Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An Analysis-First Framework for Region-Level Heterogeneity Characterization in Breast Cancer Whole-Slide Images

Mohammad Mehedi Hassan

m_hassan@kcg.edu

The Kyoto College of Graduate Studies for Informatics

Research Article

Keywords: Whole-slide images, Tissue heterogeneity, Computational pathology, Breast cancer

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11763682/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

An Analysis-First Framework for Region-Level Heterogeneity Characterization in Breast Cancer Whole-Slide Images

Mohammad Mehedi Hassan[1*]

[1*]Information Systems Department, The Kyoto College of Graduate Studies for Informatics, 7 Tanakamonzen-cho, Sakyo-ku, Kyoto, 606-8225, Kyoto, Japan.

Corresponding author(s). E-mail(s): m_hassan@kcg.ac.jp;

Abstract

Histopathological whole-slide images (WSIs) contain rich spatial information reflecting tissue organization and heterogeneity, yet many computational approaches reduce these images to patch-level predictions or coarse aggregates, limiting tissue-scale interpretability. This study presents an analysis-first, region-aware framework for characterizing tissue heterogeneity in breast cancer WSIs. Diagnostic slides from The Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA) cohort were partitioned into spatially localized tissue regions, and region-level representations were extracted using a frozen ResNet-50 encoder. Without relying on supervised learning, a region-aware heterogeneity characterization framework was developed using four representation-space measures capturing regional feature dispersion, inter-region dissimilarity, embedding-norm concentration, and clustering-derived rarity. Evaluation on 173 WSIs revealed non-degenerate metric distributions and structured correlations, indicating that tissue organization cannot be fully summarized by a single representation-space measure. Quantile-based thresholding identified recurring organizational phenotypes termed concentrated low dispersion, diffuse heterogeneous, and rare pattern rich. Representative cases illustrated visually distinct appearances associated with the metric-based phenotype definitions. Robustness analyses demonstrated stable phenotype assignments under moderate threshold perturbations. In addition, region-size ablation experiments showed that dispersion and embedding-norm concentration measures remained highly consistent across spatial scales, whereas clustering-based rarity measures were more sensitive to region granularity. A pilot encoder-sensitivity analysis further showed moderate-to-strong preservation of metric rankings between ResNet-50 and pathology-specific

UNI representations. The proposed framework explicitly separates heterogeneity characterization from predictive modeling and provides an interpretable approach for studying tissue organization in WSIs. These results provide a basis for future studies on tissue heterogeneity using computational pathology.

Keywords: Whole-slide images, Tissue heterogeneity, Computational pathology, Breast cancer

1 Introduction

Whole-slide histopathology images allow one to study in detail tissue structure and serve as the basis for cancer diagnosis and research [1, 2]. With the growing amount of WSIs that can be digitized nowadays, computational pathology stands as one of the most important domains where high-resolution histological data can be analyzed for gaining relevant biological insights. The overwhelming majority of current approaches focus on patch-level processing and, consequently, involve dividing WSIs into small tiles to apply convolutional neural networks separately to each of them [3, 4]. Such a strategy allows one to achieve good prediction results, but cannot preserve spatial relations between image regions and may reduce highly structured data to just aggregated scores. Heterogeneity [5–7] of tissues is one of the main characteristics of any solid tumors, breast cancers included, since it implies a presence of different morphologies and architectures within them. However, pathologists regularly use spatial reasoning in their work with diseases, while most computational frameworks regard WSIs simply as sets of patches with exchangeable observations [8]. There have been many attempts using multiple instance learning (MIL) models to tackle problems arising from the weakly supervised nature of WSI. Particularly, attention-based variants have gained considerable interest due to their capability of selecting informative patches based on importance scores [9, 10]. Nevertheless, most existing MIL models are still working at the patch level without any additional modeling on the coherent regions that might carry valuable information for tumor heterogeneity. Consequently, heterogeneity becomes a latent factor embedded in each specific problem instead of an explicitly defined measure. More recently, researchers started to focus on representing WSIs at a higher level, i.e., the level of region, where patches in each region are adjacent to each other as they usually reflect the same histology pattern [11]. This line of research, however, mainly serves a predictive purpose. In this work, we adopt an analysis-first perspective to characterize tissue-level heterogeneity in breast cancer WSIs from the TCGA-BRCA [12] cohort. Using spatially localized regions and frozen deep representations, we define a set of unsupervised metrics capturing regional feature dispersion, inter-region dissimilarity, embedding-norm concentration, and minimum cluster prevalence. By decoupling heterogeneity characterization from supervised prediction, we show that tissue organization cannot be fully summarized by a single representation-space measure and can be characterized independently of specific clinical endpoints. The contributions of this study are summarized as follows:

