nature portfolio

Peer Review File

# DNA methylation reprogramming of gene regulatory elements is associated with lung metastasis in triple-negative breast cancer

Corresponding Author: Dr Diego Marzese

**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 interesting manuscript focuses on the differences in DNA methylation between primary tumors and lung metastases in triple-negative breast cancer (TNBC).

Using a xenograft model of MDA-MB-231, the authors compared genome-wide DNA methylation profiles of primary tumors, lymph node metastases, and lung metastases, identifying differentially methylated sites (DMS). They demonstrated that hypomethylated regions were more abundant in lung metastases than in primary tumors or lymph node metastases. Furthermore, they identified a group of genes that harbored at least two DMS within promoter regions specifically in lung metastases, and showed that these genes are involved in invasion and metastasis. To strengthen these findings, the authors reanalyzed publicly available DNA methylation and RNA-seq data from 19 TNBC patient samples in the AURORA US cohort. They defined a core gene set of 22 genes as the overlap between (i) genes with promoter methylation changes accompanied by altered expression in the AURORA US data and (ii) genes with promoter methylation changes in the xenograft model. Some of these core genes were further correlated with lung metastasis-free survival in additional datasets.

To my knowledge, reports describing genome-wide DNA methylation differences between primary tumors and lung metastases in TNBC are scarce, and the findings of this manuscript are worthy of publication. The experimental system is simple and straightforward, the methodology is generally appropriate, and the manuscript is easy to follow. However, the association between metastasis and hypomethylation has already been reported in other cancer types, and the level of novelty in this regard is not necessarily high. Moreover, in my view, the interpretations of the data presented are, in several places, overstated or inaccurate, which diminishes the overall impact of the manuscript.

My specific concerns are listed below.

Major points:
1. I understand the main finding of this manuscript to be that “the genome-wide DNA methylation landscape differs between primary tumors and lung metastases.” However, no data are presented to demonstrate that this observation functionally contributes to the promotion of lung metastasis. While the authors show that the set of genes with promoter methylation changes is significantly enriched for metastasis-related genes, this correlation alone is insufficient to support the manuscript’s conclusion, as reflected in the title, “Epigenetic programs shaping lung metastasis in triple-negative breast cancer.” Throughout the manuscript, the authors treat correlational findings as evidence of causation, and this interpretation should be carefully revised.

2. The authors frequently use the term “epigenome.” However, the only epigenomic data presented in this manuscript are DNA methylation profiles. Previous studies have shown that, in breast cancer cells (HCC1954), there is widespread DNA hypomethylation, but most genes located in hypomethylated regions are transcriptionally silent and associated with repressive histone marks (doi: 10.1101/gr.125872.111). In the present manuscript, the number of genes with increased expression is extremely limited compared to the large number of hypomethylated regions identified in lung metastases. This suggests that complex epigenomic regulation beyond DNA methylation alone is likely involved. The authors should revise the results and discussion to reflect that epigenetic regulation of gene expression is not solely determined by DNA

methylation, and use the term “epigenome” with caution.

3. The absence of transcriptomic data from the xenograft model is, as the authors themselves state in the Discussion, a major limitation of this study. While I understand that obtaining multi-omic data from the same samples can be challenging, if RNA from the same specimens is available, validating the expression of several key genes by qPCR would significantly strengthen the manuscript. If this is not feasible, I strongly recommend that the statement in the Abstract—“To address this gap, we performed comprehensive multi-omics analysis, integrating epigenomic and transcriptomic profiles from TNBC xenograft models and patient cohorts” (lines 42–44)—be revised to a more accurate description to avoid misleading readers.

4. The authors should provide supplementary tables for the following results:
(a) the list of genes with promoter DMS in the xenograft model (lines 159–162),
(b) the corresponding list from the AURORA US dataset (lines 209–211),
(c) the overlap between these sets (Figure 3e),
(d) differential expression analysis results from the AURORA US cohort (lines 233–234).

Minor points:
5. The description of H&E staining in the xenograft model (“Representative H&E staining confirmed the histological features of primary tumors and metastatic sites”) would be strengthened by briefly specifying what histological features were observed. This would aid readers who are experts in epigenomics but not necessarily in breast cancer pathology.

6. For the analysis of the relationship between DMS and chromatin states (lines 139–146; Supplementary Fig. 2b-c), the Methods section indicates that chromatin state information was obtained from the UCSC Genome Browser. To avoid confusion, this should also be briefly mentioned in the Results section.

7. The cutoff criteria used for defining DNA methylation and gene expression changes should be explicitly stated in the main text.

8. The datasets used for survival analysis should be properly cited not only in the Methods section but also in the Results section (line 270).

This manuscript reports interesting observations; however, the results are described in an overstated manner. Although associations between hypomethylation and metastasis have been reported—for example, hypomethylation of stemness-related genes in circulating tumor cells from advanced breast cancer patients (doi: 10.1016/j.cell.2018.11.046)—detailed reports in TNBC and lung metastasis are limited, and the data presented here do have merit. I encourage the authors to adopt a more precise and balanced interpretation of their findings, which I believe would substantially improve the quality and impact of the manuscript.

Reviewer #2

(Remarks to the Author)
In this manuscript, Bedoya-López and colleagues combine DNA methylation profiling from a xenograft metastasis series with re-analyses of public patient cohorts, i.e. AURORA, to propose that lung metastases in TNBC exhibit a distinct methylation program and to nominate candidate genes/pathways with potential prognostic relevance. The work is clearly written and the analyses are generally coherent, and the attempt to pinpoint findings across mouse model and patient data is appreciated. However, there are major concerns about the incremental value and translational relevance on the current form of the study.

