a-d. We consistently observed that the proportion of expressed genes in rye introgression lines remains lower than in these six rye varieties, thereby confirming and strengthening our original conclusion.

8.) Line 243: the authors write about “suppressed expression” in different places throughout that chapter, actually referring to a lower number of expressed orthologous genes, not reduced or elevated expression strength (if I understand correctly…how would you compare expression quantities across different experiments other than the experimental approach using qRT-PCR?). I think this is somewhat confusing and authors could make clear this is about number of expressed genes (where the issues of the pan-gene atlas described above may have an influence on the results). Later qRT-PCR results are being presented for selected example genes.

Response: We sincerely thank the reviewer for this insightful comment. Your interpretation is correct, and we have revised the text to more clearly reflect our results. (1) The term "suppressed expression" in this context refers to a reduced number of expressed homologous genes from introgressed segments, rather than a decrease in expression intensity. (2) For differential expression analyses across experiments, we applied robust quantile normalization using the "normalize.quantiles.robust" function. This approach ranks gene expression within each sample, enabling comparisons based on expression rank rather than absolute values.

In our rye analysis, we observed upregulation of some resistance-related genes and downregulation of genes involved in primary metabolism. To validate this observation, we selected three resistance-related genes (*SECCE1Rv1G0002890*, *SECCE1Rv1G0003770*, *SECCE1Rv1G0008680*) and three basic metabolic genes for genomic DNA amplification and qRT-PCR validation. These resistance genes were present in introgression lines but appeared in only a subset of rye accessions. This suggests that breeders may have preferentially retained desirable rye traits in introgression lines. The apparent upregulation observed in transcriptomic analyses likely reflects the broader presence of these genes across introgression lines, rather than a genuine increase in their expression levels. These points have now been clarified in lines 273-278 and are further discussed in lines 559-566 of the revised manuscript.

Reviewer #3 (Remarks to the Author):

This is an excellent piece of work that will have a significant impact on the field. My comments are worth considering, but they can largely be addressed by discussing limitations rather than requiring extensive re-analysis or refocusing of sections.

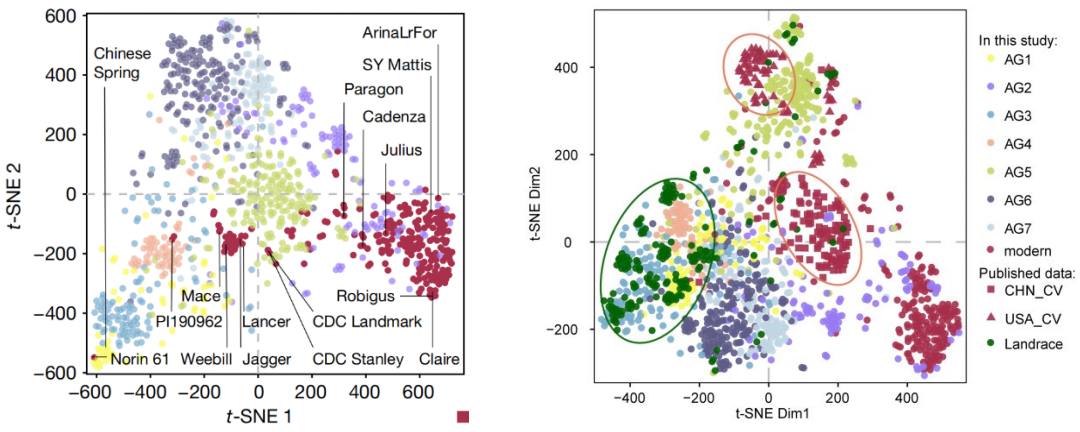
Response: We sincerely thank the reviewer for the positive evaluation and thoughtful comments, which have greatly helped us improve the clarity and completeness of our manuscript. We provide point-by-point responses below.

General Comments

The first three sections largely reiterate the arguments made by Coombes et al. (2024), demonstrating the impact of using a single reference and the advantages of a pan-transcriptome reference. Since this argument has already been made, the authors should focus on why their new Panreference atlas is an improvement over existing approaches.

A key limitation of Coombes et al. (2024) is that they only use three out of the seven ancestral groups. This study extends the approach by incorporating progenitor species into an existing 10-plus pan-transcriptome reference. However, it is crucial to assess whether the addition of 24 references adequately represents the seven ancestral groups identified in Wat-seq. Simply increasing the number of references may raise computational costs without substantially reducing bias. An iterative approach to panel design, including rounds of design testing, would be beneficial. Additionally, benchmarking should include comparisons with the existing 10-plus pan-transcriptome reference, as benchmarking solely against Chinese Spring is no longer sufficient.

Response:



(Cheng et al., 2024)

We thank the reviewer for raising this important point. The ancestral group of our wheat panel has already been assessed by in *Cheng et al.*[8], which used genotype data from *Niu et al.*[1]. In their study, CHN_CV corresponds to the MCC varieties in our panel, USA_CV to our USMC varieties, and Landraces to the landraces included in our study. As shown in their figure, MCC varieties overlap with ancestral group AG2, USMC varieties with AG5, and landraces mainly with AG1, AG3, and AG4, with a small fraction distributed across other groups. Therefore, our panel of 328 wheat accessions broadly represents five out of the seven ancestral groups (AG1-AG5). Among the 44

reference genomes used in the revised version, several correspond to key ancestral groups relevant to our panel. For example, Chinese Spring and Norin61 overlap with AG1, supporting alignment for landraces; Mace, Lancer, Jagger, CDC Landmark, and CDC Stanley align with AG5, facilitating analysis of USMC lines; and ArinaLrFor, SY Mattis, and Julius align with AG2, representing MCC lines. Thus, the inclusion of these diverse genomes significantly improves alignment accuracy across our panel by better representing ancestral variation. In addition, integration of diploid and tetraploid progenitors involved in wheat evolution further strengthens the capacity of our pan-gene atlas to capture genetic variation across wheat germplasm, including lineage-specific and introgressed alleles that are often overlooked in modern breeding. This issue has been discussed in detail in the revised manuscript (lines 540-556).

We acknowledge that *Coombes et al.* (2024) already demonstrated the benefits of a pan-transcriptome reference. However, our expanded pan-reference improves upon existing approaches by incorporating additional genomes that increase representation across multiple ancestral groups, thereby reducing reference bias more effectively in a genetically diverse panel. Given that most transcriptomic studies in wheat have relied solely on the Chinese Spring reference, we included benchmark analyses using Chinese Spring for comparison. Importantly, our analysis revealed that approximately 24,446 (30.5%) of the expressed genes in the population are derived from genomes beyond Chinese Spring and the 10+ genomes, underscoring the necessity of constructing a more comprehensive pan-gene atlas.

One indication of reference bias is the down-regulation of multiple genes within a locus. In the analysis of the diversity set, are any of these signatures still present?

Response: We thank the reviewer for this important question. Reference bias can indeed result in the underestimation of gene expression levels. In our analysis, we found that such signatures were evident when using only the Chinese Spring reference genome. For instance, as shown in Fig. 1a, in 1RS.1BL introgression lines, expression levels on chromosome 1B were reduced, reflecting misalignment of rye-derived transcripts. However, after incorporating the rye genome into the pan-gene reference (Fig. 1d), a greater number of expressed rye genes were detected, and fewer genes were spuriously mapped to the 1B chromosome of Chinese Spring. This reallocation corrects the expression underestimation caused by reference bias. Although the number of genes detected within a locus may change more substantially across genotypes, these differences reflect improved assignment of expression to the correct genomic origin, thereby mitigating reference bias in the diversity set. These results demonstrate that the inclusion of multiple donor genomes in the pan-reference framework effectively reduces reference bias and improves the accuracy of gene expression quantification across diverse genotypes.

Reference Bias in eQTL Mapping

In terms of eQTL mapping, reference bias affects two key areas:

1. RNA-seq reference bias, which this paper effectively addresses.

2. Reference bias in WGS read mapping, where WGS reads are mapped to the Chinese Spring reference to generate SNPs. This issue should be explicitly discussed.

Response: We thank the reviewer for their valuable comment. For genotype data, we utilized whole-genome-sequenced genotype data based on the CS v1.1 reference genome, which was generated by *Niu et al.*[1]. We acknowledge that reference bias still exists, particularly in regions not represented by the Chinese Spring genome. Due to the current limit of computational resource, SNPs in this study were based on a single reference. In future work, we aim to construct a wheat pan-genome and perform SNPs, structural variants, copy number variations, and mobile element insertions discovery using a multi-genome framework to better capture the full spectrum of genetic variation. We have added a discussion of this issue in lines 596-604 of the revised manuscript.

TWAS Study

The TWAS approach helps mitigate some of the challenges associated with reference bias in eQTL mapping, as it enables the identification of non-Chinese Spring genes. This is perhaps the most innovative aspect of the paper. However, the integration of GWAS and TWAS is not entirely clear, particularly given the limitations of mapping WGS reads to a Chinese Spring reference. Further clarification on this integration would improve the manuscript.

Response: We thank the reviewer for this insightful comment. We have added a more detailed explanation of the integration of GWAS and TWAS in the Methods section (lines: 791-808). Specifically, while GWAS can only identify candidate genes located within the Chinese Spring genome, TWAS first estimates expression weights for all genes, including both Chinese Spring and non-Chinese Spring genes, based on eQTL data. By integrating expression weights and SNP-phenotype associations from GWAS, TWAS enables the identification of regulatory links between SNPs and the expression of non-Chinese Spring genes that may contribute to phenotypic variation. Thus, SNPs serve as a bridge connecting GWAS and TWAS to detect associations beyond the reference genome. The TWAS analysis in this study relied on GWAS summary statistics, and the detailed procedures follow the official TWAS pipeline available at http://gusevlab.org/projects/fusion/[10].

The advantages of using TWAS and SMR to identify non-Chinese Spring candidate genes are further discussed in lines 577-582: “Because GWAS relies solely on SNP information present in the Chinese Spring genome to associate with phenotypes, it can only identify candidate genes located within the Chinese Spring genome. In contrast, by using TWAS, SMR, and correlation analyses between gene expression and phenotypes, we were able to identify candidate genes not present in the Chinese Spring genome. The use of transcriptomic data enabled us to identify a greater number of non-Chinese Spring candidate genes”.

Differential Gene Expression Analysis

The research question is well-defined: the authors aim to explore how different breeding programs and landraces have influenced differential gene regulation and networks. However, the results, methods, and figures are difficult to follow. Providing a clearer explanation of the differential expression analysis would enhance the readability of the subsequent network analysis. Using example genes—such as Vern 2 as a figure panel and its correlation between winter and spring lines—could improve clarity.

Response: We sincerely thank the reviewer for this valuable suggestion. To improve the clarity and logic of the section on differentially expressed genes, we have revised the content of the "Differentially expressed genes and modules between different sub-populations" section in the revised version to enhance readability and interpretation. In addition, we analyzed the expression correlation between *VRN1*, *VRN2*, and *VRN3* and the growth habits (spring vs. winter wheat), as shown in Supplementary Fig. 15. Together with Supplementary Figs. 16 and 17, these results provide a clearer understanding of the differential expression patterns of *VRN1*, *VRN2*, and *VRN3*.

Data Availability

The current Panreference atlas is not readily usable. It is critical that the authors provide a FASTA file for the pan-atlas rather than just a list of gene IDs.

Response: We sincerely thank the reviewer for this suggestion. We have now provided the FASTA file and made it available on Figshare database (https://figshare.com/s/ee89a74381fea08c4c4e).

