

Normalization Bias in Morpho-Transcriptomic Prediction

Swann Ruyter¹ 

SWANN.RUYTER@ICM-INSTITUTE.ORG

Reuben Dorent^{*1} 

REUBEN.DORENT@INRIA.FR

Daniel Racoceanu^{*1} 

DANIEL.RACOCEANU@SORBONNE-UNIVERSITE.FR

¹ Sorbonne Université, CNRS, Inserm, AP-HP, Inria, Paris Brain Institute - ICM, Paris, France

Editors: Under Review for MIDL 2026

Abstract

Spatial transcriptomics (ST) prediction from H&E histology has attracted growing attention as a scalable approach to infer gene expression from tissue morphology. However, existing methods are typically benchmarked after target normalization, whose impact remains insufficiently characterized. This is critical because raw ST counts partly reflect local cellularity; consequently, normalization determines how much of this morphology-linked variation is retained in the prediction target. Using STEM as a controlled diffusion-based backbone, we compare its original log-normalization with a Pearson-residual-based normalization on a HER2-positive breast cancer ST cohort. We show that log-normalization preserves substantially stronger nuclei- and depth-related signals, and that the best-predicted genes under this regime are also the most strongly associated with nuclei-related features in the observed data. These findings suggest that part of the apparent performance gain in log-transformed space arises from the preservation of an easier-to-predict cellularity-related signal, with important implications for method comparison and for the interpretation of correlation-based benchmarks.

Keywords: Spatial transcriptomics, diffusion models, normalization, H&E.

1. Introduction

Spatial transcriptomics (ST) enables *in situ* measurement of gene expression while preserving tissue architecture, thereby providing a direct link between routine H&E histology and spatially resolved molecular states (Rao et al., 2021). Given the widespread availability of H&E slides and the high cost and limited scalability of ST, recent work has focused on morpho-transcriptomic approaches that infer gene expression directly from tissue morphology (e.g., Pang et al., 2021; He et al., 2020; Zeng et al., 2022; Zhu et al., 2025).

Current methods are mostly evaluated with correlation- or error-based metrics, such as Pearson correlation coefficient (PCC) and root mean squared error (RMSE), computed on normalized transcriptomic targets. However, the effect of target normalization itself is rarely isolated (Wang et al., 2025). This issue is particularly critical in ST, where sequencing depth directly affects the stability of expression estimates, while also depending in part on the amount of biological material captured within each spot, and thus on local tissue cellularity.

As a result, target normalization can either preserve or suppress morphology-associated signals linked to cellularity and capture efficiency, thereby introducing a potential bias in the evaluation of predictive models.

* Contributed equally

In this work, we use STEM (Zhu et al., 2025) to compare the effect of its original log-transformed target with a Pearson-residual-based normalization (Lause et al., 2021). We show that the log-transformed target retains a stronger cellularity- and sequencing-depth-related axis, and that its best-predicted genes are also the most strongly associated with nuclei-related features. These findings suggest that part of the apparent correlation gain in log-transformed space arise from the preservation of easier-to-predict cellularity-related variation, potentially confounding method comparison and biasing the assessment of finer morpho-transcriptomic signals.

2. Method

We use STEM (Zhu et al., 2025) as a controlled morpho-transcriptomic backbone, in which a conditional denoising diffusion model generates spot-level transcriptomic predictions from corresponding histological patch embeddings. Local cellularity is estimated from H&E images via nuclei segmentation using TIAToolbox (Pocock et al., 2022). Nuclei counts are aggregated at the spot level and used as a proxy for captured biological material.

To quantify the effect of target normalization, we compare STEM’s original log-transformed normalization (Eq. (1)) with a Pearson-residual-based normalization (Eq. (2)) under identical *one-slide-out* predictions and matched observed targets. We then derived two complementary analysis families: (i) coupling between observed targets and nuclei count, and (ii) nuclei association among the top-ranked predicted genes.