- We introduce a region-aware analysis framework that characterizes tissue organization in whole-slide images independently of supervised prediction objectives.
- We operationalize tissue organization through complementary representation-space measures capturing dispersion, embedding-norm concentration, and clustering-derived prevalence patterns.
- We define organizational phenotypes using transparent quantile-based rules and evaluate their robustness under threshold perturbations.
- We investigate sensitivity to spatial granularity through a region-size ablation study, demonstrating that continuous dispersion measures remain substantially more stable than phenotype assignments and clustering-derived rarity measures.

2 Related Works

Computational analysis of histopathological whole-slide images (WSIs) has largely been driven by patch-based deep learning approaches, where slides are decomposed into small tiles processed independently by convolutional neural networks [3, 4]. While effective for various prediction tasks, such approaches inherently discard spatial relationships between patches, limiting their ability to capture tissue-level organization. To address weak supervision in WSI analysis, multiple instance learning (MIL) has become a dominant paradigm. Attention-based MIL methods further enable identification of informative regions by learning patch-level importance weights [9, 10]. However, most MIL approaches function at the patch-level directly without explicit consideration of mid-scale tissue organization, thus leaving heterogeneity implicitly encoded in a task-specific manner. Recently, spatial dependencies were accounted for via region-aware or graph representations where dependencies between adjacent patches were explicitly considered [11]. While such formulations facilitate better contextual reasoning, they are predominantly trained to predict labels from annotated data. Several studies have explored morphological variation and tissue organization in WSIs using clustering, representation learning, and spatial analysis approaches to characterize intra-slide diversity beyond predictive modeling [13, 14]. Such approaches often rely on heuristic definitions of heterogeneity or embed heterogeneity within predictive models, making independent interpretation of tissue organization challenging. In contrast to prediction-driven pipelines, a smaller body of work has explored unsupervised or analysis-oriented use of deep representations to study morphological variation in histopathology.

Our work builds on this line by adopting a region-aware, analysis-first framework that explicitly quantifies tissue heterogeneity using unsupervised metrics, without coupling organization characterization to a specific supervised task.

3 Methodology

3.1 Dataset and Study Design

This study is conducted on breast cancer whole-slide images (WSIs) from The Cancer Genome Atlas (TCGA-BRCA). Only diagnostic hematoxylin and eosin (H&E) stained SVS files that are physically available and verifiable were included. To avoid repeated

sampling of tissue architecture from the same patient, one diagnostic slide was retained per patient. When both DX1 and DX2 diagnostic slides were available, the DX1 slide was prioritized. After filtering, the final analysis cohort consisted of 173 WSIs from 173 unique patients.

Rather than structuring the workflow around a prediction objective, we adopt an analysis-first study design. All methodological components are developed to characterize tissue organization and heterogeneity at the slide level without requiring dense annotations or fully supervised labels.

3.2 Tissue Detection and Tiling

Each WSI was processed to remove background regions and isolate tissue areas. A low-resolution representation of the slide was generated and converted to grayscale, followed by Otsu [15] thresholding to produce a binary tissue mask. Tiles containing less than 50% tissue coverage were discarded to prevent downstream representations from being dominated by background artifacts. Tissue regions were then decomposed into non-overlapping 256×256 pixel tiles at $20\times$ magnification using a stride of 256 pixels. Tile coordinates were preserved in the slide reference frame to maintain spatial correspondence throughout subsequent processing stages. Additional tile extraction and region construction parameters are summarized in Table 1.