Major concerns

1. The study is largely descriptive, focusing on differential methylation and enrichment/overlap analyses, with limited causal evidence that the identified methylation changes mechanistically drive transcriptional programs or metastatic phenotypes. Strengthening the work would require clearer functional experimental verification rather than primarily signature-level interpretation.

2. The study’s only new experimental system appears to be a single TNBC cell line (MDA-MB-231) xenograft metastasis cascade, whereas much of the translational/clinical interpretation relies on AURORA patient cohorts. As presented, it is not sufficiently clear what unique inference the xenograft provides beyond what is already accessible directly from patient multi-omics (methylation + RNA). Given the known limitations of single-line xenografts (clonal idiosyncrasies, immunodeficient context, potential purity), the authors should explicitly justify why this model is necessary and how it strengthens the conclusions.

3. The manuscript integrates methylation changes with RNA-seq signals primarily using patient datasets and then presents overlap/candidate genes, e.g., the 22 gene intersection. However, the matched RNA-seq from the xenograft samples is missing. It is difficult to justify how the xenograft methylation alterations translate into transcription. This makes the cross-system comparison (xenograft methylome vs patient transcriptome) harder to interpret and increases the risk that model-specific methylation signals are being projected onto real patient biology.
Notably, the need to filter down to a small overlap (22 genes) suggests that methylation signals do not broadly predict transcription in a straightforward way, and therefore the xenograft discovery may contribute limited incremental value.

4. The authors pool primary + LN metastasis samples and compare them against distant lung metastases, citing low sample numbers and similarity between primary and regional metastases. This seems to be a pragmatic choice but under-justified. Even if the PT–LN contrast yields fewer DMS than LU-involving contrasts, that does not establish that PT and LN are interchangeable; hundreds to thousands of loci can still carry biologically meaningful regional dissemination signals. Prior literatures support the view that regional metastases can exhibit non-trivial methylation remodeling relative to primaries (e.g., Reyngold et al., PMID: 25083786; Urrutia et al., PMID: 25628026; Barekati et al., PMID: 22695536; Mathe et al., PMID: 27671774). Pooling PT and LN could therefore mask stepwise progression patterns or confound interpretation of "lung-specific" changes.
The authors should perform sensitivity analyses showing that key DMS sets remain consistent when analyzing LU vs PT and LU vs LN separately (in addition to the pooled LU vs PT+LN contrast) for justification.

5. In the introduction, the authors implied that the epigenetic landscape of TNBC metastasis is broadly "poorly characterized." Given the existence of prior methylation profiling studies comparing primary vs metastasis (including matched regional metastases mentioned in point 3) and large multi-omics analysis on AURORA (PMID: 36585450), the manuscript should more precisely state what is genuinely new here.

Minor points
1. In Figure 3e, the manuscript text states 284 shared genes, while the venn diagram shows 285 in the overlap.

2. For the Kaplan–Meier analyses, please specify how "high vs low" expression cutoffs were chosen, e.g. median expression?

3. It will be appreciated if the enriched methylation loci in open sea regions can be functionally linked/interpreted to chromatin states/enhancers where possible.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I would like to thank the authors for their extensive efforts in revising the manuscript. The overall tone is now well-balanced and appropriately cautious. The re-analysis utilizing various public datasets not only strengthens the authors' original claims but also provides novel insights, such as through the integrated analysis with histone modifications. Expanding the scope of the analysis to enhancer regions has also added significant depth to the paper. Furthermore, the numerous newly provided supplementary tables in the revised manuscript will serve as valuable resources for future research on DNA methylation and broader epigenomics in TNBC lung metastasis.

While the authors have addressed most of my previous concerns sufficiently, I would like them to consider the following minor points before final publication.
(1) Supplementary Figure 4a:
I understand that this heatmap illustrates expression correlations used to narrow down candidate genes for qPCR validation. However, it is currently unclear from the figure itself, the figure legend, or the main text what the underlying data source is and what the color scale represents. Although the rebuttal letter clarifies that the AURORA dataset was used, adding these details to the figure legend would greatly enhance readability. If feasible, including a representative scatter plot with a fitted regression line for a key gene pair would further facilitate the reader's understanding.
(2) Title:
Although the revised title is more appropriately cautious than the previous version, it still feels somewhat broad or vague, resembling the title of a review article. That being said, the abstract clearly conveys the core findings and context of the study. Therefore, if the authors and the editor are satisfied with the current title, I have no further objections.

Reviewer #2

(Remarks to the Author)
The revised manuscript is substantially improved. In particular, the revised title and overall framing are more appropriate. The additional analyses, including separate LU vs PT and LU vs LN sensitivity, qPCR validation, survival analysis, and additional supplementary tables, have greatly strengthened the manuscript. Overall, most of my previous concerns have been adequately addressed. However, I still like to recommend one minor text change before publication.
Although the authors state that causal interpretations between DNAm and metastasis have been removed, in the result section, line 364-367, the statement that the data "support a functional role for these genes beyond metastatic adaptation" remains, I feel, too strong for the provided evidence. I suggest revising this to a more cautious statement, such as "further nominates these genes as candidates for future functional validation."

**Point-by-point response to the reviewers' comments on the manuscript**

We thank the reviewers for their constructive comments, which have helped us to improve the clarity and rigor of the manuscript. In response, we have significantly revised the manuscript to more accurately reflect the scope of our findings, incorporated additional analyses, and expanded the discussion to contextualize the results. All new analyses are also described in the Methods and Results sections, with several rearrangements made to facilitate a comprehensive understanding.

Below, we provide a detailed, point-by-point response to each comment. Reviewers' comments are shown in **bold**, and our responses are provided in regular text. Substantial textual changes in the revised manuscript, including figure footnotes, are indicated in **blue (referenced as follows: Line XX-YY)**. All line numbers refer to the revised manuscript.