Reference

1. Niu J, *et al.* Whole-genome sequencing of diverse wheat accessions uncovers genetic changes during modern breeding in China and the United States. *Plant Cell* **35**, 4199-4216 (2023).
2. Xie J, *et al.* A biGWAS strategy reveals the genetic architecture of the interaction between wheat and *Blumeria graminis* f. sp. *tritici*[J]. *bioRxiv*, 2025.04.09.647224 (2025)
3. He F, *et al.* Genomic variants affecting homoeologous gene expression dosage contribute to agronomic trait variation in allopolyploid wheat. *Nat. Commun.* **13**, 826 (2022).
4. Ramírez-González RH, *et al.* The transcriptional landscape of polyploid wheat. *Science* **361**, (2018).
5. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med. J.* **24**, 69-71 (2012).
6. Jiao C, *et al.* Pan-genome bridges wheat structural variations with habitat and breeding. *Nature* **637**, 384-393 (2025).

7. He F, *et al.* Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nat. Genet.* **51**, 896-904 (2019).
8. Cheng S, *et al.* Harnessing landrace diversity empowers wheat breeding. *Nature* **632**, 823-831 (2024).
9. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol.* **18**, 83 (2017).
10. Gusev A, *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245-252 (2016).

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have addressed all my concerns. I have no further comments.

Response: We sincerely thank the reviewer for taking the time to review our revised manuscript and for the positive feedback.

Reviewer #2 (Remarks to the Author):

The revised manuscript by Zhang et al. addresses many of my previous concerns. In particular, the generation and (re-) analysis of native rye expression data strengthens the author's claims, and I also appreciate the inclusion of material from the latest pan-genome and following re-analysis. This also further improves the value of the data generated here as a critical resource for wheat improvement.
I have a few points of concern following the author's responses:

1.) However, this section remains important because we go beyond confirming this bias by analyzing introgressed segments specific to our wheat panel, identifying expression underestimated in approximately 124 accessions with large introgressions. Additionally, we propose potential donor genome models from existing literature that could mitigate the bias. This section thus not only validates previous findings but extends them by offering practical solutions tailored to the unique structure of our panel.

I see and appreciate this aspect, but I do feel that the paragraph needs to be modified accordingly. As it stands now, it still mainly reads as a confirmation of existing knowledge, and does not transport the message of a "practical solution tailored to the panel" well.

Response: We sincerely thank the reviewer for the valuable comment, which helped us clarify the intention of our analysis. In the revised manuscript, we have rewritten the relevant section in the Results (lines 182-198) to better describe the corrective effect of incorporating potential donor genomes on gene expression quantification. The revised text now reads as follows: "To accurately quantify gene expression in introgression regions, we merged the donor genome with the Chinese Spring reference genome (see Methods) and re-analyzed gene expression for all wheat samples. We found that when using only the Chinese Spring reference genome, the average number of expressed genes detected in the introgression lines within the corresponding translocated chromosome segments (chr1A: 0-214.4 Mb, chr1B: 0-239.3 Mb, chr2D: 570.1-619.2 Mb, chr3D: 500-615.5 Mb, chr5B: 497.2-537 Mb) was 14, 12, 39, 26 and 27 per 10 Mb window, respectively.

In contrast, when using the merged genome as reference, the average number of expressed genes mapped to these Chinese Spring segments decreased to 4, 4, 22, 12 and 9, while the corresponding donor-derived regions Rye (chr1R: 0-280 Mb), Renan (chr2D: 570-635 Mb), *Th. elongatum* (chr3E: 500-676.9 Mb), and *T. timopheevii* (chr5G: 435-485 Mb) showed higher average number of expressed genes: 15, 26, 19 and 29 per 10 Mb window. These results demonstrate that using the merged genome enables a substantial number of mis-mapped transcripts to be correctly assigned to their donor genome origins, effectively recovering actively expressed genes in the introgressed regions across each 10 Mb window (Fig. 1d and Supplementary Note 1). Thus, the reference bias can be severe when measuring expression in introgression using a single reference genome. Incorporating gene models from donor species is essential to accurately estimate the expression levels in introgressed regions."

2.) However, our work does not represent a conventional pan-genome construction. Instead, we aimed to build a pan-gene collection by directly utilizing available gene models from wheat genomes of different ploidy levels….The definitions of variable and private genes apply only within the context of our constructed atlas and are not intended as genome-wide classifications.

I understand that, and appreciate that the authors are not making claims about (pan-) genome-wide gene classifications here. However, the problem remains that the genes classified as variable and private in this non-harmonized analysis are being used with these labels in downstream analyses, and conclusions are being drawn (also) based on those unreliable classifications. Or in other words, if a substantial portion of genes is classified as "private" genes as a result of gene prediction differences only, this would still have an effect on their use and interpretation in the pan-gene atlas presented here.

Response: We thank the reviewer for the insightful and rigorous comment. Our pan-gene atlas was categorized using Orthofinder. Initially, we classified genes based on the Orthofinder output files *Orthogroups.tsv* and *Orthogroups_UnassignedGenes.tsv*, labeling them as variable genes and private genes, respectively. This differs from the classical pan-genome classification approach. As we also discussed in the revised manuscript (lines 558-563), "In our case, the detection of unassigned and assigned genes is constrained by the completeness and consistency of available gene annotations, which may vary due to differences in annotation pipelines across studies. In addition, the assignment of genes into orthogroups by OrthoFinder may be affected by sequence similarity thresholds, clustering parameters, and the complex evolutionary history of gene families, which could further influence the number of unassigned genes and their genomic distribution." and indeed, defining genes in *Orthogroups_UnassignedGenes.tsv* as private genes may cause confusion for readers. Therefore, based on the OrthoFinder classification results, we have renamed variable genes as assigned genes and private genes as unassigned genes.

Accordingly, we have updated all relevant sections of the manuscript, including figures and tables, to use the terms “assigned” and “unassigned” genes rather than “variable” and “private” genes. Importantly, this change in terminology does not affect our downstream analyses or the study’s conclusions, as the underlying classification method remains unchanged.

3.) To avoid redundancy in quantification, we selected the longest representative transcript from each orthogroup (OG), as gene expression quantification software often cannot resolve reads mapped to homologous regions with similar sequences. Retaining all homologs may lead to artificial read averaging, thereby compromising quantification accuracy. This strategy ensures more precise read assignment. For candidate gene analysis, we revisited all homologs within each OG to investigate sequence variation in detail.

I do understand the technical rational behind this approach with respect to read assignment. However, my point is that presumably in many cases, the longest transcript may not be a good representative of the orthologous group per se. This is especially true in a non-harmonized gene prediction dataset which is used in this study.

Additionally I am wondering about this: the pan-gene atlas consists of genes from a broad and diverse panel of wild, domesticated wheats, landraces, and even other triticeae genomes (? Line 205), and I am not clear how that affects the mapping accuracy of the also diverse transcriptome samples. The authors evaluated this on the basis of numbers of expressed genes, but this is not really indicative of mapping quality. For example, what happens if expression data from modern cultivars just does not map to the wild wheat representative gene model which was chosen from an OG just because it was the longest transcript? I do think that this should be evaluated more deeply, for example by looking at and comparing the mapping quality of all alignments within an orthologous group.

Response: We sincerely thank the reviewer for this thoughtful suggestion. We quantified gene expression using Salmon, a pseudo-alignment tool that does not output traditional per-read mapping quality scores (such as MAPQ from aligners like Bowtie2). Instead, Salmon integrates mapping confidence into expression estimates through a conditional probability model, rich equivalence classes, and probabilistic abundance estimation. In this framework, high expression levels primarily result from fragments with high-confidence mappings, whereas the influence of lower-quality mappings is effectively down-weighted by the model. The reviewer's concern about alignment reliability is therefore intrinsically addressed by Salmon’s statistical framework, which has demonstrated superior accuracy over comparable tools in both simulated and real datasets[1]. Consequently, the expressed genes identified in our study are expected to be of high quality.

To further assess whether choosing the longest transcript within each orthogroup is appropriate—

especially when the longest transcript originates from wild wheat and its relevance to modern cultivars might be questioned—we constructed a redundant pan-gene set by merging all genes from the 44 wheat genomes. Using two Chinese cultivars and two U.S. cultivars, we quantified expression levels across all genes. We then focused on orthogroups where the longest representative transcript was derived from diploid or tetraploid species. Among orthogroups whose longest transcript had TPM > 1, the longest transcript consistently showed significantly higher expression than its homologs within the same orthogroup (Wilcoxon test, p-value < 2.2×10^{-16}). In contrast, for orthogroups where the longest transcript had TPM $\leqslant$ 1, there was no significant difference between the longest transcript and other transcripts in the group (Supplementary Fig. 23). These results suggest that the longest transcript tends to capture a greater proportion of expression and thus serves as a better representative of the orthogroup.

Furthermore, we compared the expression of the longest transcripts from the non-redundant pan-gene atlas constructed in this study to those from the redundant pan-gene set. The longest transcripts in the non-redundant atlas generally exhibited slightly higher expression, which could be because reads that would otherwise be distributed across multiple homologous transcripts in the redundant set instead consolidate onto a single representative transcript (Supplementary Fig. 23). Overall, these findings support the choice of the longest transcript as a practical and representative measure of expression abundance within orthogroups. Using a non-redundant gene set also greatly reduces computational and time costs. However, it may not fully capture gene heterogeneity within orthogroups, a limitation that could be addressed by more comprehensive strategies in future work. These details have been described in the Methods (lines 768-769) and Supplementary Note 4.

4.) We also fully acknowledge the limitation that SNPs called solely on the CS reference may not capture regulatory variants for genes absent in CS. In such cases, the corresponding eQTLs are necessarily categorized as trans-acting, although their true physical proximity may be closer in alternative genomes.

I fully support the use of CSv1.1 as back bone for SNP detection here, for the same reasons as the authors. My main concern is also not the assembly quality here, rather the cases termed as "trans-acting" here. So it can be expected that a portion of those in fact are false positively classified eQTLs? If that is the case, this limitation should be estimated, described and discussed, as many of those potential eQTLs may be the ones most interesting.

Response: We thank the reviewer for the detailed comments. Regarding the sentence "So it can be expected that a portion of those in fact are false positively classified eQTLs?", we understand this to mean that some *cis*-eQTLs may be misclassified as *trans*-eQTLs, rather than these signals not being true eQTLs. Please correct us if our understanding is incorrect.

To address this question, we have provided a response in the Results section (lines 346-362): "Since the three alien introgression segments (Rye: chr1R (0-280 Mb), Renan: chr2D (570-635 Mb), and *T. timopheevii*: chr5G (435-485 Mb)) have undergone translocations with the wheat genome, all eQTLs located on wheat chromosomes were treated as *trans*-eQTLs. However, because these introgressed segments share high sequence homology with specific wheat chromosomes, we performed synteny analysis to clarify their relationships. The results showed strong collinearity between the introgressed segments and their homologous wheat chromosomes, whereas no collinearity was observed with non-homologous chromosomes (Supplementary Fig. 10). Given the known translocations, eQTLs located on Chinese Spring chromosomes 1B, 2D, and 5B were approximated as *cis*-eQTLs under our study's definitions, while those on other chromosomes were considered as *trans*-eQTLs.

We then compared the regulatory signals from *cis*-eQTLs, *trans*-eQTLs, as well as from eQTLs on homoeologous chromosomes. The analysis revealed that *cis*-eQTLs and eQTLs from homoeologous chromosomes had significantly stronger regulatory signals than *trans*-eQTLs (Wilcoxon test, p-value < 0.0001). Notably, *cis*-eQTL signals were not always stronger than those from homoeologous chromosomes (Supplementary Fig. 11), suggesting that both eQTLs on translocated chromosomes and those on homoeologous chromosomes play important roles in regulating introgressed genes."