3.3 Tile-Level Feature Extraction

Tile-level visual representations were extracted using a convolutional neural network with frozen parameters. Specifically, a ResNet-50 [16] backbone pretrained on ImageNet [17] was employed, and the network's final pooling layer output was used as the tile embedding. Keeping the encoder fixed separated representation extraction from heterogeneity assessment and prevented the feature space from being adapted toward a specific supervised endpoint. Mathematically, for each tile t_j, the feature vector of dimension D

$$\mathbf{f}_j = \phi(t_j) \in \mathbb{R}^D, \tag{1}$$

is obtained through a fixed feature extractor $\phi(\cdot)$. Features were extracted from the final global average pooling layer, yielding a 2048-dimensional embedding for each tile. Each tile embedding is associated with spatial coordinates (x_j, y_j) and a corresponding slide identifier, enabling spatial grouping and region-aware aggregation.

3.4 Region Construction

To account for tissue organization on scales larger than individual tiles, spatially cohesive regions were created by grouping tiles within a fixed 1024×1024 pixel grid at level-0 resolution, corresponding to a maximum grid capacity of approximately 4×4 neighboring tiles. Tile extraction and region construction statistics for the final analysis cohort are summarized in Table 1. Tiles located within the same grid cell were grouped to form spatially localized intermediate-scale regions. This grid-based construction preserves local proximity but does not assume that each region corresponds to a distinct histological unit. Region representation was created by averaging together

Table 1 Tile extraction and region construction parameters for the 173-slide analysis cohort.

Parameter	Value
Magnification	20×
Tile size	256 × 256 pixels
Tile stride	256 pixels
Tile overlap	None
Region size	1024 × 1024 pixels
Mean tiles per WSI	2474
Median tiles per WSI	2000
Mean regions per WSI	280
Median regions per WSI	219

embedding vectors of all tiles contained in the region $R_k = \{t_{k1}, \ldots, t_{km}\}$:

$$\mathbf{r}_k = \frac{1}{|R_k|} \sum_{t_j \in R_k} \mathbf{f}_j. \tag{2}$$

This transformation reduces thousands of tile representations to a more digestible number of interpretable regions. Across the 173-slide analysis cohort, each WSI contained an average of 2474 tiles and 280 regions, with median values of 2000 tiles and 219 regions, respectively.

3.5 Region-Level Heterogeneity Metrics

Region-level tissue heterogeneity is characterized using unsupervised metrics calculated from the set of all regions embedded from each slide individually. Suppose $\mathcal{R} = \{\mathbf{r}_k\}_{k=1}^{K}$ is the set of region embeddings for a slide. The proposed metrics characterize the distribution of embeddings derived from spatially localized regions. They do not explicitly incorporate adjacency, distance, or relative arrangement among regions and therefore quantify region-level representation heterogeneity rather than spatial autocorrelation or tissue topology. Region embeddings were clustered independently for each slide using KMeans with $k = 5$, 10 initializations, and a random seed of 0. The same clustering configuration was used for all slides and region-size experiments. Raw region embeddings were used without additional standardization to preserve consistency with the original feature-space metric definitions. A fixed value of $k = 5$ was selected as a pragmatic clustering resolution that provided multiple region groups while avoiding excessive fragmentation. The same clustering configuration was applied to all slides and experiments to maintain consistency. Sensitivity to clustering resolution remains a limitation of the proposed clustering-derived prevalence measure. We define four different heterogeneity metrics that measure various properties of the

distribution of region embeddings:

$$H_{\text{var}} = \frac{1}{D} \sum_{d=1}^{D} \text{Var}\left(\{\mathbf{r}_k^{(d)}\}_{k=1}^{K} \right), \tag{3}$$

which captures overall variability among regional representations within a slide. Variance was calculated using the population convention with divisor K (`ddof`=0). Inter-region dissimilarity is measured as the mean pairwise cosine distance,

$$H_{\cos} = \frac{2}{K(K-1)} \sum_{k<l} \left(1 - \frac{\mathbf{r}_k^{\top} \mathbf{r}_l}{\|\mathbf{r}_k\|_2 \|\mathbf{r}_l\|_2} \right), \tag{4}$$

quantifying global architectural diversity in the embedding space. The maximum embedding-norm proportion is defined as the contribution of the largest-norm region embedding relative to the sum of the embedding norms across all regions:

$$H_{\text{norm}} = \frac{\max_k \|\mathbf{r}_k\|_2}{\sum_{k=1}^{K} \|\mathbf{r}_k\|_2}. \tag{5}$$