3. Experiments

We evaluate our approach on the public ST cohort HER2-positive breast cancer (Andersson et al., 2021), harmonized in HEST-1K (Jaume et al., 2024) (35 slides after QC). All experiments follow a *one-slide-out* protocol, ensuring strict separation between training and test slides. A *one-slide-out* protocol ensures strict separation between training and test data. A panel of 300 genes is selected by combining per-slide variability with global ranking by expression and dispersion. All experimental settings are fixed, and only the target normalization is varied.

Figure 1 panels A–C, reveal that the effect is already present in the observed data. Nuclear density correlates with sequencing depth across slides, with a median slide-level Spearman correlation of 0.505, indicating that nuclei count captures a meaningful proxy for the amount of material captured per spot. This dependence is substantially better preserved in the log-transformed target than in the Pearson-residual target, both in terms of median gene-wise nuclei association (panel B; median $|\rho(\text{true log expr}, n_{\text{nuclei}})| = 0.337$ vs. 0.085) and the proportion of nuclei-associated genes (panel C; 0.68 vs. 0.03 for $|\rho| > 0.3$). These results show that normalization affects both scaling and the amount of retained cellularity-related signal (Supplementary Table 1).

Figure 1 panels D–F shows that the apparent performance advantage of log-space prediction is not uniform across genes. Although log-transformed targets yield higher predictive correlations for the top-ranked genes (panel D), these genes are also more strongly associated with nuclear density in the observed data than the corresponding top-ranked genes in Pearson space (panel E). This effect is stable across all the top- k thresholds con-

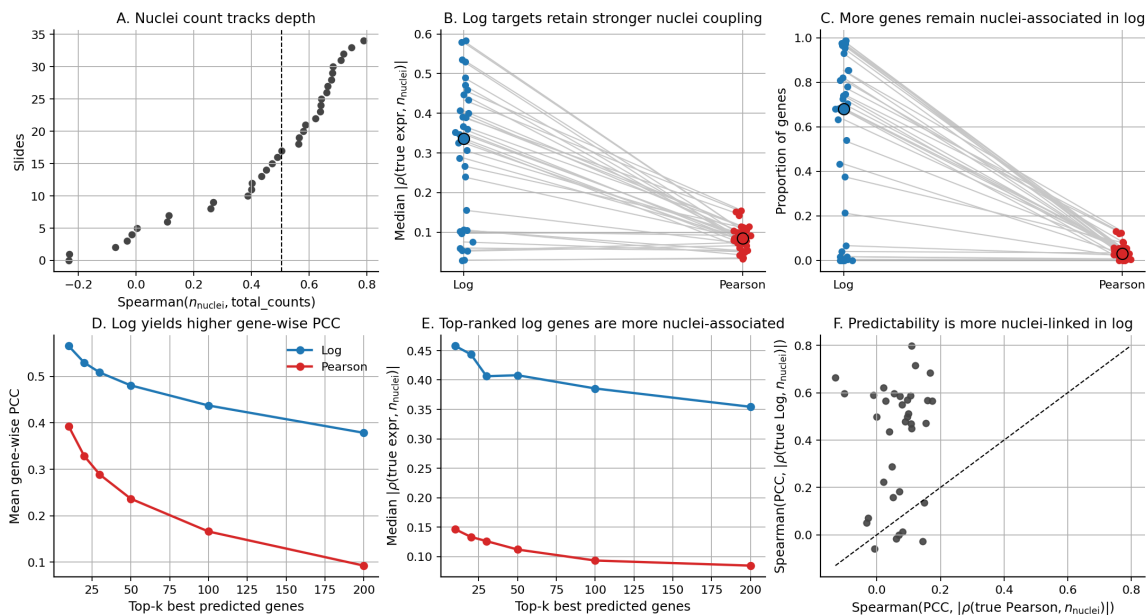


Figure 1: **Target normalization alters observed signal and apparent predictive gain.** **A–C:** Log targets retain stronger cellularity-related signal than Pearson targets. **D–F:** Higher correlations in log space concentrate on genes more associated with nuclei count.