**Reviewer #1 (Remarks to the Author):**

**This interesting manuscript focuses on the differences in DNA methylation between primary tumors and lung metastases in triple-negative breast cancer (TNBC).**

**Using a xenograft model of MDA-MB-231, the authors compared genome-wide DNA methylation profiles of primary tumors, lymph node metastases, and lung metastases, identifying differentially methylated sites (DMS). They demonstrated that hypomethylated regions were more abundant in lung metastases than in primary tumors or lymph node metastases. Furthermore, they identified a group of genes that harbored at least two DMS within promoter regions specifically in lung metastases, and showed that these genes are involved in invasion and metastasis. To strengthen these findings, the authors reanalyzed publicly available DNA methylation and RNA-seq data from 19 TNBC patient samples in the AURORA US cohort. They defined a core gene set of 22 genes as the overlap between (i) genes with promoter methylation changes accompanied by altered expression in the AURORA US data and (ii) genes with promoter methylation changes in the xenograft model. Some of these core genes were further correlated with lung metastasis-free survival in additional datasets.**

**To my knowledge, reports describing genome-wide DNA methylation differences between primary tumors and lung metastases in TNBC are scarce, and the findings of this manuscript are worthy of publication. The experimental system is simple and straightforward, the methodology is generally appropriate, and the manuscript is easy to follow. However, the association between metastasis and hypomethylation has already been reported in other cancer types, and the level of novelty in this regard is not necessarily high. Moreover, in my view, the interpretations of the data presented are, in several places, overstated or inaccurate, which diminishes the overall impact of the manuscript.**

**My specific concerns are listed below.**

**Major points:**

**1. I understand the main finding of this manuscript to be that "the genome-wide DNA methylation landscape differs between primary tumors and lung metastases." However, no data are presented to demonstrate that this observation functionally contributes to the promotion of lung metastasis.**

**While the authors show that the set of genes with promoter methylation changes is significantly enriched for metastasis-related genes, this correlation alone is insufficient to support the manuscript's conclusion, as reflected in the title, "Epigenetic programs shaping lung metastasis in triple-negative breast cancer." Throughout the manuscript, the authors treat correlational findings as evidence of causation, and this interpretation should be carefully revised.**

**Response to reviewer:** We thank the reviewer for this critical comment and fully agree that our study does not provide direct functional evidence demonstrating that DNA methylation changes causally drive lung metastasis. In response, we have substantially revised the manuscript to eliminate any causal interpretation of our findings. Specifically, we have modified the title, abstract, and relevant sections of the Results and Discussion to consistently frame our observations as associations between DNA methylation patterns and metastatic progression, rather than evidence of causation.

We explicitly present our manuscript as a characterization of DNA methylation alterations associated with lung metastasis in TNBC, and as a resource for identifying candidate regulatory regions and genes that may contribute to metastatic phenotypes. We have also incorporated additional text in the Discussion to clearly acknowledge that functional validation will be required to establish causality. Importantly, we have revised statements that previously implied mechanistic roles, for example: "epigenetic programs shaping metastasis" to more accurate descriptions reflecting the descriptive and integrative nature of the study. We believe these changes address the reviewer's concern and result in a more precise and balanced interpretation of our findings.

**2. The authors frequently use the term "epigenome." However, the only epigenomic data presented in this manuscript are DNA methylation profiles. Previous studies have shown that, in breast cancer cells (HCC1954), there is widespread DNA hypomethylation, but most genes located in hypomethylated regions are transcriptionally silent and associated with repressive histone marks (doi: 10.1101/gr.125872.111). In the present manuscript, the number of genes with increased expression is extremely limited compared to the large number of hypomethylated regions identified in lung metastases. This suggests that complex epigenomic regulation beyond DNA methylation alone is likely involved. The authors should revise the results and discussion to reflect that epigenetic regulation of gene expression is not solely determined by DNA methylation, and use the term "epigenome" with caution.**

**Response to reviewer:** We thank the reviewer for this important observation and agree that the term "epigenome" was used too broadly in the original manuscript, given that our data are limited to DNA methylation profiling. We have revised the terminology throughout the manuscript to more accurately reflect the scope of our analyses, avoiding generalized use of the term "epigenome" where not appropriate. We also agree that DNA methylation alone does not fully explain transcriptional regulation. To address this, we expanded our analyses to incorporate chromatin context. Specifically, we integrated publicly available ChIP-sequencing data (GSE85158) profiling key histone marks (H3K4me3, H3K27ac, H3K9ac, H3K9me3, H3K27me3) from MDA-MB-231 cells, the TNBC cell line used to generate our xenograft model (Methods: Line 591-595).

Based on these data, we classified promoter regions into active, repressed, bivalent, and inactive chromatin states (Methods: Line 609-622), and validated these annotations using independent chromatin accessibility data (ATAC-sequencing; E-MTAB-12821; **Supplementary Figure 3d**).

Using this framework, we found that the relationship between DNA methylation and gene expression is strongly dependent on chromatin context (Results: Line 254-266). Promoter hypomethylation was associated with increased gene expression primarily in regions marked by active histone modifications, whereas genes linked to repressive or bivalent chromatin states showed limited transcriptional response despite DNA methylation loss (**Supplementary Figure 3e-f**). These results support a context-dependent model in which DNA methylation acts in conjunction with pre-existing chromatin states rather than as a sole determinant of gene expression.