This metric provides a representation-space measure of feature-magnitude concentration. Larger values indicate that a single region contributes a comparatively large embedding magnitude, whereas smaller values indicate that embedding magnitude is distributed across more regions. Because the measure is influenced by the number of regions and embedding scale, it should not be interpreted as the spatial prevalence or biological dominance of a tissue pattern. The minimum cluster proportion is defined as

$$H_{\text{min}} = \frac{\min_c |\mathcal{C}_c|}{K} \tag{6}$$

where $\mathcal{C}_c$ denotes the set of regions assigned to cluster c. Lower values indicate that the least populated clustering-derived representation pattern occupies a smaller proportion of the slide regions. This measure characterizes feature-space cluster prevalence and does not independently establish the presence of a biologically distinct structural component. These metrics are calculated independently for each slide and do not depend on any labeling information from outside the process.

3.6 Heterogeneity Phenotype Definition

To characterize recurring organizational patterns, heterogeneity phenotypes were defined using quantile-based thresholds on the computed metrics. In particular, the 20th and 80th percentiles were chosen as the criteria for the presence of extreme behavior along each heterogeneity measure. Slides that demonstrated weak dispersion and a comparatively concentrated embedding-norm distribution were considered *Concentrated low dispersion*, while slides demonstrating strong dispersion and dissimilarity between regions were called *diffuse heterogeneous*. The third type, defined by the presence of small region subsets, was called *rare-pattern rich*. Slides that did not meet

extreme criteria along these dimensions were assigned to an intermediate category. Let $Q_{20}(H)$ and $Q_{80}(H)$ denote the 20th and 80th percentile thresholds of metric H in the baseline cohort. Phenotypes were assigned using the following threshold-based rules:

$$\text{Concentrated low dispersion:} \quad H_{\text{var}} \leq Q_{20}(H_{\text{var}}) \wedge H_{\text{norm}} \geq Q_{80}(H_{\text{norm}}), \tag{7}$$

$$\text{Diffuse heterogeneous:} \quad H_{\text{var}} \geq Q_{80}(H_{\text{var}}) \wedge H_{\cos} \geq Q_{80}(H_{\cos}), \tag{8}$$

$$\text{Rare pattern rich:} \quad H_{\min} \leq Q_{20}(H_{\min}). \tag{9}$$

Slides not satisfying these conditions were assigned to the intermediate category. If a slide satisfied multiple phenotype criteria, the rules were applied sequentially in the order concentrated low dispersion, diffuse heterogeneous, and rare pattern rich.

3.7 Representative Case Selection

For qualitative illustration, three representative slides were selected for each phenotype by ranking eligible slides according to the metrics defining that phenotype. Concentrated low dispersion cases were ranked using low feature variance together with a high maximum embedding-norm proportion. Diffuse-heterogeneous cases were ranked using high feature variance and high mean pairwise cosine distance. Rare-pattern-rich cases were ranked using a low minimum cluster proportion. The selected slides were used only for qualitative visualization and did not influence the metric definitions or phenotype thresholds.

3.8 Statistical Analysis

Descriptive statistics were used to characterize the distributions of the proposed metrics. Spearman rank correlations were used to evaluate associations among the metrics and their relationships with the numbers of retained tiles and constructed regions. Threshold robustness was assessed using exact phenotype agreement and Cohen's κ. No supervised predictive modeling was performed.

4 Experimental Results

4.1 Distribution of Heterogeneity Metrics

We started our investigations by studying the empirical behavior of the suggested metrics for measuring tissue heterogeneity at the region level within the TCGA-BRCA collection. The empirical distributions of region feature variance, mean pairwise cosine distance, maximum embedding-norm proportion, and minimum cluster proportion are illustrated in Figure 1. All four metrics have non-trivial distributions with considerable dynamic ranges, which means that there is significant variation in tissue structure between slides. Indeed, we can see that the empirical distribution of region feature variance and mean pairwise cosine distance has long right tails, implying that some of the slides have highly varied region features. The maximum embedding-norm proportion occurred on a smaller numerical scale but exhibited measurable variation across slides. Across metrics, smooth transitions and identifiable extreme cases are observed