We described the method for synteny analysis in the Methods section, lines 818-825: "Synteny analysis was conducted using JCVI[109] to investigate the chromosomal correspondence between alien introgression segments and the Chinese Spring reference genome. Gene sequences from three introgressed segments (Rye: (chr1R : 0-280 Mb), Renan: (chr2D : 570-635 Mb), and *T. timopheevii:* (chr5G : 435-485 Mb)) were subjected to pairwise synteny searches against all annotated genes in the Chinese Spring genome using JCVI[109] with default parameters. The resulting paired gene files were then used to generate macrosynteny visualizations, also employing default settings."

5.) in the context of TWAS and SMR analyses, we approximated the syntenic location of non-Chinese Spring genes by using their most significant eQTL peak mapped to the Chinese Spring genome. As shown in Fig. 4e, the strongest eQTL signals tend to cluster near the transcription start site, allowing this approach to serve as an approximate of the physical position.

Is there any reference for this approach demonstrating the overall validity? The fact that the signals cluster near transcription start sites does not appear to be a good indicator that indeed syntenic locations are being hit. I would expect that in many cases syntenic relationships are much more complex to resolve and the most significant eQTL peak hit may be just the next best random match.

Response: We thank the reviewer for this insightful comment. We fully agree that the syntenic relationships between non-Chinese Spring genes and the Chinese Spring genome can be complex, particularly in the presence of structural variation or introgressed segments. The reason we approximated the positions of non-Chinese Spring (non-CS) genes using the most significant eQTL signals is that these non-CS genes in our pan-gene atlas correspond to genes that lack homologous copy in the Chinese Spring reference, as identified by OrthoFinder. As a result, we did not specifically perform synteny analysis between these genes and the Chinese Spring genome. However, positional information was required for downstream TWAS and SMR analyses.

To support this approach, we examined the strongest eQTL signals for Chinese Spring genes and found that 65.19% were *cis*-eQTLs, suggesting that the most significant signal could provide a reasonably accurate proxy for at least part of the genes (Supplementary Figs. 8a, b). Using this strategy, we then performed TWAS by evaluating the effects of genetic variants within ±2 Mb of each gene on gene expression weights, and conducted SMR analyses separately for *cis*- and *trans*-eQTL windows to prioritize non-CS candidate genes.

Nevertheless, we acknowledge that this approximation only partially addresses the challenge, and further methodological improvements will be needed in future studies to identify additional non-Chinese Spring candidate genes more comprehensively. These revisions have been added to the manuscript in the Results section (lines 322-323: “Analysis of the strongest eQTL signals for Chinese Spring genes revealed that 65.19% were *cis*-eQTLs (Supplementary Figs. 8a, b).”) and in the Discussion section (lines 613-619: “However, as the non-Chinese Spring genes could not be assigned to the same orthogroups as Chinese Spring genes, we did not perform synteny analysis for them with Chinese Spring. Instead, we approximated their positions using the most significant eQTL signals, with the accuracy of this approach potentially reaching up to 65.19% (Supplementary Figs. 8a, b). Nevertheless, this strategy only partially addresses the limitation, and further methodological improvements will be needed in future studies to identify additional non-Chinese Spring candidate genes.”).

Reviewer #3 (Remarks to the Author):

After reviewing this new version of the manuscript, I have no further comments to make. I am satisfied the authors have addressed our initial comments as far as could reasonably be expected.

Response: We greatly appreciate the reviewer’s thorough evaluation of our revised manuscript and are pleased to know that the revisions have addressed the initial concerns

Reference

1 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417-419 (2017).

Point-by-point response to Reviewers

Response Date: Sep 10th, 2025

Zhang et al., Population-scale gene expression analysis reveals the contribution of expression diversity to the modern wheat improvement

Reviewer #2 (Remarks to the Author):

This revised version of the manuscript by Zhang et al. addresses most of my previous concerns, although mostly not by new or updated analyses. In particular I appreciate that the authors now include both evaluation data/background information for critical analyses steps, and discuss obvious limitations with some of the approaches. These include:

Point 2: "Therefore, based on the OrthoFinder classification results, we have renamed variable genes as assigned genes and private genes as unassigned genes."

My intention here was not primarily to suggest changes in terminology (although these updated terms help to navigate the potential problems) but to either take measures to harmonise gene predictions used here (which I acknowledge to be a load of work) or to estimate and discuss potential errors introduced by the current strategy. As the authors lean towards option 2, I do think this is a welcomed improvement, although the extend of downstream consequences of miss-classifications remain unclear to me (it is not about the effect of change of terminology on downstream analyses, but of genes wrongly labeled as e.g unassigned/private).

Response: We sincerely appreciate the reviewer's thoughtful comments, which helped us clarify both the motivation and potential limitations of our approach. The reviewer raised a concern that some genes labeled as "unassigned" might not be unique to a single genome, which could affect downstream interpretations. We agree that consistent annotation is essential for conventional pan-genome construction. However, the main goal of our study differs from standard pan-genome analysis. Previous transcriptomic studies in wheat have largely relied on a single reference genome[1, 2], which overlooks much of the population's transcript diversity. By integrating gene models from multiple assemblies, we aimed to expand the reference gene set to improve transcript capture and candidate gene identification. OrthoFinder was used to classify the pan-gene atlas to trace candidate gene orthogroups. It is not designed to replace rigorous pan-genome heterogeneity analysis.

Re-annotating all genomes with a unified pipeline to construct a pan-genome is undoubtedly a promising and robust approach, but it also requires substantial computational resources and time. This strategy could therefore be more appropriately implemented in future studies, where a

standardized pan-genome would provide an even stronger basis for population genetic analyses. In contrast, the lightweight pan-gene atlas strategy adopted in this study already provides a clear advantage over single-reference approaches.

The classification of gene category does not affect our main conclusions for several reasons:

1) **The definition for gene category**: All Chinese Spring genes were retained as the baseline reference, and additional gene sets from other published genomes were subsequently incorporated. If a unified annotation pipeline were applied, some of these 'unassigned' genes might indeed appear in multiple genomes. However, these genes may still be clustered into the same orthogroup by OrthoFinder, from which we select one representative gene.

2) **Impact of misclassification of gene category on expression quantification**: Using the pan-gene atlas, we detected 80,128 expressed genes, which is 23,296 more than with the Chinese Spring reference alone (including 10,357 unassigned genes). This improved detection resulted from the expanded reference gene models, which enabled greater transcript capture.

3) **Impact of misclassification of gene category on eQTL mapping**: The identification of eQTLs depends on expression variation rather than gene classification per se, so misclassification would alter counts but not major conclusions. Introgression-specific analyses were restricted to known donor regions which did not include unassigned genes.

4) **Impact of misclassification of gene category on GWAS/TWAS/SMR analyses**: Candidate genes were defined at the orthogroup level. We identified 71 non-Chinese Spring agronomic genes and 3 non-Chinese Spring disease-resistance genes. If an 'unassigned' candidate gene presents in multiple assembled genomes, its genetic association with agronomic traits should not change.

5) **Differential expression & network analyses**: These analyses focused on population-level selection signals, and the classification of unassigned genes had no impact in this section.

In summary, the misclassification of unassigned genes has minimal impact on downstream analyses. While re-annotating all 44 genomes with a uniform pipeline would indeed reduce errors in gene classification, it would also require substantial computational resources and beyond the scope of this study. Our approach leverages publicly available genomic data to investigate the landscape of expression diversity in wheat, which not only achieves new insights but also demonstrates the power of pan-genome resources generated by the international community[3].

Accordingly, we revised the Discussion sections (Lines 553-582) to explicitly clarify these points and to discuss the impact of gene misclassification on downstream analyses.

Point 3: "The reviewer's concern about alignment reliability is therefore intrinsically addressed by Salmon's statistical framework, which has demonstrated superior accuracy over comparable tools in both simulated and real datasets."

My concerns were not about the alignment reliability and the quantification tool used, but rather on what gene model (from what wheat line) was selected as the (main and only) template to map against. I do think this is a very critical step and the choice matters for both overall quantification and interpretation of the results obtained. The authors now present an evaluation demonstrating some of the effect of choosing the longest isoform per OG, and found "Among orthogroups whose longest transcript had TPM > 1, the longest transcript consistently showed significantly higher expression than its homologs within the same orthogroup".

To me, this demonstrates that indeed the longest transcript captures a greater proportion of expression, but I do not agree that this necessarily means it serves as a better representative of the orthogroup (thinking of gene prediction artefacts etc…one could also claim that an OG representative close to the OG mean expression could serve as the "best"). However, since the authors now include information and discussion about this potential concern, readers will have all required information.

Response: We sincerely appreciate the reviewer's comments, which have provided us with valuable insights and inspiration. As pointed out by the reviewer, gene models derived from different wheat lines may be affected by population heterogeneity, potentially influencing expression quantification. Having a reference genome for every wheat line would be ideal for understanding the pan-transcriptome diversity[3,4]; however, eQTL studies typically relied on a single reference genome[1, 2, 5, 6, 7, 8].

Selecting the longest isoform was also used in previous genome annotation practice and pan-transcriptome pipelines. APPRIS and MANE initiatives have compared multiple representative strategies and still consider the longest transcript to be acceptable in most cases[9]. In clinical genomics, longest isoform remains one of recommended mappings for variant interpretation[10], and UniProt similarly adopts the longest protein as the canonical model[10]. Moreover, in transcriptome assembly workflows such as EvidentialGene, longest CDS filtering retains high accuracy of recovered transcripts, further supporting this practical choice[11].

The strategy of using the longest transcript as the representative of an OG is not perfect but represents a step forward compared to traditional population-scale RNA-seq studies. Different selecting strategies such as the one mentioned by the reviewer, i.e., selecting the one close to the OG mean expression could be considered and evaluated in our further studies. Currently, we want to emphasize that our work was made possible because the release of pan-genome annotations from

the wheat community, such as the release of *de novo* annotation from the 10+ Wheat Genome Project[3]. Our work here highlighted the necessities of releasing genome annotation for more wheat varieties and encourage our peers to leverage those valuable resources in understanding this important crop.

Accordingly, we have discussed the rationale for selecting the longest transcript in the Discussion section (lines 583-591) and refined the methodological description in the Methods section (lines 793-798).

Point 4 and 5: I was indeed concerned that those signals may not represent true eQTLs at all when they are lacking a physical position in the reference genome (and not so much about a misclassification as trans-QTL). This is mainly because of the introgressed regions and translocation as also pointed out by the authors. I may not have understood the full methodology here, so I appreciate the added discussion especially on the limitations of placing the non-reference eQTLs on the genome

Response: We sincerely appreciate the reviewer's thoughtful comments. In the Results section (lines 334-346), we have highlighted our interpretation of the eQTL signals for introgressed segments: "These signals largely reflected the presence-absence variation of introgressed segments, rather than true regulatory interactions. By contrast, eQTLs calculated from only introgression lines were more evenly distributed across the genome (Figs 4f, g and Supplementary Figs. 9c, e), providing a more accurate view of how the wheat genome regulates introgressed genes." This approach was necessary because, as only a subset of accessions carries the introgressed segments, population-level eQTL mapping is strongly confounded by co-inheritance of introgressed regions. In contrast, focusing on lines with the same introgression minimizes this structural bias and more reliably captures eQTL signals from the wheat genomic background.

We revised and expanded the Discussion (lines 624-629) to emphasize this point: "Notably, when analyzing eQTLs for introgressed genes at both the population level and within introgression lines alone, we found that because introgressed segments are present only in a subset of samples, population-level eQTLs exhibit pronounced population structure, whereas eQTLs computed using only introgression lines are distributed more uniformly across the genome. Therefore, regulatory loci for introgressed segments calculated using all samples should be interpreted with caution."