In addition, given the limited number of genes directly associated with promoter methylation changes, we hypothesized that other gene regulatory elements would be affected by these alterations. Thus, by leveraging GeneHancer annotations (Methods: Line 603-607), we identified 224 differentially methylated enhancers associated with 226 dysregulated genes across our xenograft model and the AURORA US dataset (Results: Line 296-319; **Figure 5a-b**). Notably, enhancer-associated methylation changes showed a stronger relationship with gene expression than promoter-based analyses. Similar to promoters, the functional impact of DNA methylation at enhancers was also dependent on chromatin context, with hypomethylation in active regions associated with gene upregulation, whereas repressive contexts showed limited or inverse transcriptional effects (**Figure 5c**). These findings further support a model in which DNA methylation changes contribute to transcriptional regulation in a context-dependent manner, rather than being a definitive determinant.

For instance, these patterns were exemplified by previously characterized lung metastasis-associated genes such as *ID1*, whose role in lung metastatic colonization has been consistently demonstrated across multiple studies (*Gupta et al., PNAS 2007*; *Minn et al., Nature 2005)*. In our study, the active enhancer regulating *ID1* expression was consistently hypomethylated in lung metastases (**Figure 5d**), accompanied by increased *ID1* expression as confirmed by targeted qPCR (**Figure 5e**). Although this represents an illustrative example, similar regulatory reprogramming may affect additional genes (**Supplementary Table 7**).

Finally, we have revised the Discussion to explicitly acknowledge that transcriptional regulation in this context cannot be attributed to DNA methylation alone but reflects the interplay between multiple epigenetic layers (Discussion: Line 409-423).

**3. The absence of transcriptomic data from the xenograft model is, as the authors themselves state in the Discussion, a major limitation of this study. While I understand that obtaining multi-omic data from the same samples can be challenging, if RNA from the same specimens is available, validating the expression of several key genes by qPCR would significantly strengthen the manuscript. If this is not feasible, I strongly recommend that the statement in the Abstract—"To address this gap, we performed comprehensive multi-omics analysis, integrating epigenomic and transcriptomic profiles from TNBC xenograft models and patient cohorts" (Liness 42–44)—be revised to a more accurate description to avoid misleading readers.**

**Response to reviewer:** We thank the reviewer for this important comment and agree that the original version of the manuscript could be misleading regarding the extent of multi-omic integration in the xenograft model. We acknowledge that matched transcriptomic data were not generated for the xenograft samples, which represents a limitation of the study. In response, we have revised the Abstract and relevant sections of the manuscript to more accurately describe the scope of our analyses. Specifically, we now clarify that the integrative multi-omics framework is primarily based on patient-derived datasets (AURORA US cohort), while the xenograft model contributes genome-wide DNA methylation profiling.

To partially address the lack of transcriptomic data in the xenograft model, we performed targeted gene expression validation using qPCR on matched primary tumor and lung metastasis samples (Methods: Line 547-Line 567), and primers were designed to amplify short amplicons (60–90 bp; **Supplementary Table 9**). Due to the limited and partially degraded RNA obtained from FFPE sections, we focused on a selected set of candidate genes based on their expression patterns in the AURORA dataset. This analysis revealed distinct expression clusters (**Supplementary Figure 4a**), from which representative genes were chosen for validation, prioritizing robustly expressed transcripts. The final panel included *FKBP4*, *TPI1*, *CAPS*, *MEIS2*, *LYNX1*, and *AK1*, with at least one gene validated per cluster (except for *CD5*; Results: Line 267-268). Gene validation for *CD5* and *HOXA5* was also considered, but could not be reliably detected, likely due to low transcript abundance or RNA degradation.

Across models, several genes showed consistent upregulation in lung metastases relative to primary tumors, although the magnitude of change varied between samples (**Figure 4c**; Results: Line 271-274). Additionally, we validated the expression of *ID1*, a gene with well-established relevance in lung metastatic colonization, which showed consistent upregulation across datasets and in our model (**Figure 5e-f**; Results: Line 326-329). While these analyses do not replace transcriptomic profiling, they provide support for the transcriptional variations of selected candidates. Finally, we have revised the manuscript to explicitly acknowledge this limitation and to avoid implying that transcriptomic data were generated from the xenograft samples (Discussion: Line 457-460). We believe these changes improve the clarity and accuracy of the manuscript and address the reviewer's concerns.

**4. The authors should provide supplementary tables for the following results:**

**(a) the list of genes with promoter DMS in the xenograft model (Liness 159–162),**

**(b) the corresponding list from the AURORA US dataset (Liness 209–211),**

**(c) the overlap between these sets (Figure 3e),**

**(d) differential expression analysis results from the AURORA US cohort (Liness 233–234).**

**Response to reviewer:** We agree that clear and transparent reporting of these results is essential. In response, we have expanded the supplementary material to include the requested datasets. Specifically:

(a) The list of genes with promoter DMS in the xenograft model is now provided as new **Supplementary Table 1**.

(b) The corresponding list from the AURORA US dataset is included as new **Supplementary Table 3**.

(c) The overlap between these gene sets is reported as new **Supplementary Table 4**.

(d) Differential gene expression analysis results from the AURORA US cohort are now provided as new **Supplementary Table 5**.

**Minor points:**

**5. The description of H&E staining in the xenograft model ("Representative H&E staining confirmed the histological features of primary tumors and metastatic sites") would be strengthened by briefly specifying what histological features were observed. This would aid readers who are experts in epigenomics but not necessarily in breast cancer pathology.**

**Response to reviewer:** We thank the reviewer for this suggestion. We have revised the manuscript to provide additional detail on the histological features used to identify primary tumors and metastatic lesions in the xenograft model. Specifically, metastatic regions were defined based on established histopathological criteria, including disruption of normal tissue architecture, invasive growth patterns, increased cellular density, nuclear atypia, and elevated nuclear-to-cytoplasmic ratios, consistent with breast cancer metastases in distant organs. The description accompanying the H&E images has been updated accordingly (Results: Line 119-122).