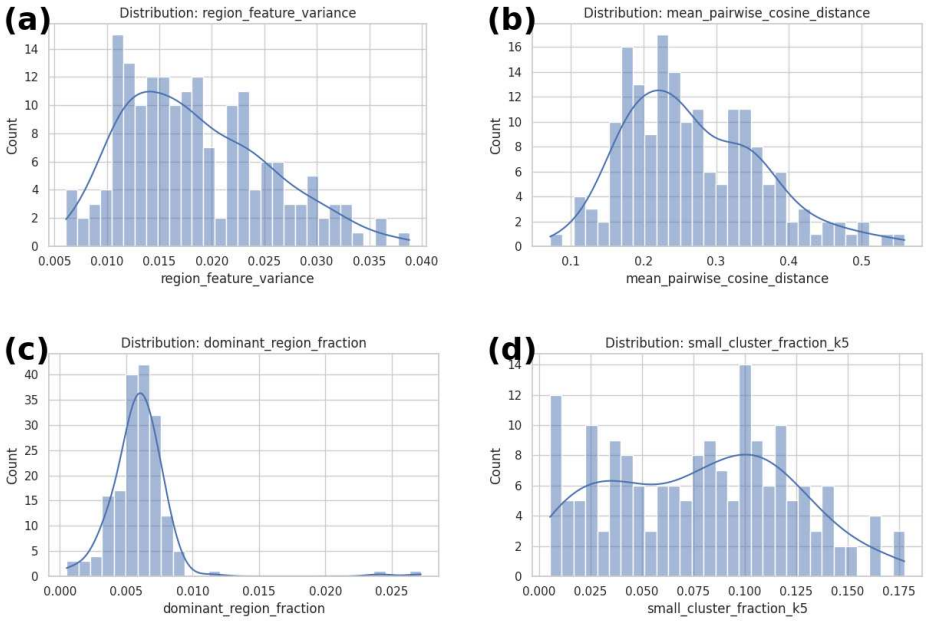


Fig. 1 Heterogeneity distribution by region-level measures across 173 WSIs: (a) variance of regional features, (b) mean cosine distance between regions, (c) maximum embedding-norm proportion, and (d) minimum cluster proportion.

Table 2 Summary statistics of region-level heterogeneity metrics across the patient-independent TCGA-BRCA cohort of 173 WSIs.

Metric	Mean	Std.	Min	Max
Region feature variance	0.0183	0.0068	0.0061	0.0388
Mean pairwise cosine distance	0.2666	0.0928	0.0713	0.5592
Maximum embedding-norm proportion	0.0059	0.0023	0.0005	0.0272
Minimum cluster proportion	0.0781	0.0434	0.0054	0.1744

across slides, supporting the interpretation of heterogeneity as a continuous organizational property rather than a discrete categorization. Summary statistics for all heterogeneity metrics are reported in Table 2.

4.2 Correlation Structure of Heterogeneity Metrics

Pairwise Spearman correlations were used to evaluate redundancy and differences among the proposed heterogeneity metrics. The resulting correlation matrix is shown in Figure 2.

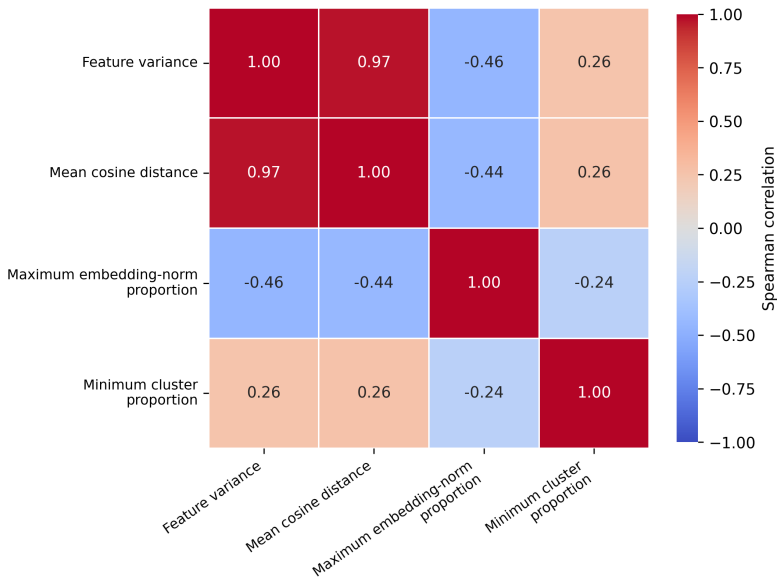


Fig. 2 Spearman correlation matrix among region-level heterogeneity metrics. A strong positive correlation was observed between the two dispersion measures, while weaker associations were observed between the remaining representation-space measures.