sidered, with a median paired excess in nuclei association ranging from 0.257 (top-10) to 0.273 (top-200). All paired Wilcoxon tests are significant ($p \leq 2.9 \times 10^{-7}$) and permutation tests were consistent ($p \leq 2.0 \times 10^{-4}$). The same trend holds at slide level, where gene predictability is more tightly coupled to nuclei dependence in log-transformed space (panel F). Together, these results indicate that part of the correlation gain in log space reflects preserved cellularity-related variation rather than finer morpho-transcriptomic recovery (Supplementary Table 2).

4. Conclusion

In summary, transcriptomic target normalization changes not only what is predicted, but also how predictive performance should be interpreted. In our setting, log-normalized targets preserve a stronger cellularity-related signal than Pearson-residual-normalized targets, and this signal is concentrated among the genes that appear best predicted. As a result, correlation-based benchmarks may partly reward coarse, easier-to-predict cellularity-related variation rather than finer morpho-transcriptomic recovery, depending on the biological question of interest. Although this study is limited to a single backbone and cohort, it demonstrates that normalization should be treated as a central experimental factor rather than a secondary preprocessing choice.

Acknowledgments

The research leading to these results has received funding from Agence Nationale de la Recherche as part of the “France 2030” program (reference ANR-23-IACL-0008, PRAIRIE-PSAI) and as part of the “Investissements d’avenir” program (reference ANR-19-P3IA-0001, PRAIRIE 3IA Institute; and reference ANR-10-IAIHU-0006). The ARAMIS Lab is affiliated with DIM C-BRAINS, funded by the Conseil Régional d’Île-de-France. This work was performed using HPC resources from GENCI–IDRIS (Grant 2025-AD011016416). R.D. received a Marie Skłodowska-Curie grant No 101154248 (project: SafeREG).

References

- Alma Andersson, Ludvig Larsson, Linnea Stenbeck, Fredrik Salmén, Anna Ehinger, Sunny Z Wu, Ghamdan Al-Eryani, Daniel Roden, Alex Swarbrick, Åke Borg, Jonas Frisé, Camilla Engblom, and Joakim Lundeberg. Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nature communications*, 12(1):6012, October 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-26271-2.
- Bryan He, Ludvig Bergenstråhle, Linnea Stenbeck, Abubakar Abid, Alma Andersson, Åke Borg, Jonas Maaskola, Joakim Lundeberg, and James Zou. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature Biomedical Engineering*, 4(8):827–834, August 2020. ISSN 2157-846X. doi: 10.1038/s41551-020-0578-x.
- Guillaume Jaume, Paul Doucet, Andrew H. Song, Ming Y. Lu, Cristina Almagro-Pérez, Sophia J. Wagner, Anurag J. Vaidya, Richard J. Chen, Drew F. K. Williamson, Ahrong Kim, and Faisal Mahmood. HEST-1k: A Dataset for Spatial Transcriptomics and Histology Image Analysis, November 2024. arXiv:2406.16192 [cs].
- Jan Lause, Philipp Berens, and Dmitry Kobak. Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data. *Genome biology*, 22(1):258, September 2021. ISSN 1474-760X. doi: 10.1186/s13059-021-02451-7.
- Minxing Pang, Kenong Su, and Mingyao Li. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors, November 2021. Pages: 2021.11.28.470212 Section: New Results.
- Johnathan Pocock, Simon Graham, Quoc Dang Vu, Mostafa Jahanifar, Srijay Deshpande, Giorgos Hadjigeorgiou, Adam Shephard, Raja Muhammad Saad Bashir, Mohsin Bilal, Wenqi Lu, David Epstein, Fayyaz Minhas, Nasir M. Rajpoot, and Shan E. Ahmed Raza. TIAToolbox as an end-to-end library for advanced tissue image analytics. *Communications Medicine*, 2(1):120, September 2022. ISSN 2730-664X. doi: 10.1038/s43856-022-00186-5.
- Anjali Rao, Dalia Barkley, Gustavo S. França, and Itai Yanai. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871):211–220, August 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03634-9.

