Integrative Analysis from the Epigenome to Translatome Uncovers Patterns of Dominant Nuclear Regulation during Transient Stress

Travis A. Lee and Julia Bailey-Serres

*Plant Cell. Advance Publication September 13, 2019; doi: 10.1105/tpc.19.00463*

*Corresponding author: Julia Bailey-Serres serres@ucr.edu*

### Review timeline:

<table>
<thead>
<tr>
<th>TPC2019-00463-LSB</th>
<th>Submission received:</th>
<th>June 20, 2019</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st Decision:</td>
<td>July 22, 2019 revision requested</td>
</tr>
<tr>
<td>TPC2019-00463-LSB1</td>
<td>1st Revision received:</td>
<td>Aug. 21, 2019</td>
</tr>
<tr>
<td></td>
<td>2nd Decision:</td>
<td>Sept. 2, 2019 acceptance pending, sent to science editor</td>
</tr>
<tr>
<td></td>
<td>Final acceptance:</td>
<td>Sept. 12, 2019</td>
</tr>
<tr>
<td></td>
<td>Advance publication:</td>
<td>Sept. 13, 2019</td>
</tr>
</tbody>
</table>

### REPORT:

(The report shows the major requests for revision and author responses. Minor comments for revision and miscellaneous correspondence are not included. The original format may not be reflected in this compilation, but the reviewer comments and author responses are not edited, except to correct minor typographical or spelling errors that could be a source of ambiguity.)

TPC2019-00463-LSB 1st Editorial decision – revision requested  July 22, 2019

Please address all Reviewer comments as substantially as possible. Additionally, as noted by the majority of reviewers, it would be great if you can make the writing more accessible to non-experts in the field that is the focus of your manuscript. Once you have addressed these two points, please resubmit your manuscript with a comprehensive 'Response to Reviewers' document.

Annette is sending to you separately some suggestions for the figures, which you can try to correct in the revision. Please apply the comments to all figures in the paper. You can use your judgment with the goal of visual appeal on one hand, and legibility on the other. Please also consider displaying the right number of significant figures (first uncertain digit); the others may be retained and hidden.

Reviewer comments:

[Reviewer comments shown below along with author responses]

TPC2019-00463-LSB 1st Revision received  Aug. 21, 2019

Reviewer comments and **author responses**:

**Reviewer #1:**

This research paper by Lee and Bailey-Serres provides a highly robust and comprehensive analysis of a large range of chromatin, nuclear and RNA changes that occur at the genome-wide scale in response to transient stress, namely hypoxia. Through analysing and cross comparing eight chromatin assays (including several histone modifications, H2A.Z occupancy, chromatin accessibility, and active RNAPII etc.), as well as RNA assays (including nRNA, polyA and TRAP), the authors provide key and high-resolution insight into how different genes are differentially regulated in response to oxygen deprivation. The resulting data sets, and accompanying analysis (which includes more targeted assessment of specific gene groupings, such as the 49 hypoxia regulated genes, and cytosolic ribosome proteins) provide a holistic overview of nuclear regulation, which (to my knowledge) has not been done before to such a degree (at least not in plants). The data generated in this manuscript will be invaluable to
researchers investigating hypoxia tolerance, but will also be relevant to the broader scientific community that is interested in signal-responsive changes in nuclear regulation, as it uncovers novel and unexpected patterns and (anti)correlations between different processes that are likely to be applicable across diverse stresses.

Considering the complexity of this study, the manuscript is exceptionally well written and largely clear. The Introduction is appropriate, and the Discussion and Conclusions are well rounded. The data presentation is precise, and generally easy to follow (which is not always easy with studies of this kind). I think the authors should be commended for the breadth of the study, and particularly with its successful attempt to integrate such disparate datasets and produce meaningful insight and conclusions. I have only very minor comments (minor comments addressed).

Reviewer #2:

This article has an ambitious goal, which I have not seen tackled elsewhere. Using as a model system, the response of Arabidopsis seedlings to hypoxia and reaeration, it set out to characterize much of the entire gene expression pipeline from (five) histone modifications and chromatin accessibility over transcription factor binding, transcription status of Pol II, and nuclear RNA accumulation all the way to traditional polyadenylated RNA accumulation and ribosome association of the mRNA. Genes were classified into 16 clusters with particular attention on hypoxia-induced mRNAs (e.g. ADH1, also HSP70) and "growth-associated" mRNAs (e.g. ribosomal protein, RPs).