Reference

1. He F, *et al.* Genomic variants affecting homoeologous gene expression dosage contribute to agronomic trait variation in allopolyploid wheat. *Nat. Commun.* **13**, 826 (2022).
2. Zhao P, *et al.* Integration of genome-wide association study, linkage analysis, and

population transcriptome analysis to reveal the TaFMO1-5B modulating seminal root growth in bread wheat. *Plant J.* **116**, 1385-1400 (2023).

3. White B*, et al.* De novo annotation of the wheat pan-genome reveals complexity and diversity of the hexaploid wheat pan-transcriptome. *bioRxiv*, 2024.01.09.574802 (2024).
4. Guo W*, et al.* A barley pan-transcriptome reveals layers of genotype-dependent transcriptional complexity. *Nat. Genet.* **57**, 441-450 (2025).
5. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature* 550, 204-213 (2017).
6. Li X, et al. The impact of rare variation on gene expression across tissues. *Nature* 550, 239-243 (2017).
7. Kremling KAG, et al. Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* 555, 520-523 (2018).
8. You J, et al. Regulatory controls of duplicated gene expression during fiber development in allotetraploid cotton. *Nat. Genet.* 55, 1987-1997 (2023).
9. Pozo F, Rodriguez JM, Martínez Gómez L, Vázquez J, Tress ML. APPRIS principal isoforms and MANE Select transcripts define reference splice variants. *Bioinformatics* **38**, ii89-ii94 (2022).
10. Pozo F, Rodriguez JM, Vázquez J, Tress ML. Clinical variant interpretation and biologically relevant reference transcripts. *NPJ Genom. Med.* **7**, 59 (2022).
11. Gilbert D. *Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes?* (2019).

Point-by-point response to Reviewers

Zhang et al., Population-scale gene expression analysis reveals the contribution of expression diversity to the modern wheat improvement

REVIEWER COMMENTS (revision 1)

Reviewer #1 (Remarks to the Author):

This study performs population RNA-seq on a diverse wheat panel with existing genome sequences. The authors present complex datasets effectively, though some sections are overly verbose. Key advances include using pan-genome resources for RNA-seq alignment (identifying >20,000 novel transcripts and improving alignment by 20%) and integrating GWAS, SMR, and TWAS to prioritize pan-genome candidates. However, critical concerns regarding developmental-stage mismatches and overstated breeding implications must be addressed.

Response: We sincerely thank the reviewer for the insightful comments and constructive suggestions. Below we address each concern in detail.

Major Concerns

1. Temporal disconnect between RNA-seq and trait data

The 34 field-measured agronomic traits (e.g., grain weight/number) do not align temporally with RNA-seq data from 2-week seedlings. Gene networks governing reproductive traits are unlikely to be established at this early stage. The authors must:

- Provide biological justification for linking seedling transcriptomes to adult phenotypes

- Acknowledge limitations of this approach or present evidence for conserved regulatory signatures across development

Response: We thank the reviewer for pointing out this important limitation. We have addressed it with the following clarifications: 1) We performed RNA-seq on 2-week-old seedlings from 328 wheat accessions, and integrated these transcriptomes with the published whole-genome resequencing data[1] to infer regulatory relationships. 2) Gene expression is inherently spatiotemporally specific, and transcriptomes from the seedling stage may not fully capture expression dynamics relevant to the 34 field traits, which are mostly measured during reproductive stages. However, modern wheat lines carry numerous introgressed segments (e.g., 1RS from rye, chromosomal fragments from *T. ponticum* and *T. timopheevii*, as well as from diploid and tetraploid relatives) that create presence/absence variation (PAV) at the expression level. Genes located in

these introgressions often show binary expression patterns, being present in carriers and absent in non-carriers, even in seedling-stage tissue, especially for highly specific genes. Thus, our expression-based associations largely reflect PAV-driven genetic regulation. 3) Our regulatory network analysis is based on seedling-stage expression data. We selected candidate genes involved in reproductive traits to illustrate how co-expression networks have changed from landraces to modern cultivars. 4) We now explicitly acknowledge the limitation in the manuscript (lines 582-585): "As the RNA-seq data used in this study were collected at the seedling stage, the detection of expression changes associated with developmental traits was limited. However, introgressed genes often exhibit presence/absence variation in expression, which was effectively captured." Future studies will aim to profile multiple tissues and developmental stages to complement and extend these findings.

2. Overextended breeding claims

Statements about direct breeding applications (e.g., "optimizing regulatory networks") lack empirical support given the study's focus on transcriptional mechanisms. While the pan-genome resource has breeding relevance, the authors should:

- Substantiate breeding claims with specific testable hypotheses

- Reframe discussions to distinguish demonstrated findings from speculative applications

Response: We thank the reviewer for the valuable comment regarding the interpretation of regulatory network changes. Our investigation of differences between landraces and cultivars in gene regulatory networks was based on the number of co-expression pairs between eGenes and active-eQTGs. We found that both modern Chinese cultivars and U.S. cultivars exhibited significantly more co-expression pairs ($|SCC| > 0.3$, p-value < 0.05) involving the same sets of eGenes and active-eQTGs compared to landraces. This suggests that the extent of co-expression relationships is reduced in landraces. However, we agree that this observation alone cannot confirm whether regulatory networks have been optimized during breeding. It is more accurate to state that the co-expression network structure has changed.

To address this concern, we revised two relevant statements in the manuscript: 1) The sentence: "Therefore, future breeding strategies should not only focus on the selection of specific genes but also on optimizing and utilizing the overall regulatory networks." has been revised to (lines 617-621): "In addition to individual expression changes, breeding has altered gene co-expression and regulatory network architecture. The number of co-expression pairs involving active-eQTGs and eGenes was significantly lower in landraces than in cultivars, suggesting that modern breeding has increased network complexity. Future breeding strategies may benefit from considering not only gene-specific selection but also optimization of broader regulatory networks." 2) The sentence: "These findings also indicate that modern breeding has reshaped co-expression regulatory networks,

potentially enhancing the ability to cope with adverse environmental conditions and improving population stability." has been revised to (lines 513-514): "Taken together, these findings suggest that modern breeding has substantially rewired co-expression and regulatory networks." These changes aim to avoid overinterpretation and better reflect the evidence provided in our study.

Minor Concerns

Data & Methods

3. Clearly cite sources of reused agronomic trait data. For novel traits, detail field trials (locations, replicates, phenotyping protocols).

Response: We thank the reviewer for the helpful suggestion. We now clearly distinguish between 20 previously published agronomic traits and 14 newly collected ones in the Methods section (lines 628-649), and provide comprehensive information on field locations, years, planting methods, and phenotyping protocols.

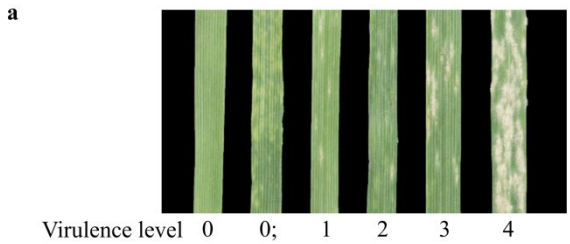
4. Clarify disease resistance scoring:

- Number of plants assessed per accession

- Handling of intra-accession variability

- Validation of the 0-4 scale (published standard or study-specific?)

Response: We appreciate the reviewer's comment. For powdery mildew scoring, 3-4 seedlings were evaluated per accession, and the average value was used. The 0-4 scoring standard was based on a published study[2], where infection types 0-2 were considered as resistant and 3-4 as susceptible. Representative images of the phenotypic standard are provided in Supplementary Fig. 11a.



Analytical Rigor

5. Reduce excessive abbreviations; define all non-standard acronyms at first use.

Response: Thank you for this valuable suggestion. We have replaced nearly all abbreviations with their full forms—for example, "CS" is now written as "Chinese Spring", "KN9204" as "Kenong 9204", "AK58" as "Aikang 58", and "HC" as "high-confidence". We retain only the most frequently

used terms such as “landraces (LRs)”, “modern Chinese cultivars (MCCs)”, and “modern U.S. cultivars (USMCs)”, which are defined at first use.

6. Specify whether the SNP matrix used in PEER analysis matches the eQTL detection matrix.

Response: We thank the reviewer for the suggestion. This is now clarified in the Methods section (lines 757-758), where we specify that the SNP matrix used in PEER analysis is consistent with that used for eQTL detection.

7. Justify expression thresholds (TPM > 0.5, SCC > 0.3) with empirical validation or field-standard references.

Response: Thank you for the reviewer’s insightful comments. The threshold of TPM > 0.5 for defining gene expression was adopted in our study based on two previous transcriptomic studies in wheat: *Ramírez-González RH et al*. (Science, 2018) and *He et al.* (Nature Communications, 2022), both of which employed this cutoff[3, 4]. We have now cited these references accordingly in the revised manuscript.

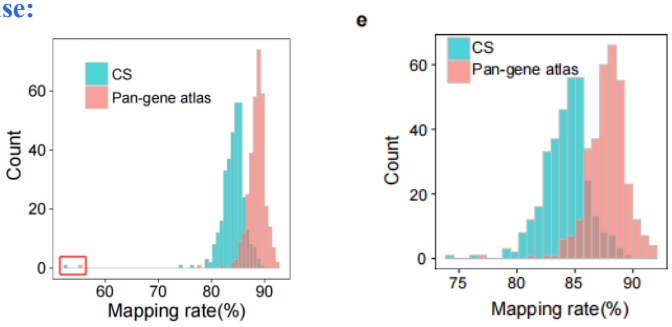
To define expression correlation, we considered gene pairs with |SCC| > 0.3 and $p < 0.05$ as significantly correlated. This threshold is supported by *Mukaka MM* (Malawi Medical Journal, 2012), who suggested that SCC values below 0.2 indicate negligible correlation, while values above 0.3 indicate positive correlation[5]. We have now cited these references accordingly in the revised manuscript.

Presentation

8. Figures:

- Fig 2E: Replace overlapping lines with histograms/marginal plots.

Response:



We thank the reviewer for this detailed suggestion. We have replaced the overlapping line plot with a frequency histogram, which better visualizes the distribution of alignment rates. In doing so, we noticed that one accession (1699B) had an unusually low mapping rate. Analysis of unmapped reads revealed contamination by Barley Yellow Mosaic Virus, so this sample was removed. Fig. 2e now

shows the alignment rates for the remaining 327 samples used in downstream analysis.

- Fig 4G: Annotate data point counts.

Response: Thank you. The figure has been revised to include the number of data points.

- Fig 5F: State randomization iterations in figure/legend.

Response: Thank you. We have added the number of randomization iterations to the legend of Fig. 5f.

9. Text:

- Fix section header formatting in "Detection of eQTL".

Response: We appreciate the reviewer's careful reading. The header formatting has been corrected.

- Tighten verbose phrasing (e.g., Introduction: "SNP diversity has been extensively characterized" vs original).

Response: Thank you for this writing suggestion. We have carefully revised the manuscript to eliminate verbose phrasing and improve clarity and logical flow throughout.

Reviewer #2 (Remarks to the Author):

This study by Zhang et al describes a large-scale gene expression analysis in bread wheat, utilizing diverse material, mainly landraces and cultivars from China and US.

The work presented contributes to a highly relevant topic, which is to better understand plasticity in gene expression patterns and their influence on important traits. Another aspect is the investigation of reference genome biases, and especially the influence of introgressions and alien material, on gene expression patterns.

While the authors generated an impressive set and amount of data, I have a few major concerns especially related to data analysis, and consequently interpretation of the results.

Response: We sincerely thank the reviewer for the thoughtful comments and constructive suggestions, which have helped us to improve the overall quality and clarity of our manuscript. Below, we provide detailed responses to each point.