**6. For the analysis of the relationship between DMS and chromatin states (Liness 139–146; Supplementary Fig. 2b-c), the Methods section indicates that chromatin state information was obtained from the UCSC Genome Browser. To avoid confusion, this should also be briefly mentioned in the Results section.**

**Response to reviewer:** We thank the reviewer for this comment and agree that the previous description could lead to confusion. In the revised manuscript, we have replaced the initial analysis based on chromatin state annotations obtained from the UCSC Genome Browser with a more context-specific approach. Specifically, we now integrate ChIP-sequencing profiles of key histone modifications derived from the MDA-MB-231 TNBC cell line, as described in the Methods (Methods: Line 591-622).

This updated strategy provides a more biologically relevant framework to interpret the relationship between DNA methylation changes and chromatin context. The resulting chromatin state annotations are now explicitly incorporated into the Results (Results: Line 254-266, 313-319) and further discussed in the Discussion section (Discussion: Line 409-423), with supporting visualizations provided in **Supplementary Figure 3d-f and Figure 5b-d**.

**7. The cutoff criteria used for defining DNA methylation and gene expression changes should be explicitly stated in the main text.**

**Response to reviewer:** Thanks for this observation. In our revised manuscript, we have explicitly stated the cutoff criteria used to define DNA methylation and gene expression changes in the Results section (Results: Line 134, 150, 203, 247-248).

**8. The datasets used for survival analysis should be properly cited not only in the Methods section but also in the Results section (Lines 270).**

**Response to reviewer:** The datasets used for survival analysis are now properly cited not only in the Methods section but also in the Results section (Results: Line 342). In addition, we have expanded the analyzed cohort from 88 to 173 patients to further improve the robustness of the results. More details can be found in the results section (Results: Line 336-367) and their associated methods (Methods: Line 579-581, 649-665).

**This manuscript reports interesting observations; however, the results are described in an overstated manner. Although associations between hypomethylation and metastasis have been reported—for example, hypomethylation of stemness-related genes in circulating tumor cells from advanced breast cancer patients (doi: 10.1016/j.cell.2018.11.046)—detailed reports in TNBC and lung metastasis are limited, and the data presented here do have merit. I encourage the authors to adopt a more precise and balanced interpretation of their findings, which I believe would substantially improve the quality and impact of the manuscript.**

**Response to reviewer:** We hope that the updated version of the manuscript fulfills the reviewer's suggestions. From our side, we thank the reviewer for the time invested. We believe that his/her comments, together with the newly performed analyses and the implemented changes, improved the overall quality of the manuscript and provided a more reliable interpretation of our data.

**Reviewer #2 (Remarks to the Author):**

**In this manuscript, Bedoya-López and colleagues combine DNA methylation profiling from a xenograft metastasis series with re-analyses of public patient cohorts, i.e. AURORA, to propose that lung metastases in TNBC exhibit a distinct methylation program and to nominate candidate genes/pathways with potential prognostic relevance. The work is clearly written and the analyses are generally coherent, and the attempt to pinpoint findings across mouse model and patient data is appreciated. However, there are major concerns about the incremental value and translational relevance on the current form of the study.**

**Major concerns**

**1. The study is largely descriptive, focusing on differential methylation and enrichment/overlap analyses, with limited causal evidence that the identified methylation changes mechanistically drive transcriptional programs or metastatic phenotypes. Strengthening the work would require clearer functional experimental verification rather than primarily signature-level interpretation.**

**Response to reviewer:** We thank the reviewer for this important comment and agree that the original version of the manuscript relied primarily on descriptive analyses and did not provide direct functional evidence linking DNA methylation changes to metastatic phenotypes. We have revised the manuscript to remove causal interpretations and to more clearly present our findings as associations supported by multiple complementary lines of evidence. While our study does not establish direct mechanistic causality, we have strengthened the functional relevance of our findings through the following additions:

- **Gene expression validation:** We performed targeted qPCR evaluation of key genes identified in our cross-cohort analyses. Notably, we observed specific transcriptional changes between primary tumors and lung metastases (**Figure 4c**, **Figure 5e;** Results: Line 271-274, 325-329), supporting the relevance of our findings. Moreover, treatment with 5-azacytidine in TNBC cells supported the mechanistic role of DNA methylation in regulating the expression of these genes (**Supplementary Figure 4c**, **Figure 5g;** Results: Line 279-283, 330-332). While these experiments do not demonstrate full causality, they provide orthogonal evidence supporting the transcriptional relevance of the identified methylation changes.

- **Analysis of distal regulatory regions:** We expanded our analyses beyond promoter regions to include distal regulatory elements (Results: Line 296-334). In the revised manuscript, we show that TNBC lung metastases underwent widespread DNA methylation remodeling at enhancer regions (**Figure 5a**), which correlated with gene expression changes (**Figure 5b**, **Supplementary Table 7**)**.** As a representative example, we describe alterations associated with *ID1*, a gene previously implicated in the transition from primary tumors to lung metastases. While promoter DNA methylation at *ID1* remained unchanged, associated enhancer regions were markedly hypomethylated (**Figure 5d**). This hypomethylation was accompanied by increased *ID1* expression in our model and in independent datasets (**Figure 5e-f**), suggesting that distal regulatory elements constitute a key layer of gene regulation in the lung-adaptive transcriptional program.