One of the stronger conclusions from the paper is that nuclear RNA (nRNA), mostly fully spliced, and polyA RNA can be regulated differentially by hypoxia. Certain genes are hypoxia induced at the nRNA level but not the polyA level, and vice versa (Figure 6, cluster 5-6). This suggests degradation of polyA RNAs under hypoxia, which is reasonable if they have been sequestered in stress granules for a while, or perhaps some hidden control over nuclear polyadenylation.

Another striking finding is that certain genes are hypoxia inducible at the level of RNA PolII Ser2-phosphorylation (i.e. transcription elongation is taking place), but these do not accumulate into polyA RNA etc. until reaeration (cluster 9 genes). Interestingly, this latent transcriptional activation appears to be mediated by Heat Shock Transcription Factors (HSFs).

The classical hypoxia-induced genes are characterized by increased H3K9-acetylation and decreased occupancy by H2A.Z as well as increased Ser2P-RNA. Whereas the RP mRNAs are characterized by a variable and mostly slight increase in H3K27me3, increased nuclear RNA and a variable and generally slight decrease in ribosome associated (TRAP) mRNA. To address the role of ERF-type hypoxia-sensing transcription factors, they performed a CHIP-Seq experiment with a stabilized allele of HRE2. Extending single-gene data in the literature, they find that HRE2 does not bind the classical hypoxia-responsive promoter element but binds genome-wide to a motif known as the EBP box with GCC repeats.

Overall, these data generally match with prior datasets from the Bailey-Serres lab, but provide a dramatically more complex and comprehensive picture. The data are presented and visualized in thoughtful ways. The quality of the data appears to be good (but see note on ATAC-Seq below). Perhaps this is because 'investigator' was not a variable in this work? Publishing these results is important not so much because of the major new insights into the mechanisms of hypoxia response but to document the outcome of this highly innovative approach. One might have imagined that genome-wide data from 11 different stages of gene expression would allow us to understand the precise behavior of every gene to some degree, but this is not the case. I also expected that there would be stronger evidence for translational repression from stress granule formation - this effect appears to be weaker than expected and is only discussed relatively late in the paper. I would like the authors to address the major comments in constructive ways.

RESPONSE: Thank you for your comments on our manuscript. We appreciate the more critical remark about the relationship to stress granules and have tackled this on page 19, lines 553-557.
Point 1. All that said, this paper is not exactly easy to read. Given the complex data, I'm not sure my suggestions for improvement will make a big difference, but I will try: Make sentences shorter, and say it in as plain language as possible with a meaningful verb in active voice.

**RESPONSE:** We have edited the manuscript so that it is more in an active voice. Minor edits of this sort were not marked as changed text.

Point 2. Line 166. On the ATAC-Seq data, a more nuanced interpretation may be warranted. Looking at Figure 2, I do see a 2-fold increase in transposase (hyper)sensitive sites (THS) after 2 hours hypoxia, as stated in the text. However, in response to 9h of hypoxia, the number of THSs dropped compared to the control, and the normoxia control at 9h already has a high number of THSs. What do you make of this? Might the trend during normoxia be due to diurnal regulation?

**RESPONSE:** Over ~30% of the Arabidopsis transcriptome is regulated by the clock, and transcription of phasing genes is influenced by histone modifications. Likely, promoter chromatin accessibility is affected by circadian phasing/diurnal regulation. In mouse liver, chromatin accessibility assayed by DNase-seq revealed that circadian phasing affects ~25% of protein-coding genes, with up to 8-fold changes in chromatin accessibility for individual genes (Sobel et al., 2017). To our knowledge, diurnal regulation of global chromatin accessibility has not been evaluated by DNase-, MNase-, or ATAC-seq in Arabidopsis and was not tested here, except that both 2NS and 9NS controls were performed. We are not confident in concluding that the differences involved are reflective of circadian cycling or the one hour of extended darkness the 9NS experienced to match the 9HS light regime (hypoxia initiated at the end of day when CHO reserves are maximal). We have clarified this in the text (lines 1,342-1,343). Future research is needed to understand if global reduction in chromatin accessibility may be modulated by the clock or cellular energy levels. We feel this is outside of the aims of this manuscript.

Point 3. Fig. 2d shows that the overall trend masks dramatic differences at the single gene level, where hypoxia causes both increases and decreases in transposase accessibility.

**RESPONSE:** This observation is correct and indicates a challenge associated with presentation of this type of data. We hope that readers will make use of a genome browser to view our data. We added: “These examples also illustrate that individual genes demonstrate distinct changes in the chromatin and transcript readouts”, when the genome browser views of Figure 5 are first mentioned. (Lines 244-245)

Point 4. While there is clearly a set of hypoxia-specific THSs, there is an almost equally large set of normoxia-specific sites (Fig. 2f). This should be considered in the interpretation of the data.