1.) I feel the chapter "The impact of single reference in RNA-seq reads mapping leads to underestimate expression for introgression" does not really add much novelty but mainly confirms what has been shown before by Coombes et al.

Response: We thank the reviewer for raising this point. We agree that the observation of expression bias caused by using a single reference genome is consistent with the findings of *Coombes et al.* (2021), as shown in Figs. 1a, b and Supplementary Fig. 2. However, this section remains important because we go beyond confirming this bias by analyzing introgressed segments specific to our wheat panel, identifying expression underestimated in approximately 124 accessions with large introgressions. Additionally, we propose potential donor genome models from existing literature that could mitigate the bias. This section thus not only validates previous findings but extends them by offering practical solutions tailored to the unique structure of our panel.

2.) The construction of the pan-gene atlas has major flaws:

The authors combine data (gene predictions) from highly heterogenous resources, genome assemblies (long and short read) and annotation pipelines with no attempt of harmonization or making the resources comparable. As a consequence, the reported variable and private gene portions will without doubt at least in part be a result of technical differences rather than biological differences. This also renders the downstream finding based on the pan-gene atlas at least in part unreliable.

The simple selection of the longest gene per OG ignores substantial variation possible within some of the OGs.

Response: We sincerely thank the reviewer for the constructive suggestions regarding our methodology. We fully acknowledge that in most pan-genome studies, genome annotations are harmonized using a unified pipeline to minimize technical variation. Therefore, the observed proportions of variable and private genes in our study may partly result from differences in annotation strategies, a concern we greatly appreciate. However, our work does not represent a conventional pan-genome construction. Instead, we aimed to build a pan-gene collection by directly utilizing available gene models from wheat genomes of different ploidy levels.

Because Chinese Spring is the most widely used reference genome in the community, all genes from this genome were retained. To address the limitations of relying on a single reference, additional genes from other genomes that are absent from or lack homologs in Chinese Spring were also included. The definitions of variable and private genes apply only within the context of our constructed atlas and are not intended as genome-wide classifications. We clarified it in lines 530-536 of the revised version: “In our case, the detection of private and variable genes is constrained by the completeness and consistency of available gene annotations, which may vary due to differences in annotation pipelines across studies. Nevertheless, the goal of this work was to establish a usable reference model. Importantly, the definitions of variable and private genes in this work are specific to the gene models within our pan-gene atlas, rather than the entire genome space, aiming to classify the atlas itself rather than to comprehensively categorize all genome-encoded genes.”

Our objectives were threefold: 1) to provide a lightweight, easily reproducible method to construct a pan-gene set from published models; 2) to maximize the number of genes whose expression could be quantified, thereby improving read alignment rates compared to the traditional use of a single genome as reference; and 3) to identify more candidate genes absent from the Chinese Spring reference thus provide an evaluation of the contribution of expression diversity to the agronomic improvement.

To avoid redundancy in quantification, we selected the longest representative transcript from each orthogroup (OG), as gene expression quantification software often cannot resolve reads mapped to homologous regions with similar sequences. Retaining all homologs may lead to artificial read averaging, thereby compromising quantification accuracy. This strategy ensures more precise read assignment. For candidate gene analysis, we revisited all homologs within each OG to investigate sequence variation in detail.

3.) Selection of the wheat reference material: Is there a reason why the recently published genotypes from this wheat pan-genome where not used: https://doi.org/10.1038/s41586-024-08277-0? These data add valuable reference genome resources especially for the so-far under-represented material from Asia, which is an important part of the population genomics study in the second half of this manuscript.

Response: We thank the reviewer for this excellent suggestion. The newly published wheat pan-genome[6] indeed provides valuable resources, particularly for underrepresented Asian accessions. At the time this manuscript was initially completed, the pan-genome had not yet been published, and therefore was not included in our original analyses.

To improve the quality and scope of our work, we have now incorporated the gene models from this recent release, along with additional annotations from other wheat genomes with different ploidy levels. Specifically, we expanded our reference set from 24 to 44 genomes - we used only genome assemblies with the gene models supported by transcriptomic evidence. Because the population RNA-seq were processed using this updated reference, all analyses following Fig. 2 were re-done. All corresponding figures and results were re-generated in our revised version. While the core conclusions of the study remain unchanged, the updated version provides a more complete and informative perspective. All modifications have been clearly track-marked in the revised manuscript. Here we list some of the important changes:

1) The number of genes in the pan-gene atlas increased from 201,223 to 232,736.
2) The number of genes expressed at the population level increased from 77,447 to 80,128.
3) The number of candidate genes identified increased from 231 (including 54 non-Chinese Spring genes) to 299 (including 74 non-Chinese Spring genes).
4) The number of candidate genes validated using the Kenong 9204 mutant library increased from 58 to 86.

Therefore, expanding the reference set to include genomes suggested by this reviewer enabled the detection of more expressed genes and better evaluate the contribution of introgression to the agronomic traits.

4.) eQTL map: While the newly constructed pan-gene atlas is being used to determine gene expression levels, all the SNPs are based on the single reference Chinese Spring. While I expect a certain portion of the association to be robust nevertheless this approach contradicts somewhat the concept of reference aware resources propagated here. This should especially be true for genes and loci not being present in CS, either for technical (v1.1 predictions refer to the short read assembly version…especially the intergenic regions may be problematic) or biological reasons.

I do not really understand how "private genes" not present in CS could be associated with eQTLs based on CS SNPs…including this statement (line 301): "So, the syntenic chromosome location of the non-CS genes can be anchored by the strongest eQTL signal located at the Chinese Spring genome".

Response: We sincerely thank the reviewer for this important and insightful comment. To investigate the regulatory mechanisms of gene expression across a genetically diverse wheat population, we constructed a pan-gene atlas using a set of published genome assemblies, enabling us to capture a broader range of gene expression, including genes absent from the Chinese Spring (CS) reference. For genotype data, we utilized whole-genome-sequenced genotype data based on the CS v1.1 reference genome, which was generated by *Niu et al.*[1].

We agree with the reviewer that CS v1.1, which is based on short-read sequencing, may contain assembly errors, particularly in intergenic regions. We chose this version because it remains the most widely used reference in the wheat research community, and it provides compatibility with existing large-scale genotyping resources such as the 1000 Wheat Exomes Project[7] and the Watkins panel[8]. Moreover, most merged eQTLs are located within genic regions, with 84.8% of genes having at least one eQTL detected within the gene (Figs. 4c, d), which suggests that the potential assembly errors located in highly repeated regions in Chinese Spring v1.1 do not likely affect the main features of our eQTL map.

We also fully acknowledge the limitation that SNPs called solely on the CS reference may not capture regulatory variants for genes absent in CS. In such cases, the corresponding eQTLs are necessarily categorized as *trans*-acting, although their true physical proximity may be closer in alternative genomes. Due to current technical and computational limitations, we have not yet developed a pan-genome-based variant map. In future work, we aim to construct a wheat tribe pan-genome and perform reference-free variant calling—including SNPs, structural variants, copy number variations, and mobile element insertions—to better capture the full regulatory landscape. We have now discussed this limitation explicitly in lines 596-604 of the revised manuscript.

Regarding the reviewer's question on the association between Chinese Spring-based SNPs and private genes not present in Chinese Spring: in the context of TWAS and SMR analyses, we approximated the syntenic location of non-Chinese Spring genes by using their most significant eQTL peak mapped to the Chinese Spring genome. As shown in Fig. 4e, the strongest eQTL signals tend to cluster near the transcription start site, allowing this approach to serve as an approximate of the physical position. Since this is a methodological assumption and not a main conclusion, the original sentence "So, the syntenic chromosome location of the non-Chinese Spring genes can be anchored by the strongest eQTL signal located at the Chinese Spring genome"' has been moved from the main text to the Methods section (lines 798-800):"For non-Chinese Spring genes, the genomic positions of their most significant eQTL signals in the Chinese Spring reference genome were used as proxies", to avoid confuse the readers.

5.) Line 347: I would term this analysis an integrative approach rather than "Multi-omics modeling". No additional true omics data is introduced here to my understanding.

Response: We thank the reviewer for this observation. Our use of the term "multi-omics modeling" refers to the integration of genome-wide SNP data (genomics), gene expression levels (transcriptomics), and phenotypic traits. According to accepted definitions[9], integration of two or more omics layers qualifies as a multi-omics approach. Therefore, we consider our usage of the term to be reasonable in the context of this study. However, given that our study primarily emphasizes population-scale transcriptomic data, we have revised the term "multi-omics modeling" to "integrative modeling" to improve clarity and better reflect the focus of the study.

6.) This manuscript needs some substantial editing for language and also accuracy. Text for example in Line 659 and Line 712-713 is broken completely and that makes it hard to read and understand. There may even be something missing here.

Response: We thank the reviewer for carefully pointing this out. We have thoroughly revised the manuscript to correct unclear phrasing and incomplete sentences, including those in lines 659 and 712-713. The revised version is now more concise and readable.

7.) Figure 3a-c: how can the authors be sure that observed differences in % of expressed genes are not originating from different experimental setups and/or deviations in sampling time etc.? If I understand correctly, publicly available Rye expression data was used for this comparison from 10.1371/journal.pone.0288520. While as a genomic reference Rye Lo7 is being used, the Rye expression data in Krepski et al. was sampled from three Rye inbred lines. The results may be the same, but this approach is somewhat contradicting the call for suitable reference genomes made in this same study.

Response: We sincerely thank the reviewer for this helpful suggestion. In response to the concern regarding the comparability of public rye expression data, we generated new RNA-seq data from

six rye varieties at the seedling stage. The sequencing data have been submitted to the National Genomics Data Center (GSA: CRA025944). We reanalyzed this section using the new dataset, and the results are presented in revised Fig. 3a-d. We consistently observed that the proportion of expressed genes in rye introgression lines remains lower than in these six rye varieties, thereby confirming and strengthening our original conclusion.

8.) Line 243: the authors write about "suppressed expression" in different places throughout that chapter, actually referring to a lower number of expressed orthologous genes, not reduced or elevated expression strength (if I understand correctly…how would you compare expression quantities across different experiments other than the experimental approach using qRT-PCR?). I think this is somewhat confusing and authors could make clear this is about number of expressed genes (where the issues of the pan-gene atlas described above may have an influence on the results). Later qRT-PCR results are being presented for selected example genes.

Response: We sincerely thank the reviewer for this insightful comment. Your interpretation is correct, and we have revised the text to more clearly reflect our results. (1) The term "suppressed expression" in this context refers to a reduced number of expressed homologous genes from introgressed segments, rather than a decrease in expression intensity. (2) For differential expression analyses across experiments, we applied robust quantile normalization using the "normalize.quantiles.robust" function. This approach ranks gene expression within each sample, enabling comparisons based on expression rank rather than absolute values.

In our rye analysis, we observed upregulation of some resistance-related genes and downregulation of genes involved in primary metabolism. To validate this observation, we selected three resistance-related genes (*SECCE1Rv1G0002890*, *SECCE1Rv1G0003770*, *SECCE1Rv1G0008680*) and three basic metabolic genes for genomic DNA amplification and qRT-PCR validation. These resistance genes were present in introgression lines but appeared in only a subset of rye accessions. This suggests that breeders may have preferentially retained desirable rye traits in introgression lines. The apparent upregulation observed in transcriptomic analyses likely reflects the broader presence of these genes across introgression lines, rather than a genuine increase in their expression levels. These points have now been clarified in lines 273-278 and are further discussed in lines 559-566 of the revised manuscript.

Reviewer #3 (Remarks to the Author):

This is an excellent piece of work that will have a significant impact on the field. My comments are worth considering, but they can largely be addressed by discussing limitations rather than requiring extensive re-analysis or refocusing of sections.