- **Integration of chromatin context:** Following the reviewers' suggestions, we incorporated chromatin context into the interpretation of DNA methylation changes. Using publicly

available ChIP-sequencing (histone marks) and ATAC-sequencing data from TNBC cells, we classified regulatory regions as active (H3K4me3, H3K27ac, H3K9ac), repressed (H3K27me3, H3K9me3), or bivalent if both marks were present. Regions lacking any mark were classified as inactive at the basal state and were not considered in further analyses (**Supplementary Figure 3d**; Methods: Line 591-622). Importantly, we found that the relationship between DNA methylation and gene expression was strongly dependent on chromatin state (Results: Line 254-266). Promoter hypomethylation was associated with gene upregulation primarily in regions enriched for active marks, whereas genes linked to bivalent or repressive states showed limited transcriptional response despite methylation loss (**Supplementary Figure 3e–f**). A similar, and even more pronounced, pattern was observed at enhancer regions (**Figure 5c**). These results indicated that DNA methylation alone was insufficient to explain transcriptional changes and instead acted within a pre-existing chromatin landscape that constrained or enabled gene activation.

- **Expanded lung metastasis-free survival analysis:** We expanded our search for publicly available datasets generated using comparable platforms, increasing the cohort size from 88 to 173 patients (Methods: Line 579-581, 649-665). Notably, a subset of the identified genes showed consistent associations with lung dissemination risk when dysregulated in primary tumors (**Supplementary Table 8**), supporting the notion that the transcriptomic alterations observed in metastases may have originated earlier during tumor evolution (Results: Line 342-367).

**2. The study's only new experimental system appears to be a single TNBC cell line (MDA-MB-231) xenograft metastasis cascade, whereas much of the translational/clinical interpretation relies on AURORA patient cohorts. As presented, it is not sufficiently clear what unique inference the xenograft provides beyond what is already accessible directly from patient multi-omics (methylation + RNA). Given the known limitations of single-line xenografts (clonal idiosyncrasies, immunodeficient context, potential purity), the authors should explicitly justify why this model is necessary and how it strengthens the conclusions.**

**Response to reviewer:** We thank the reviewer for this important comment and agree that the use of a single TNBC xenograft model has inherent limitations, including its clonal origin and immunodeficient context. In the revised manuscript, we have clarified the specific role of the xenograft system within our study design. Rather than serving as a standalone translational model, the xenograft provides a controlled framework in which DNA methylation differences between primary tumors and matched metastatic sites can be assessed within a consistent genetic background and reduced microenvironmental variability. We acknowledge that patient-derived datasets offer broader clinical relevance and multi-omic integration. Accordingly, we have revised the manuscript to explicitly present the xenograft as a discovery platform for identifying candidate DNA methylation alterations, which are subsequently evaluated for consistency and relevance using independent patient cohorts (AURORA US dataset).

We have also expanded the Discussion to more explicitly address the limitations of this model, including its single-cell-line origin, lack of immune context, and potential model-specific effects

(Discussion: Line 453-462). We believe this revised framing more accurately reflects the complementary roles of the experimental model and patient-derived data in our study.

**3. The manuscript integrates methylation changes with RNA-seq signals primarily using patient datasets and then presents overlap/candidate genes, e.g., the 22 gene intersection. However, the matched RNA-seq from the xenograft samples is missing. It is difficult to justify how the xenograft methylation alterations translate into transcription. This makes the cross-system comparison (xenograft methylome vs patient transcriptome) harder to interpret and increases the risk that model-specific methylation signals are being projected onto real patient biology.**

**Notably, the need to filter down to a small overlap (22 genes) suggests that methylation signals do not broadly predict transcription in a straightforward way, and therefore the xenograft discovery may contribute limited incremental value.**

**Response to reviewer:** We thank the reviewer for this insightful comment and agree that the absence of matched genome-wide transcriptomic data in the xenograft model represents an important limitation. In the revised manuscript, we have clarified that the integration of DNA methylation and gene expression data is primarily based on patient-derived datasets, while the xenograft model contributes genome-wide DNA methylation profiling.

To mitigate the limitations of cross-system comparisons, we applied stringent filtering criteria, including requiring consistent methylation patterns across datasets and prioritizing genes with multiple DMS within promoter regions. We have revised the manuscript to explicitly acknowledge that this approach may retain only a subset of biologically relevant signals.

To provide additional support for the transcriptional relevance of selected candidates, we performed targeted qPCR validation in the xenograft model using matched primary tumor and lung metastasis samples (Methods: Line 547-567). Primers designed to generate short amplicons (60–90 bp) are described in **Supplementary Table 9**. Due to limited sample material, candidate genes were prioritized based on their expression profiles in the AURORA US dataset. This approach identified distinct expression clusters (**Supplementary Figure 4a**), from which representative genes were selected for validation. The final panel comprised *FKBP4*, *TPI1*, *CAPS*, *MEIS2*, *LYNX1*, and *AK1*. *CD5* and *HOXA5* were also evaluated but could not be consistently detected, likely owing to low transcript abundance or RNA degradation. This analysis demonstrated concordant upregulation for these genes, supporting the biological relevance of our model system (**Figure 4c**). While not equivalent to genome-wide transcriptomic profiling, these analyses confirm that several key candidates identified through cross-cohort integration are transcriptionally engaged in the model system.

Importantly, we further refined our analytical framework to better understand why only a subset of methylation changes translated into transcriptional differences (Results: Line 254-266). By integrating publicly available chromatin data, we found that the impact of DNA methylation on gene expression was strongly dependent on chromatin context. In particular, hypomethylation at promoters marked by active chromatin was more consistently associated with gene upregulation, whereas similar methylation changes in inactive or repressed regions showed limited transcriptional impact (**Supplementary Figure 3e-f**).