**RESPONSE:** Although interesting dynamics exist for normoxia-specific regions of chromatin accessibility, we decided to focus on hypoxia-specific chromatin dynamics. Several ATAC-seq datasets exist for control Arabidopsis seedlings (Maher et al., 2018; Lu et al., 2016; Harris et al., 2018), thus we wanted to focus on novel regions of hypoxia induced chromatin accessibility.

Point 5. Overall, I wonder whether the apparent global increase in transposase accessible sites could be due to a technical reason, perhaps a more efficient transposase reaction overall or more efficient library prep (rather than an actual change at the chromatin level)? Is there some kind of calibration method that would rule this out?

**RESPONSE:** We are not aware of a calibration or QC metric for plant ATAC-seq data, but several measures were taken in an attempt to ensure that the samples were of comparable quality.

1. The quality of libraries and sequencing results was comparable between experiments (average alignment: 85% +/- 3.8% standard deviation).
2. High regions of background variability (reads) were masked prior to normalization and analysis of the data (Supplemental Arabidopsis Blacklisted Chromatin).
3. Following this normalization, the genome-wide intra-genic levels of background were similar for all time points (regions downstream of TSS and upstream of TES). The generally similar distribution of peaks identified from the datasets also suggests that the datasets are of comparable quality.

We analyzed the genomic-wide distribution of chromatin accessibility observed between individual replicates for each time point. Some level of variance was observed between individual replicates, with 9HS demonstrating the most pronounced variation. Based on other publications, these levels of variation were within the acceptable range.
for Arabidopsis ATAC-seq datasets (Harris et al., 2018).

Point 6. Throughout the manuscript, I kept wondering whether it would be justified to collapse some of the 16 gene clusters because they are probably not sufficiently distinct. A smaller number of clusters might make the paper easier to digest with no loss in conclusions.

RESPONSE: We agree with the comment that a smaller number of clusters could make the paper easier to digest. Despite the thoughtfulness of this comment, acting on it would dramatically impact the presentation of the entire manuscript. All text presenting the clusters would change. All diagrams of clusters and the data profile plots would change for the short-term and long-term data. The likely outcome would be that instead of mentioning clusters 1 and 2 together, we would mention only cluster 1. Yet, in our evaluation of the data, these two clusters are subtly different. Moreover, our initial analyses included varying the number of clusters utilized (i.e., 8, 10, 12, 14 clusters). In each of our cluster analyses, we selected the number of clusters based on the informative patterns of gene regulation, GO analysis, profiling of the histone marks (Figure 4D) as well as plots with the same clusters with the 9HS data. Our decision was based on the observation that we resolved differences that were interesting.

Point 7. Specifically, Line 201, also 223 and comment below. Here you are testing the hypothesis: For a gene that is upregulated by hypoxia at the polyA-RNA level, is there evidence that the gene is unresponsive at the nRNA (or Ser2P- or TRAP-) level? In order to test this, I would argue that one should not apply an FDR filter or significance filter to the nRNA response (or Ser2P or TRAP), just a statistically informed fold-change cutoff. I suspect that this reanalysis would classify a lot more genes as co-regulated across multiple scales. For example, looking at Fig. 4c, not only cluster 16, but also clusters 12, 14, and 15 seem to have many coordinately downregulated genes, etc.

RESPONSE: -- The Circos diagrams provide the knowledge that the number of coordinately differentially regulated genes is relatively small.

-- The cluster analysis required a DEG in one of the four comparisons. This is routine practice when comparing conditions or genotypes.

-- As the reviewer suggests, there are moderately more continuously regulated DEGs if we relax the criterion for the remaining categories (e.g. 213 coordinate up-DRGs expanded to 252 by removal of FDR cutoff). The Circos plots, for example, expand and there are more arcs between the four readouts.

Point 8. Figure S9. This figures gives evidence that the 16 clusters are unique. But it also gives evidence that some clusters are similar to each other and should be combined. Did you control for multiple comparisons in the Wilcoxon signed rank test?

RESPONSE: To clarify, the statistical comparisons between clusters in Figure S9 was performed for the H2A.Z, H3K4me3, and H3K9ac modifications for each cluster to determine if histone modifications occur in association with the Ser2P and RNA based readouts for which the clustering was performed for. A similar analysis was performed for the Ser2P and RNA-based readouts for which the clustering was performed against, which revealed that each cluster is significant against at clusters with similar levels of regulation in at least one of these readouts. As requested, we have performed Wilcoxon signed rank sum tests on the clusters to gain support for the decision to use 16 clusters. In response to the reviewer’s request, these plots are provided in Supplemental Figure 9G-N.