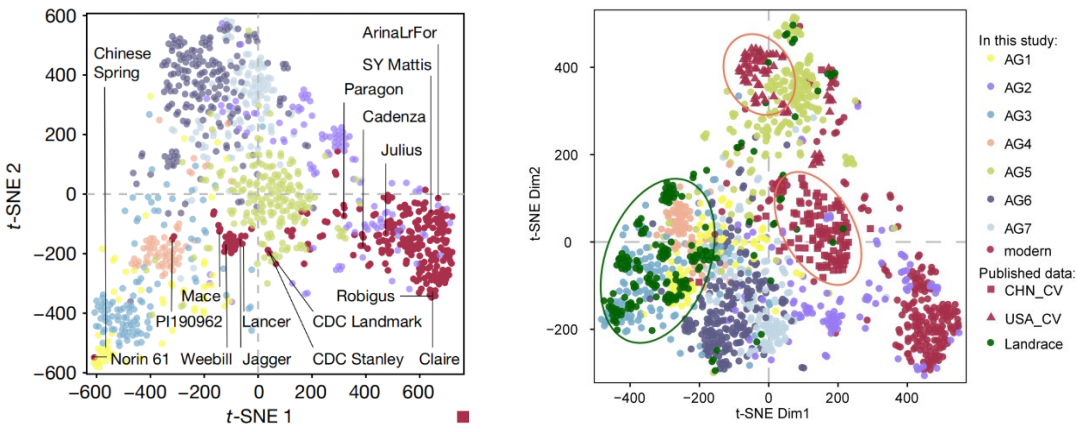
Response: We sincerely thank the reviewer for the positive evaluation and thoughtful comments, which have greatly helped us improve the clarity and completeness of our manuscript. We provide point-by-point responses below.

General Comments

The first three sections largely reiterate the arguments made by Coombes et al. (2024), demonstrating the impact of using a single reference and the advantages of a pan-transcriptome reference. Since this argument has already been made, the authors should focus on why their new Panreference atlas is an improvement over existing approaches.

A key limitation of Coombes et al. (2024) is that they only use three out of the seven ancestral groups. This study extends the approach by incorporating progenitor species into an existing 10-plus pan-transcriptome reference. However, it is crucial to assess whether the addition of 24 references adequately represents the seven ancestral groups identified in Wat-seq. Simply increasing the number of references may raise computational costs without substantially reducing bias. An iterative approach to panel design, including rounds of design testing, would be beneficial. Additionally, benchmarking should include comparisons with the existing 10-plus pan-transcriptome reference, as benchmarking solely against Chinese Spring is no longer sufficient.

Response:



(Cheng et al., 2024)

We thank the reviewer for raising this important point. The ancestral group of our wheat panel has already been assessed by in *Cheng et al.*[8], which used genotype data from *Niu et al.*[1]. In their study, CHN_CV corresponds to the MCC varieties in our panel, USA_CV to our USMC varieties, and Landraces to the landraces included in our study. As shown in their figure, MCC varieties overlap with ancestral group AG2, USMC varieties with AG5, and landraces mainly with AG1, AG3, and AG4, with a small fraction distributed across other groups. Therefore, our panel of 328 wheat accessions broadly represents five out of the seven ancestral groups (AG1-AG5). Among the 44

reference genomes used in the revised version, several correspond to key ancestral groups relevant to our panel. For example, Chinese Spring and Norin61 overlap with AG1, supporting alignment for landraces; Mace, Lancer, Jagger, CDC Landmark, and CDC Stanley align with AG5, facilitating analysis of USMC lines; and ArinaLrFor, SY Mattis, and Julius align with AG2, representing MCC lines. Thus, the inclusion of these diverse genomes significantly improves alignment accuracy across our panel by better representing ancestral variation. In addition, integration of diploid and tetraploid progenitors involved in wheat evolution further strengthens the capacity of our pan-gene atlas to capture genetic variation across wheat germplasm, including lineage-specific and introgressed alleles that are often overlooked in modern breeding. This issue has been discussed in detail in the revised manuscript (lines 540-556).

We acknowledge that *Coombes et al.* (2024) already demonstrated the benefits of a pan-transcriptome reference. However, our expanded pan-reference improves upon existing approaches by incorporating additional genomes that increase representation across multiple ancestral groups, thereby reducing reference bias more effectively in a genetically diverse panel. Given that most transcriptomic studies in wheat have relied solely on the Chinese Spring reference, we included benchmark analyses using Chinese Spring for comparison. Importantly, our analysis revealed that approximately 24,446 (30.5%) of the expressed genes in the population are derived from genomes beyond Chinese Spring and the 10+ genomes, underscoring the necessity of constructing a more comprehensive pan-gene atlas.

One indication of reference bias is the down-regulation of multiple genes within a locus. In the analysis of the diversity set, are any of these signatures still present?

Response: We thank the reviewer for this important question. Reference bias can indeed result in the underestimation of gene expression levels. In our analysis, we found that such signatures were evident when using only the Chinese Spring reference genome. For instance, as shown in Fig. 1a, in 1RS.1BL introgression lines, expression levels on chromosome 1B were reduced, reflecting misalignment of rye-derived transcripts. However, after incorporating the rye genome into the pan-gene reference (Fig. 1d), a greater number of expressed rye genes were detected, and fewer genes were spuriously mapped to the 1B chromosome of Chinese Spring. This reallocation corrects the expression underestimation caused by reference bias. Although the number of genes detected within a locus may change more substantially across genotypes, these differences reflect improved assignment of expression to the correct genomic origin, thereby mitigating reference bias in the diversity set. These results demonstrate that the inclusion of multiple donor genomes in the pan-reference framework effectively reduces reference bias and improves the accuracy of gene expression quantification across diverse genotypes.

Reference Bias in eQTL Mapping

In terms of eQTL mapping, reference bias affects two key areas:

1. RNA-seq reference bias, which this paper effectively addresses.

2. Reference bias in WGS read mapping, where WGS reads are mapped to the Chinese Spring reference to generate SNPs. This issue should be explicitly discussed.

Response: We thank the reviewer for their valuable comment. For genotype data, we utilized whole-genome-sequenced genotype data based on the CS v1.1 reference genome, which was generated by *Niu et al.*[1]. We acknowledge that reference bias still exists, particularly in regions not represented by the Chinese Spring genome. Due to the current limit of computational resource, SNPs in this study were based on a single reference. In future work, we aim to construct a wheat pan-genome and perform SNPs, structural variants, copy number variations, and mobile element insertions discovery using a multi-genome framework to better capture the full spectrum of genetic variation. We have added a discussion of this issue in lines 596-604 of the revised manuscript.

TWAS Study

The TWAS approach helps mitigate some of the challenges associated with reference bias in eQTL mapping, as it enables the identification of non-Chinese Spring genes. This is perhaps the most innovative aspect of the paper. However, the integration of GWAS and TWAS is not entirely clear, particularly given the limitations of mapping WGS reads to a Chinese Spring reference. Further clarification on this integration would improve the manuscript.

Response: We thank the reviewer for this insightful comment. We have added a more detailed explanation of the integration of GWAS and TWAS in the Methods section (lines: 791-808). Specifically, while GWAS can only identify candidate genes located within the Chinese Spring genome, TWAS first estimates expression weights for all genes, including both Chinese Spring and non-Chinese Spring genes, based on eQTL data. By integrating expression weights and SNP-phenotype associations from GWAS, TWAS enables the identification of regulatory links between SNPs and the expression of non-Chinese Spring genes that may contribute to phenotypic variation. Thus, SNPs serve as a bridge connecting GWAS and TWAS to detect associations beyond the reference genome. The TWAS analysis in this study relied on GWAS summary statistics, and the detailed procedures follow the official TWAS pipeline available at http://gusevlab.org/projects/fusion/[10].

The advantages of using TWAS and SMR to identify non-Chinese Spring candidate genes are further discussed in lines 577-582: “Because GWAS relies solely on SNP information present in the Chinese Spring genome to associate with phenotypes, it can only identify candidate genes located within the Chinese Spring genome. In contrast, by using TWAS, SMR, and correlation analyses between gene expression and phenotypes, we were able to identify candidate genes not present in the Chinese Spring genome. The use of transcriptomic data enabled us to identify a greater number of non-Chinese Spring candidate genes”.

Differential Gene Expression Analysis

The research question is well-defined: the authors aim to explore how different breeding programs and landraces have influenced differential gene regulation and networks. However, the results, methods, and figures are difficult to follow. Providing a clearer explanation of the differential expression analysis would enhance the readability of the subsequent network analysis. Using example genes—such as Vern 2 as a figure panel and its correlation between winter and spring lines—could improve clarity.

Response: We sincerely thank the reviewer for this valuable suggestion. To improve the clarity and logic of the section on differentially expressed genes, we have revised the content of the "Differentially expressed genes and modules between different sub-populations" section in the revised version to enhance readability and interpretation. In addition, we analyzed the expression correlation between *VRN1*, *VRN2*, and *VRN3* and the growth habits (spring vs. winter wheat), as shown in Supplementary Fig. 15. Together with Supplementary Figs. 16 and 17, these results provide a clearer understanding of the differential expression patterns of *VRN1*, *VRN2*, and *VRN3*.

Data Availability

The current Panreference atlas is not readily usable. It is critical that the authors provide a FASTA file for the pan-atlas rather than just a list of gene IDs.

Response: We sincerely thank the reviewer for this suggestion. We have now provided the FASTA file and made it available on Figshare database (https://figshare.com/s/ee89a74381fea08c4c4e).

Reference

1. Niu J, *et al.* Whole-genome sequencing of diverse wheat accessions uncovers genetic changes during modern breeding in China and the United States. *Plant Cell* **35**, 4199-4216 (2023).
2. Xie J, *et al.* A biGWAS strategy reveals the genetic architecture of the interaction between wheat and *Blumeria graminis* f. sp. *tritici*[J]. *bioRxiv*, 2025.04.09.647224 (2025)
3. He F, *et al.* Genomic variants affecting homoeologous gene expression dosage contribute to agronomic trait variation in allopolyploid wheat. *Nat. Commun.* **13**, 826 (2022).
4. Ramírez-González RH, *et al.* The transcriptional landscape of polyploid wheat. *Science* **361**, (2018).
5. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med. J.* **24**, 69-71 (2012).
6. Jiao C, *et al.* Pan-genome bridges wheat structural variations with habitat and breeding. *Nature* **637**, 384-393 (2025).

7. He F, *et al.* Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nat. Genet.* **51**, 896-904 (2019).
8. Cheng S, *et al.* Harnessing landrace diversity empowers wheat breeding. *Nature* **632**, 823-831 (2024).
9. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol.* **18**, 83 (2017).
10. Gusev A, *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245-252 (2016).

REVIEWER COMMENTS (revision 2)

Reviewer #1 (Remarks to the Author):

The authors have addressed all my concerns. I have no further comments.

Response: We sincerely thank the reviewer for taking the time to review our revised manuscript and for the positive feedback.

Reviewer #2 (Remarks to the Author):

The revised manuscript by Zhang et al. addresses many of my previous concerns. In particular, the generation and (re-) analysis of native rye expression data strengthens the author's claims, and I also appreciate the inclusion of material from the latest pan-genome and following re-analysis. This also further improves the value of the data generated here as a critical resource for wheat improvement.
I have a few points of concern following the author's responses:

1.) However, this section remains important because we go beyond confirming this bias by analyzing introgressed segments specific to our wheat panel, identifying expression underestimated in approximately 124 accessions with large introgressions. Additionally, we propose potential donor genome models from existing literature that could mitigate the bias. This section thus not only validates previous findings but extends them by offering practical solutions tailored to the unique structure of our panel.