Chuhan Wang, Adam S. Chan, Xiaohang Fu, Shila Ghazanfar, Jinman Kim, Ellis Patrick, and Jean Y. H. Yang. Benchmarking the translational potential of spatial gene expression prediction from histology. *Nature Communications*, 16(1):1544, February 2025. ISSN 2041-1723. doi: 10.1038/s41467-025-56618-y.

Yuansong Zeng, Zhuoyi Wei, Weijiang Yu, Rui Yin, Yuchen Yuan, Bingling Li, Zhonghui Tang, Yutong Lu, and Yuedong Yang. Spatial transcriptomics prediction from histology jointly through Transformer and graph neural networks. *Briefings in Bioinformatics*, 23(5):bbac297, September 2022. ISSN 1477-4054. doi: 10.1093/bib/bbac297.

Sichen Zhu, Yuchen Zhu, Molei Tao, and Peng Qiu. Diffusion Generative Modeling for Spatially Resolved Gene Expression Inference from Histology Images, January 2025. arXiv:2501.15598 [cs].

Supplementary Material

Logarithmic normalization

For the logarithmic target used in the original STEM setting, raw counts are transformed element-wise without library-size normalization:

$$Y_{ij}^{\text{log}} = \log_2(X_{ij} + 1). \quad (1)$$

This representation compresses the dynamic range of count values but does not explicitly correct for sequencing-depth effects.

Pearson Residual Normalization

Let $X \in \mathbb{R}_{\geq 0}^{N \times G}$ be a raw count matrix, where N and G are respectively the number of spots and genes. We compute Pearson residuals under a negative binomial (NB) model with fixed dispersion θ . First, we define the library size $n_i = \sum_j X_{ij}$, the gene proportion $p_j = \frac{\sum_i X_{ij}}{\sum_{i,j} X_{ij}}$, and the expected mean $\mu_{ij} = n_i p_j$. Then, Pearson residuals are computed as

$$R_{ij} = \frac{X_{ij} - \mu_{ij}}{\sqrt{\mu_{ij} + \mu_{ij}^2/\theta}}, \quad (2)$$

where θ denotes a fixed dispersion parameter and a numerical floor of 10^{-8} is applied to the denominator for stability. This formulation leads to approximately variance-stabilized residuals across genes and expression levels.

Supplementary Statistical Tables

Table 1: **Observed targets retain different amounts of nuclei- and depth-related signal.** Slide-level summary statistics computed from the observed transcriptomic targets. Strong nuclei association is defined as $|\rho(\text{expression}, n_{\text{nuclei}})| > 0.3$.

Metric	Log	Pearson
Median $ \rho(\text{expr}, n_{\text{nuclei}}) $	0.337	0.085
Median proportion with $ \rho(\text{expr}, n_{\text{nuclei}}) > 0.3$	0.680	0.030
Depth coupling		
Median Spearman(n_{nuclei} , total_counts)	0.505	

Table 2: **Top- k best-predicted genes are more nuclei-associated in log space.** For each k , we report the slide-level median paired difference $\Delta_k = |\rho(\text{true log}, n_{\text{nuclei}})| - |\rho(\text{true Pearson}, n_{\text{nuclei}})|$ computed on the top- k best-predicted genes, together with paired Wilcoxon and permutation tests.

Top- k	Median Δ_k	Wilcoxon p	Permutation p
10	0.257	2.9×10^{-7}	2.0×10^{-4}
20	0.266	7.1×10^{-7}	1.0×10^{-4}
50	0.269	6.9×10^{-8}	1.0×10^{-4}
200	0.273	6.0×10^{-9}	1.0×10^{-4}