Point 9. Line 230. What is the spatial resolution in bp for the CHIP technique? The Methods section does not speak to this. I am surprised that ATAC-Seq and H3K9ac are still increased at -1kb 5’ and +1kb 3’ of the TSS. Is that an artifact of the technique? Where does this signal eventually drop off? Are you really scoring effects on the flanking genes here?

RESPONSE: The spatial resolution of the ChIP is the fragmentation size of the chromatin (200-400bp) used for immunoprecipitation. This information was added to the Methods.

Point 10. Line 226. Ser2P is also increased throughout the -1kb to +1kb region and presumably beyond. This I find very surprising; isn’t Ser2P an RNA-Seq technique; why is there no step function at the transcription start sites?

RESPONSE: The Ser2P data are Ser2P ChIP and not an RNA based method. This is clear in the Methods and also indicated in Figure 1A. Similar to the Histone ChIP, the method captures chromatin of regions flanking any site of polymerase binding. As a result, one does not expect the step function near the TSS. Given that transcription extends beyond the site of cleavage and polyadenylation, it is not unexpected to see 3’ reads in this assay. Our
pipeline counted Ser2P reads that map within coding regions (and not those outside).

We hypothesize that some regions with highly stress-activated genes may become broadly transcriptionally active, perhaps as chromatin loops. Future experiments that combine an RNA-based transcriptional assay such as GRO-seq (reported after we began this work) and RNAPII ChIP and Hi-C could provide further insights into our observations. One additional challenge is that the transcription units of some genes overlap in Arabidopsis. An RNA based method would be required to understand when reads flanking a gene of interest are derived from a different transcription unit.

Point 11. Line 26-30. It may seem obvious now, but still should be clarified: The generalizations in the abstract do not apply to all of the genes in a given group. The qualifier "generally, but not in every case" applies.

RESPONSE: We agree with the reviewer’s conclusion, but struggle with adding “generally but not in every case” for a genomic study. This seems implicit. Nonetheless, we have added “generally” in the abstract and made other modifications to qualify that “the majority” of members of a group display a particular characteristic.

Point 12. I am not well informed what t-SNE does, and it is not explained well in the Results or Methods. However, the t-SNE results in Fig. 1b suggest that the variation between biological replicates is quite substantial as compared to the variation between treatments, especially for the nuclear RNA (nRNA). What can we conclude from the t-SNE? Why not do a more traditional PCA and report PC1 and PC2.

RESPONSE: t-Distributed Stochastic Neighbor Embedding (t-SNE) is a machine-learning algorithm for dimensionality reduction, similar to PCA and MDS. In initial analyses, PCA and MDS were performed for dimensionality reduction, but samples were predominantly separated by assay type due to various biases and backgrounds (ChIP and three distinct methods of RNA-seq) of each particular assay (PC1, > 65% variance explained; PC2, > 12% variance explained). For this particular analysis, t-SNE was utilized to determine local relationships between the various datasets. To perform this analysis, principal components 1-10 from PCA were utilized as input for t-SNE dimensionality reduction. The two-dimension output of t-SNE then allowed for the separation of samples based on condition when several distinct assays were analyzed in combination. This method of analysis was described in the Methods, but additional details have been added in this revision.

Point 13. The TRAP results here for RPs seem to be a bit at odds with prior studies. Do you have an explanation? This could be due to different ways to normalize signal intensities.

RESPONSE: The reviewer is correct that the polysomal data were not normalized, yet reduction in polysomal levels of RPs is evident at the more global scale (see Figure 3A). Our prior research normalized for the reduction in polysome levels in the tissue due to a consistent 45% reduction in ribosomes translating mRNAs after 2HS in our system [Branco-Price et al., 2008; Juntawong et al., 2014]. We did not perform any normalization here to reduce the complexity of the analysis and because one cannot correctly normalize when TRAP or other cell-specific method is applied. We have added a comment in the Methods about this point.

Reviewer #3:

In this ms, Lee and Bailey-Serres performed an accurate survey of the epigenome by studying the hypoxic translatome of Arabidopsis. Seedlings were treated with short (2h) and long (9h) term hypoxia. The 2h hypoxia was also followed by a recovery by re-oxygenation. The authors present a large amount of data. Given the complexity of the figures, the interpretation of the results is not always easy to follow. There is no doubt, however, that these datasets will be of great interest to the scientific community, also beyond scientists interested in plant responses to hypoxia.