In addition, extending our analysis to distal regulatory elements revealed a broader set of enhancer-associated methylation changes linked to gene expression (Results: Line 296-334). Using enhancer annotations, we identified a larger set of genes showing concordant methylation and expression changes across systems (224 differentially methylated enhancers associated with 226 dysregulated genes; **Figure 5a**, **Supplementary Figure 5a**, **Supplementary Table 7**). As a representative example, we validated the expression of *ID1*, a gene with a well-established role in lung metastatic colonization. In our model, enhancer regions associated with *ID1* were hypomethylated in lung metastases, and this was accompanied by increased expression (**Figure 5d–e**).

Taken together, these analyses support a context-dependent model in which DNA methylation contributes to transcriptional regulation in combination with other regulatory layers. We have revised the manuscript accordingly to avoid implying a direct or universal correspondence between methylation changes and transcriptional outputs.

**4. The authors pool primary + LN metastasis samples and compare them against distant lung metastases, citing low sample numbers and similarity between primary and regional metastases. This seems to be a pragmatic choice but under-justified. Even if the PT–LN contrast yields fewer DMS than LU-involving contrasts, that does not establish that PT and LN are interchangeable; hundreds to thousands of loci can still carry biologically meaningful regional dissemination signals. Prior literatures support the view that regional metastases can exhibit non-trivial methylation remodeling relative to primaries (e.g., Reyngold et al., PMID: 25083786; Urrutia et al., PMID: 25628026; Barekati et al., PMID: 22695536; Mathe et al., PMID: 27671774). Pooling PT and LN could therefore mask stepwise progression patterns or confound interpretation of "lung-specific" changes.**

**The authors should perform sensitivity analyses showing that key DMS sets remain consistent when analyzing LU vs PT and LU vs LN separately (in addition to the pooled LU vs PT+LN contrast) for justification.**

**Response to reviewer:** We thank the reviewer for this comment and agree that primary tumors and regional lymph node metastases represent biologically distinct stages of disease progression. Indeed, we showed in the AURORA dataset that local metastases exhibit a DNA methylation landscape that differs from that of primary tumors (**Supplementary Figure 3a-b**). In our xenograft model, primary tumor and lymph node samples displayed highly similar profiles, potentially due to the short time from primary to lymph node metastases (**Figure 2a**, **Supplementary Figure 1**). Based on these similarities and the small sample size, we pooled these two sample types together. However, we acknowledge that this approach could potentially mask biologically meaningful differences between these compartments.

To address this concern, we performed additional sensitivity analyses comparing lung metastases independently against primary tumors (LU vs PT) and lymph node metastases (LU vs LN), in addition to the pooled comparison (LU vs PT+LN) (Lines 153-159). As shown in the new **Supplementary Figure 2a**, the majority of DMS detected in the pooled comparison were also recovered in both independent contrasts, indicating that the pooled analysis captures a shared lung-specific methylation signature.

Importantly, because each comparison also contained a subset of unique DMS, we assessed the directionality of these CpGs. Effect-size estimates for the pooled comparisons were highly concordant with those from LU vs PT and LU vs LN (Spearman's rho > 0.9 in both cases). Overall, the high concordance in both DMS overlap and effect-size directionality across comparisons supports the robustness of our findings and indicates that the pooled PT+LN reference captures a consistent lung-associated methylation program. Taken together, these results indicate that the pooled analysis captures a robust lung-associated DNA methylation pattern.

**5. In the introduction, the authors implied that the epigenetic landscape of TNBC metastasis is broadly "poorly characterized." Given the existence of prior methylation profiling studies comparing primary vs metastasis (including matched regional metastases mentioned in point 3) and large multi-omics analysis on AURORA (PMID: 36585450), the manuscript should more precisely state what is genuinely new here.**

**Response to reviewer:** We thank the reviewer for this comment and agree that the original wording in the Introduction may have overstated the extent to which the epigenetic landscape of TNBC metastasis is unexplored. In the revised manuscript, we have refined this statement to more accurately reflect the current state of the field and to clearly define the specific scope and contribution of our study (Introduction: Line 86-94). While previous studies have characterized DNA methylation differences between primary tumors and metastatic sites, our work addresses several aspects that have not been systematically explored:

- **Metastatic site specificity:** Prior studies have predominantly analyzed regional or mixed metastatic sites, whereas our study focuses specifically on lung metastasis, enabling the identification of DNA methylation patterns associated with this clinically relevant metastatic niche (Result sections 1 and 2: Line 112 and 144).
- **Regulatory context integration:** We extend beyond differential methylation analysis by incorporating chromatin state annotations and enhancer-associated regulatory elements, providing a context-dependent framework to interpret DNA methylation changes (Results: Line 254-266, 296-334).
- **Cross-system convergence**: By integrating a controlled xenograft model with independent patient-derived datasets, we identify DNA methylation alterations that are reproducible across systems, reducing potential confounding effects from tumor heterogeneity and microenvironmental admixture (Results: Line 216-240).

**Minor points**

**1. In Figure 3e, the manuscript text states 284 shared genes, while the venn diagram shows 285 in the overlap.**

**Response to reviewer:** We thank the reviewer for noting this discrepancy. The inconsistency between the manuscript text and the Venn diagram has been corrected in the revised figure and corresponding text (Results: Line 235-240; **Figure 3e**).

**2. For the Kaplan–Meier analyses, please specify how "high vs low" expression cutoffs were chosen, e.g. median expression?**

**Response to reviewer:** We thank the reviewer for this suggestion. We modified the manuscript to provide a detailed description of how expression cutoffs were defined in the Methods section (Methods: Line 656-662), and the selected cutoff values for each gene are reported in **Supplementary Table 8**.