I see and appreciate this aspect, but I do feel that the paragraph needs to be modified accordingly. As it stands now, it still mainly reads as a confirmation of existing knowledge, and does not transport the message of a "practical solution tailored to the panel" well.

Response: We sincerely thank the reviewer for the valuable comment, which helped us clarify the

intention of our analysis. In the revised manuscript, we have rewritten the relevant section in the Results (lines 182-198) to better describe the corrective effect of incorporating potential donor genomes on gene expression quantification. The revised text now reads as follows: “To accurately quantify gene expression in introgression regions, we merged the donor genome with the Chinese Spring reference genome (see Methods) and re-analyzed gene expression for all wheat samples. We found that when using only the Chinese Spring reference genome, the average number of expressed genes detected in the introgression lines within the corresponding translocated chromosome segments (chr1A: 0-214.4 Mb, chr1B: 0-239.3 Mb, chr2D: 570.1-619.2 Mb, chr3D: 500-615.5 Mb, chr5B: 497.2-537 Mb) was 14, 12, 39, 26 and 27 per 10 Mb window, respectively. In contrast, when using the merged genome as reference, the average number of expressed genes mapped to these Chinese Spring segments decreased to 4, 4, 22, 12 and 9, while the corresponding donor-derived regions Rye (chr1R: 0-280 Mb), Renan (chr2D: 570-635 Mb), *Th. elongatum* (chr3E: 500-676.9 Mb), and *T. timopheevii* (chr5G: 435-485 Mb) showed higher average number of expressed genes: 15, 26, 19 and 29 per 10 Mb window. These results demonstrate that using the merged genome enables a substantial number of mis-mapped transcripts to be correctly assigned to their donor genome origins, effectively recovering actively expressed genes in the introgressed regions across each 10 Mb window (Fig. 1d and Supplementary Note 1). Thus, the reference bias can be severe when measuring expression in introgression using a single reference genome. Incorporating gene models from donor species is essential to accurately estimate the expression levels in introgressed regions.”

2.) However, our work does not represent a conventional pan-genome construction. Instead, we aimed to build a pan-gene collection by directly utilizing available gene models from wheat genomes of different ploidy levels….The definitions of variable and private genes apply only within the context of our constructed atlas and are not intended as genome-wide classifications.

I understand that, and appreciate that the authors are not making claims about (pan-) genome-wide gene classifications here. However, the problem remains that the genes classified as variable and private in this non-harmonized analysis are being used with these labels in downstream analyses, and conclusions are being drawn (also) based on those unreliable classifications. Or in other words, if a substantial portion of genes is classified as “private” genes as a result of gene prediction differences only, this would still have an effect on their use and interpretation in the pan-gene atlas presented here.

Response: We thank the reviewer for the insightful and rigorous comment. Our pan-gene atlas was categorized using Orthofinder. Initially, we classified genes based on the Orthofinder output files *Orthogroups.tsv* and *Orthogroups_UnassignedGenes.tsv*, labeling them as variable genes and private genes, respectively. This differs from the classical pan-genome classification approach. As we also discussed in the revised manuscript (lines 558-563), “In our case, the detection of

unassigned and assigned genes is constrained by the completeness and consistency of available gene annotations, which may vary due to differences in annotation pipelines across studies. In addition, the assignment of genes into orthogroups by OrthoFinder may be affected by sequence similarity thresholds, clustering parameters, and the complex evolutionary history of gene families, which could further influence the number of unassigned genes and their genomic distribution." and indeed, defining genes in *Orthogroups_UnassignedGenes.tsv* as private genes may cause confusion for readers. Therefore, based on the OrthoFinder classification results, we have renamed variable genes as assigned genes and private genes as unassigned genes.

Accordingly, we have updated all relevant sections of the manuscript, including figures and tables, to use the terms "assigned" and "unassigned" genes rather than "variable" and "private" genes. Importantly, this change in terminology does not affect our downstream analyses or the study's conclusions, as the underlying classification method remains unchanged.

3.) To avoid redundancy in quantification, we selected the longest representative transcript from each orthogroup (OG), as gene expression quantification software often cannot resolve reads mapped to homologous regions with similar sequences. Retaining all homologs may lead to artificial read averaging, thereby compromising quantification accuracy. This strategy ensures more precise read assignment. For candidate gene analysis, we revisited all homologs within each OG to investigate sequence variation in detail.

I do understand the technical rational behind this approach with respect to read assignment. However, my point is that presumably in many cases, the longest transcript may not be a good representative of the orthologous group per se. This is especially true in a non-harmonized gene prediction dataset which is used in this study.

Additionally I am wondering about this: the pan-gene atlas consists of genes from a broad and diverse panel of wild, domesticated wheats, landraces, and even other triticeae genomes (? Line 205), and I am not clear how that affects the mapping accuracy of the also diverse transcriptome samples. The authors evaluated this on the basis of numbers of expressed genes, but this is not really indicative of mapping quality. For example, what happens if expression data from modern cultivars just does not map to the wild wheat representative gene model which was chosen from an OG just because it was the longest transcript? I do think that this should be evaluated more deeply, for example by looking at and comparing the mapping quality of all alignments within an orthologous group.

Response: We sincerely thank the reviewer for this thoughtful suggestion. We quantified gene expression using Salmon, a pseudo-alignment tool that does not output traditional per-read mapping quality scores (such as MAPQ from aligners like Bowtie2). Instead, Salmon integrates mapping

confidence into expression estimates through a conditional probability model, rich equivalence classes, and probabilistic abundance estimation. In this framework, high expression levels primarily result from fragments with high-confidence mappings, whereas the influence of lower-quality mappings is effectively down-weighted by the model. The reviewer's concern about alignment reliability is therefore intrinsically addressed by Salmon's statistical framework, which has demonstrated superior accuracy over comparable tools in both simulated and real datasets[1]. Consequently, the expressed genes identified in our study are expected to be of high quality.

To further assess whether choosing the longest transcript within each orthogroup is appropriate—especially when the longest transcript originates from wild wheat and its relevance to modern cultivars might be questioned—we constructed a redundant pan-gene set by merging all genes from the 44 wheat genomes. Using two Chinese cultivars and two U.S. cultivars, we quantified expression levels across all genes. We then focused on orthogroups where the longest representative transcript was derived from diploid or tetraploid species. Among orthogroups whose longest transcript had TPM > 1, the longest transcript consistently showed significantly higher expression than its homologs within the same orthogroup (Wilcoxon test, p-value $< 2.2 \times 10^{-16}$). In contrast, for orthogroups where the longest transcript had TPM $\leqslant$ 1, there was no significant difference between the longest transcript and other transcripts in the group (Supplementary Fig. 23). These results suggest that the longest transcript tends to capture a greater proportion of expression and thus serves as a better representative of the orthogroup.

Furthermore, we compared the expression of the longest transcripts from the non-redundant pan-gene atlas constructed in this study to those from the redundant pan-gene set. The longest transcripts in the non-redundant atlas generally exhibited slightly higher expression, which could be because reads that would otherwise be distributed across multiple homologous transcripts in the redundant set instead consolidate onto a single representative transcript (Supplementary Fig. 23). Overall, these findings support the choice of the longest transcript as a practical and representative measure of expression abundance within orthogroups. Using a non-redundant gene set also greatly reduces computational and time costs. However, it may not fully capture gene heterogeneity within orthogroups, a limitation that could be addressed by more comprehensive strategies in future work. These details have been described in the Methods (lines 768-769) and Supplementary Note 4.

4.) We also fully acknowledge the limitation that SNPs called solely on the CS reference may not capture regulatory variants for genes absent in CS. In such cases, the corresponding eQTLs are necessarily categorized as trans-acting, although their true physical proximity may be closer in alternative genomes.

I fully support the use of CSv1.1 as back bone for SNP detection here, for the same reasons as the authors. My main concern is also not the assembly quality here, rather the cases termed as "trans-

acting" here. So it can be expected that a portion of those in fact are false positively classified eQTLs? If that is the case, this limitation should be estimated, described and discussed, as many of those potential eQTLs may be the ones most interesting.

Response: We thank the reviewer for the detailed comments. Regarding the sentence "So it can be expected that a portion of those in fact are false positively classified eQTLs?", we understand this to mean that some *cis*-eQTLs may be misclassified as *trans*-eQTLs, rather than these signals not being true eQTLs. Please correct us if our understanding is incorrect.

To address this question, we have provided a response in the Results section (lines 346-362): "Since the three alien introgression segments (Rye: chr1R (0-280 Mb), Renan: chr2D (570-635 Mb), and *T. timopheevii*: chr5G (435-485 Mb)) have undergone translocations with the wheat genome, all eQTLs located on wheat chromosomes were treated as *trans*-eQTLs. However, because these introgressed segments share high sequence homology with specific wheat chromosomes, we performed synteny analysis to clarify their relationships. The results showed strong collinearity between the introgressed segments and their homologous wheat chromosomes, whereas no collinearity was observed with non-homologous chromosomes (Supplementary Fig. 10). Given the known translocations, eQTLs located on Chinese Spring chromosomes 1B, 2D, and 5B were approximated as *cis*-eQTLs under our study's definitions, while those on other chromosomes were considered as *trans*-eQTLs.

We then compared the regulatory signals from *cis*-eQTLs, *trans*-eQTLs, as well as from eQTLs on homoeologous chromosomes. The analysis revealed that *cis*-eQTLs and eQTLs from homoeologous chromosomes had significantly stronger regulatory signals than *trans*-eQTLs (Wilcoxon test, p-value < 0.0001). Notably, *cis*-eQTL signals were not always stronger than those from homoeologous chromosomes (Supplementary Fig. 11), suggesting that both eQTLs on translocated chromosomes and those on homoeologous chromosomes play important roles in regulating introgressed genes."

We described the method for synteny analysis in the Methods section, lines 818-825: "Synteny analysis was conducted using JCVI[109] to investigate the chromosomal correspondence between alien introgression segments and the Chinese Spring reference genome. Gene sequences from three introgressed segments (Rye: (chr1R : 0-280 Mb), Renan: (chr2D : 570-635 Mb), and *T. timopheevii:* (chr5G : 435-485 Mb)) were subjected to pairwise synteny searches against all annotated genes in the Chinese Spring genome using JCVI[109] with default parameters. The resulting paired gene files were then used to generate macrosynteny visualizations, also employing default settings."

5.) in the context of TWAS and SMR analyses, we approximated the syntenic location of non-Chinese Spring genes by using their most significant eQTL peak mapped to the Chinese Spring genome. As shown in Fig. 4e, the strongest eQTL signals tend to cluster near the transcription start

site, allowing this approach to serve as an approximate of the physical position.

Is there any reference for this approach demonstrating the overall validity? The fact that the signals cluster near transcription start sites does not appear to be a good indicator that indeed syntenic locations are being hit. I would expect that in many cases syntenic relationships are much more complex to resolve and the most significant eQTL peak hit may be just the next best random match.

Response: We thank the reviewer for this insightful comment. We fully agree that the syntenic relationships between non-Chinese Spring genes and the Chinese Spring genome can be complex, particularly in the presence of structural variation or introgressed segments. The reason we approximated the positions of non-Chinese Spring (non-CS) genes using the most significant eQTL signals is that these non-CS genes in our pan-gene atlas correspond to genes that lack homologous copy in the Chinese Spring reference, as identified by OrthoFinder. As a result, we did not specifically perform synteny analysis between these genes and the Chinese Spring genome. However, positional information was required for downstream TWAS and SMR analyses.