While most of the data are presented using an 'omic' representation of results, other figures (e.g. Fig. 4C; Fig 5; Fig. 8) provide a very clear representation of the regulation of gene expression under hypoxia. The authors show that a cluster of genes that includes several of the core anaerobic genes is upregulated by hypoxia and its transcripts are also preferentially translated because of increased chromatin accessibility, RNAPII engagement and reduced Histone 2A.Z association. Remarkably HSP genes are transcribed, but probably translated only during re-
oxygenation. This latter result is in line with the reported increase in the mRNA level of several HSP genes without a corresponding increase in the relative protein level as detected by western blot (Banti et al., 2010).

Point 1. A few aspects are unclear in Fig. 9 (which, overall, is very useful). Why are translated mRNAs absent in the Coordinated Upregulated HRG graphics at 9h hypoxia? I could not find TRAP data at 9HS.

RESPONSE: We did not include the translation state data for the 9HS time point because it was not performed in this study. It was, however, performed with plants grown in the same chambers and deprived of oxygen in a nearly identical set-up by a prior lab member (Branco-Price et al., 2008). That analysis was performed using polyA and TRAP RNA and the Affymetrix GeneChips, rather than RNA-sequencing (8,978 genes).

Point 2. Which is the mechanism inducing degradation of HRG mRNAs during re-oxygenation (Fig. 9)?

RESPONSE: The mechanisms of mRNA degradation under hypoxic stress and reoxygenation are presently not known but are a topic for future study. Several mechanisms of mRNA degradation exist that likely contribute to the post-transcriptional regulation of hypoxia-induced mRNAs upon re-oxygenation. We did not want to speculate on this mechanism of regulation at this time, as this is a goal for future research.

Point 3. Data in Fig. 8 indicate that re-aeration has a different impact on PCO2 and ADH1 (see difference in PolyA and TRAP between PCO2 and ADH1 during R). Why? Given the possible role of PCOs in the oxidation of Cys in ERF-VII proteins during re-oxygenation, I would have expected that PCO2 was still associated with polysomes during re-aeration. A comment on this would be useful.

RESPONSE: Thank you for your perceptive observation. We also noted this distinct regulation of PCO2 and ADH1 mRNAs upon re-oxygenation; a similar pattern of regulation is also observed for PCO1 (viewable in the jbrowse session built for this data: https://cluster.hpc.ucr.edu/~tlee012/JBrowse-1.16.6/). The role of the PCOs in ERF-VII modification and turnover is very convincing (Weits et al., 2014; White et al., 2017), yet to the best of our knowledge there remains opportunity to evaluate the temporal dynamics of their accumulation and function during re-aeration. The recent study on ADOs in mammals suggests that the cysteine dioxygenases that oxidize N-terminal cysteines to create N-degrons are active during reaeration (Masson et al., 2019). PCO1/2-mediated modification of the ERF-VII N-terminal cysteine is likely to rapidly increase upon re-oxygenation (less than 1 h) due to the presence of the enzyme and as evidenced by the rapid degradation of N-terminal MC-containing peptides in response to re-oxygenation following hypoxia (Licausi et al., 2011). Our data point to destabilization of the PCO2 transcript during re-aeration. Given that the different PCOs of Arabidopsis have distinct in vitro biochemical properties and preferences for the different ERFVIIIs (White et al., 2018), one can speculate that the decline in PCO2 transcript may decline rapidly because its activity is tailored to HRE1 and HRE2 rather RAP2.12 and RAP2.3. It would be of value to look at the dynamics of the protein and its target(s) in greater detail, but anticipate this will be done by others. We do plan to evaluate mRNA decay kinetics to address mechanisms of dynamics of turnover.

The continued abundance and ribosome association of ADH1 mRNA upon re-oxygenation may reflect a more sluggish restoration of metabolic homeostasis. Regulation of Pyruvate Decarboxylase enzyme levels/activity would provide a more effect means to limit fermentation. Again, we feel mRNA turnover in the context of oxygen deprivation is an important topic for future study.

We are pleased to inform you that your paper entitled "Integrative analysis from the epigenome to translatome exposes patterns of dominant nuclear regulation during transient stress" has been accepted for publication in The Plant Cell, pending a final minor editorial review by journal staff. At this stage, your manuscript will be evaluated by a Science Editor with respect to scientific content presentation, compliance with journal policies, and presentation for a broad readership.

Final acceptance from Science Editor

Sept. 12, 2019