Briefly, cutoffs were determined using a data-driven approach analogous to that implemented in KMplotter. All possible thresholds between the 25th and 75th percentiles of expression were evaluated, and the optimal cutoff was defined as the one yielding the lowest false discovery rate (Benjamini–Hochberg corrected) for lung metastasis-free survival.

**3. It will be appreciated if the enriched methylation loci in open sea regions can be functionally linked/interpreted to chromatin states/enhancers where possible.**

**Response to reviewer:** Thank you for this suggestion. In the revised manuscript, we have extended our analyses to functionally interpret methylation changes by integrating chromatin state annotations and enhancer information across all CpGs. Using publicly available ChIP-sequencing and chromatin accessibility data, we defined the regulatory context of these loci and incorporated GeneHancer annotations to identify distal regulatory elements (Methods: Line 591-622). Within this framework, we found that a substantial fraction of DMS maps to enhancer regions, and that enhancer-associated genes exhibit a higher frequency of expression changes compared to promoter-based analyses (Results: Line 296-334).

In total, we identified 224 differentially methylated enhancers linked to 226 dysregulated genes between lung metastases and primary tumors (**Figure 5a**, **Supplementary Table 7**). Hypermethylated enhancers were predominantly associated with gene downregulation, whereas hypomethylation showed a context-dependent effect, with a general trend toward increased expression (**Figure 5b-c**).

Importantly, when stratifying enhancers by chromatin state, hypomethylated enhancers in active contexts were primarily associated with gene upregulation, whereas those in repressive contexts showed limited or inverse transcriptional effects. These results highlight that the functional impact of DNA methylation changes is strongly dependent on chromatin context and support a context-dependent model of gene regulation.

**Point-by-point response to the reviewers' comments on the manuscript**

We thank the reviewers for their thoughtful and constructive comments, and for their positive assessment of the revised manuscript. We are pleased that the revisions were considered to have substantially improved the manuscript. In this final revision, we have carefully addressed the remaining minor points raised by the reviewers.

Below, we provide a point-by-point response to each comment.

**REVIEWERS' COMMENTS:**

**Reviewer #1 (Remarks to the Author):**

**I would like to thank the authors for their extensive efforts in revising the manuscript. The overall tone is now well-balanced and appropriately cautious. The re-analysis utilizing various public datasets not only strengthens the authors' original claims but also provides novel insights, such as through the integrated analysis with histone modifications. Expanding the scope of the analysis to enhancer regions has also added significant depth to the paper. Furthermore, the numerous newly provided supplementary tables in the revised manuscript will serve as valuable resources for future research on DNA methylation and broader epigenomics in TNBC lung metastasis.**

**While the authors have addressed most of my previous concerns sufficiently, I would like them to consider the following minor points before final publication.**

**(1) Supplementary Figure 4a:**

**I understand that this heatmap illustrates expression correlations used to narrow down candidate genes for qPCR validation. However, it is currently unclear from the figure itself, the figure legend, or the main text what the underlying data source is and what the color scale represents. Although the rebuttal letter clarifies that the AURORA dataset was used, adding these details to the figure legend would greatly enhance readability. If feasible, including a representative scatter plot with a fitted regression line for a key gene pair would further facilitate the reader's understanding.**

**Response to reviewer:** We thank the reviewer for this helpful suggestion. We have revised the legend of Supplementary Figure 4 to clarify that the expression correlation heatmap was generated using RNA-seq normalized count matrices from the AURORA US cohort. We now also specify that the color scale represents pairwise Pearson correlation coefficients between candidate gene expression levels.

In addition, we have included a representative scatter plot illustrating the correlation between *ZBTB17* and *AK1* expression across samples (**Supplementary Figure 4b**), as suggested by the reviewer. These additions improve the readability and interpretation of **Supplementary Figure 4a**.

**(2) Title:**

**Although the revised title is more appropriately cautious than the previous version, it still feels somewhat broad or vague, resembling the title of a review article. That being said, the**

**abstract clearly conveys the core findings and context of the study. Therefore, if the authors and the editor are satisfied with the current title, I have no further objections.**

**Response to reviewer:** We thank the reviewer for this comment and for acknowledging that the revised title was more appropriately cautious than the previous version. Following the reviewer's suggestion, we further reconsidered the title and have now modified it to "*DNA methylation reprogramming of gene regulatory elements is associated with lung metastasis in triple-negative breast cancer*". We believe that this title more precisely reflects the main findings of this study, highlighting the specific epigenetic mechanism investigated while maintaining an appropriately cautious interpretation of the data.

**Reviewer #2 (Remarks to the Author):**

**The revised manuscript is substantially improved. In particular, the revised title and overall framing are more appropriate. The additional analyses, including separate LU vs PT and LU vs LN sensitivity, qPCR validation, survival analysis, and additional supplementary tables, have greatly strengthened the manuscript. Overall, most of my previous concerns have been adequately addressed. However, I still like to recommend one minor text change before publication.**

**Although the authors state that causal interpretations between DNAm and metastasis have been removed, in the result section, line 364-367, the statement that the data "support a functional role for these genes beyond metastatic adaptation" remains, I feel, too strong for the provided evidence. I suggest revising this to a more cautious statement, such as "further nominates these genes as candidates for future functional validation."**

**Response to reviewer:** We thank the reviewer for the positive evaluation of the revised manuscript and for recognizing the improvements made during the revision process.

We agree that the highlighted sentence could still be interpreted as implying a functional role that is not directly demonstrated by the data. Accordingly, we have revised the text to adopt a more cautious interpretation, in line with the reviewer's suggestion. The sentence now states that the identification of these alterations in primary tumors and their association with lung dissemination risk "*further nominates these genes as candidates for future functional validation."*

We thank the reviewer again for this final suggestion and for the constructive feedback provided throughout the review process.