To support this approach, we examined the strongest eQTL signals for Chinese Spring genes and found that 65.19% were *cis*-eQTLs, suggesting that the most significant signal could provide a reasonably accurate proxy for at least part of the genes (Supplementary Figs. 8a, b). Using this strategy, we then performed TWAS by evaluating the effects of genetic variants within ±2 Mb of each gene on gene expression weights, and conducted SMR analyses separately for *cis*- and *trans*-eQTL windows to prioritize non-CS candidate genes.

Nevertheless, we acknowledge that this approximation only partially addresses the challenge, and further methodological improvements will be needed in future studies to identify additional non-Chinese Spring candidate genes more comprehensively. These revisions have been added to the manuscript in the Results section (lines 322-323: “Analysis of the strongest eQTL signals for Chinese Spring genes revealed that 65.19% were *cis*-eQTLs (Supplementary Figs. 8a, b).”) and in the Discussion section (lines 613-619: “However, as the non-Chinese Spring genes could not be assigned to the same orthogroups as Chinese Spring genes, we did not perform synteny analysis for them with Chinese Spring. Instead, we approximated their positions using the most significant eQTL signals, with the accuracy of this approach potentially reaching up to 65.19% (Supplementary Figs. 8a, b). Nevertheless, this strategy only partially addresses the limitation, and further methodological improvements will be needed in future studies to identify additional non-Chinese Spring candidate genes.”).

Reviewer #3 (Remarks to the Author):

After reviewing this new version of the manuscript, I have no further comments to make. I am

satisfied the authors have addressed our initial comments as far as could reasonably be expected.

Response: We greatly appreciate the reviewer's thorough evaluation of our revised manuscript and are pleased to know that the revisions have addressed the initial concerns

Reference

1 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417-419 (2017).

REVIEWER COMMENTS (revision 3)

Reviewer #2 (Remarks to the Author):

This revised version of the manuscript by Zhang et al. addresses most of my previous concerns, although mostly not by new or updated analyses. In particular I appreciate that the authors now include both evaluation data/background information for critical analyses steps, and discuss obvious limitations with some of the approaches. These include:

Point 2: "Therefore, based on the OrthoFinder classification results, we have renamed variable genes as assigned genes and private genes as unassigned genes."

My intention here was not primarily to suggest changes in terminology (although these updated terms help to navigate the potential problems) but to either take measures to harmonise gene predictions used here (which I acknowledge to be a load of work) or to estimate and discuss potential errors introduced by the current strategy. As the authors lean towards option 2, I do think this is a welcomed improvement, although the extend of downstream consequences of miss-classifications remain unclear to me (it is not about the effect of change of terminology on downstream analyses, but of genes wrongly labeled as e.g unassigned/private).

Response: We sincerely appreciate the reviewer's thoughtful comments, which helped us clarify both the motivation and potential limitations of our approach. The reviewer raised a concern that some genes labeled as "unassigned" might not be unique to a single genome, which could affect downstream interpretations. We agree that consistent annotation is essential for conventional pan-genome construction. However, the main goal of our study differs from standard pan-genome analysis. Previous transcriptomic studies in wheat have largely relied on a single reference genome[1, 2], which overlooks much of the population's transcript diversity. By integrating gene models from multiple assemblies, we aimed to expand the reference gene set to improve transcript capture and

candidate gene identification. OrthoFinder was used to classify the pan-gene atlas to trace candidate gene orthogroups. It is not designed to replace rigorous pan-genome heterogeneity analysis.

Re-annotating all genomes with a unified pipeline to construct a pan-genome is undoubtedly a promising and robust approach, but it also requires substantial computational resources and time. This strategy could therefore be more appropriately implemented in future studies, where a standardized pan-genome would provide an even stronger basis for population genetic analyses. In contrast, the lightweight pan-gene atlas strategy adopted in this study already provides a clear advantage over single-reference approaches.

The classification of gene category does not affect our main conclusions for several reasons:

1) **The definition for gene category**: All Chinese Spring genes were retained as the baseline reference, and additional gene sets from other published genomes were subsequently incorporated. If a unified annotation pipeline were applied, some of these 'unassigned' genes might indeed appear in multiple genomes. However, these genes may still be clustered into the same orthogroup by OrthoFinder, from which we select one representative gene.

2) **Impact of misclassification of gene category on expression quantification**: Using the pan-gene atlas, we detected 80,128 expressed genes, which is 23,296 more than with the Chinese Spring reference alone (including 10,357 unassigned genes). This improved detection resulted from the expanded reference gene models, which enabled greater transcript capture.

3) **Impact of misclassification of gene category on eQTL mapping**: The identification of eQTLs depends on expression variation rather than gene classification per se, so misclassification would alter counts but not major conclusions. Introgression-specific analyses were restricted to known donor regions which did not include unassigned genes.

4) **Impact of misclassification of gene category on GWAS/TWAS/SMR analyses**: Candidate genes were defined at the orthogroup level. We identified 71 non-Chinese Spring agronomic genes and 3 non-Chinese Spring disease-resistance genes. If an 'unassigned' candidate gene presents in multiple assembled genomes, its genetic association with agronomic traits should not change.

5) **Differential expression & network analyses**: These analyses focused on population-level selection signals, and the classification of unassigned genes had no impact in this section.

In summary, the misclassification of unassigned genes has minimal impact on downstream analyses. While re-annotating all 44 genomes with a uniform pipeline would indeed reduce errors in gene classification, it would also require substantial computational resources and beyond the scope of

this study. Our approach leverages publicly available genomic data to investigate the landscape of expression diversity in wheat, which not only achieves new insights but also demonstrates the power of pan-genome resources generated by the international community[3].

Accordingly, we revised the Discussion sections (Lines 553-582) to explicitly clarify these points and to discuss the impact of gene misclassification on downstream analyses.

Point 3: “The reviewer's concern about alignment reliability is therefore intrinsically addressed by Salmon’s statistical framework, which has demonstrated superior accuracy over comparable tools in both simulated and real datasets.”

My concerns were not about the alignment reliability and the quantification tool used, but rather on what gene model (from what wheat line) was selected as the (main and only) template to map against. I do think this is a very critical step and the choice matters for both overall quantification and interpretation of the results obtained. The authors now present an evaluation demonstrating some of the effect of choosing the longest isoform per OG, and found “Among orthogroups whose longest transcript had TPM > 1, the longest transcript consistently showed significantly higher expression than its homologs within the same orthogroup”.

To me, this demonstrates that indeed the longest transcript captures a greater proportion of expression, but I do not agree that this necessarily means it serves as a better representative of the orthogroup (thinking of gene prediction artefacts etc…one could also claim that an OG representative close to the OG mean expression could serve as the "best"). However, since the authors now include information and discussion about this potential concern, readers will have all required information.

Response: We sincerely appreciate the reviewer’s comments, which have provided us with valuable insights and inspiration. As pointed out by the reviewer, gene models derived from different wheat lines may be affected by population heterogeneity, potentially influencing expression quantification. Having a reference genome for every wheat line would be ideal for understanding the pan-transcriptome diversity[3,4]; however, eQTL studies typically relied on a single reference genome[1, 2, 5, 6, 7, 8].

Selecting the longest isoform was also used in previous genome annotation practice and pan-transcriptome pipelines. APPRIS and MANE initiatives have compared multiple representative strategies and still consider the longest transcript to be acceptable in most cases[9]. In clinical genomics, longest isoform remains one of recommended mappings for variant interpretation[10], and UniProt similarly adopts the longest protein as the canonical model[10]. Moreover, in transcriptome assembly workflows such as EvidentialGene, longest CDS filtering retains high accuracy of

recovered transcripts, further supporting this practical choice[11].

The strategy of using the longest transcript as the representative of an OG is not perfect but represents a step forward compared to traditional population-scale RNA-seq studies. Different selecting strategies such as the one mentioned by the reviewer, i.e., selecting the one close to the OG mean expression could be considered and evaluated in our further studies. Currently, we want to emphasize that our work was made possible because the release of pan-genome annotations from the wheat community, such as the release of *de novo* annotation from the 10+ Wheat Genome Project[3]. Our work here highlighted the necessities of releasing genome annotation for more wheat varieties and encourage our peers to leverage those valuable resources in understanding this important crop.

Accordingly, we have discussed the rationale for selecting the longest transcript in the Discussion section (lines 583-591) and refined the methodological description in the Methods section (lines 793-798).

Point 4 and 5: I was indeed concerned that those signals may not represent true eQTLs at all when they are lacking a physical position in the reference genome (and not so much about a misclassification as trans-QTL). This is mainly because of the introgressed regions and translocation as also pointed out by the authors. I may not have understood the full methodology here, so I appreciate the added discussion especially on the limitations of placing the non-reference eQTLs on the genome

Response: We sincerely appreciate the reviewer's thoughtful comments. In the Results section (lines 334-346), we have highlighted our interpretation of the eQTL signals for introgressed segments: "These signals largely reflected the presence-absence variation of introgressed segments, rather than true regulatory interactions. By contrast, eQTLs calculated from only introgression lines were more evenly distributed across the genome (Figs 4f, g and Supplementary Figs. 9c, e), providing a more accurate view of how the wheat genome regulates introgressed genes." This approach was necessary because, as only a subset of accessions carries the introgressed segments, population-level eQTL mapping is strongly confounded by co-inheritance of introgressed regions. In contrast, focusing on lines with the same introgression minimizes this structural bias and more reliably captures eQTL signals from the wheat genomic background.

We revised and expanded the Discussion (lines 624-629) to emphasize this point: "Notably, when analyzing eQTLs for introgressed genes at both the population level and within introgression lines alone, we found that because introgressed segments are present only in a subset of samples, population-level eQTLs exhibit pronounced population structure, whereas eQTLs computed using only introgression lines are distributed more uniformly across the genome. Therefore, regulatory

loci for introgressed segments calculated using all samples should be interpreted with caution."

Reference

1. He F, *et al.* Genomic variants affecting homoeologous gene expression dosage contribute to agronomic trait variation in allopolyploid wheat. *Nat. Commun.* **13**, 826 (2022).
2. Zhao P, *et al.* Integration of genome-wide association study, linkage analysis, and population transcriptome analysis to reveal the TaFMO1-5B modulating seminal root growth in bread wheat. *Plant J.* **116**, 1385-1400 (2023).
3. White B, *et al.* De novo annotation of the wheat pan-genome reveals complexity and diversity of the hexaploid wheat pan-transcriptome. *bioRxiv*, 2024.01.09.574802 (2024).
4. Guo W, *et al.* A barley pan-transcriptome reveals layers of genotype-dependent transcriptional complexity. *Nat. Genet.* **57**, 441-450 (2025).
5. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature* 550, 204-213 (2017).
6. Li X, et al. The impact of rare variation on gene expression across tissues. *Nature* 550, 239-243 (2017).
7. Kremling KAG, et al. Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* 555, 520-523 (2018).
8. You J, et al. Regulatory controls of duplicated gene expression during fiber development in allotetraploid cotton. *Nat. Genet.* 55, 1987-1997 (2023).
9. Pozo F, Rodriguez JM, Martínez Gómez L, Vázquez J, Tress ML. APPRIS principal isoforms and MANE Select transcripts define reference splice variants. *Bioinformatics* **38**, ii89-ii94 (2022).
10. Pozo F, Rodriguez JM, Vázquez J, Tress ML. Clinical variant interpretation and biologically relevant reference transcripts. *NPJ Genom. Med.* **7**, 59 (2022).
11. Gilbert D. *Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes?* (2019).

REVIEWER COMMENTS (revision 4)

Reviewer #2 (Remarks to the Author):

I appreciate the detailed responses and the inclusion of a discussion of the limitations in the latest revised manuscript version.
I have no further comments.

Response: We sincerely appreciate the reviewer's time and positive assessment of our work.