Region feature variance and mean pairwise cosine distance were strongly correlated ($\rho = 0.970$), indicating substantial redundancy between the two dispersion measures. The maximum embedding-norm proportion showed moderate negative correlations with feature variance ($\rho = -0.460$) and mean pairwise cosine distance ($\rho = -0.439$). The minimum cluster proportion exhibited weaker correlations with feature variance ($\rho = 0.256$), cosine distance ($\rho = 0.259$), and maximum embedding-norm proportion ($\rho = -0.238$), suggesting that it captures a less redundant aspect of the region-embedding distribution.

4.3 Associations with Slide Sampling Properties

To assess whether the proposed heterogeneity measures were influenced by slide sampling characteristics, Spearman correlations were calculated between each metric and the numbers of retained tiles and constructed regions. The results are summarized in Table 3. The two dispersion measures, region feature variance and mean pairwise cosine distance, showed no meaningful association with tile count ($\rho = -0.0931$, $p = 0.2230$ and $\rho = -0.0726$, $p = 0.3422$, respectively). However, both exhibited moderate positive correlations with region count ($\rho = 0.4955$ and $\rho = 0.4968$, respectively). The minimum cluster proportion also showed no meaningful association with tile count ($\rho = -0.0931$, $p = 0.2231$), but demonstrated a weaker positive correlation with region count ($\rho = 0.2377$, $p = 0.0016$). In contrast, the maximum embedding-norm proportion exhibited a weak negative association with tile count ($\rho = -0.2041$, $p = 0.0071$) and a very strong negative correlation with region count

Table 3 Associations between heterogeneity metrics and slide sampling properties in the 173-slide cohort. Values are Spearman correlation coefficients with corresponding two-sided p-values.

Metric	Number of tiles	Number of regions
Region feature variance	$\rho = -0.0931$, $p = 0.2230$	$\rho = 0.4955$, $p = 4.16 \times 10^{-12}$
Mean pairwise cosine distance	$\rho = -0.0726$, $p = 0.3422$	$\rho = 0.4968$, $p = 3.60 \times 10^{-12}$
Maximum embedding-norm proportion	$\rho = -0.2041$, $p = 0.0071$	$\rho = -0.9683$, $p = 6.35 \times 10^{-105}$
Minimum cluster proportion	$\rho = -0.0931$, $p = 0.2231$	$\rho = 0.2377$, $p = 0.0016$

Table 4 Robustness of phenotype assignments under alternative quantile thresholds in the 173-slide cohort.

Threshold comparison	Agreement (%)	Cohen's κ
15/85 vs. 20/80	87.9	0.777
25/75 vs. 20/80	89.0	0.826

($\rho = -0.9683$, $p = 6.35 \times 10^{-105}$). This finding indicates that the measure is substantially influenced by the number of region-level observations and should therefore be interpreted as a region-count-dependent feature-space concentration statistic rather than an independent estimate of biological pattern dominance.

4.4 Heterogeneity Phenotypes

Using quantile-based thresholds (20th and 80th percentiles) computed from the cohort-wide metric distributions, we defined three recurring organizational phenotypes: concentrated low dispersion, diffuse-heterogeneous, and rare-pattern-rich. Slides not meeting extreme criteria were assigned to an intermediate category. This procedure yields reproducible phenotype groupings under moderate threshold perturbations without reliance on clinical labels or supervision. Under the baseline 20th/80th percentile definition, 18 slides were classified as concentrated low dispersion, 30 as diffuse-heterogeneous, 21 as rare-pattern-rich, and 104 as intermediate. To assess robustness, phenotype assignments were recomputed using alternative quantile thresholds of 15th/85th and 25th/75th percentiles. The results are reported in Table 4.For the 15th/85th setting, all lower-tail conditions were evaluated at the 15th percentile and all upper-tail conditions at the 85th percentile. For the 25th/75th setting, the corresponding lower- and upper-tail thresholds were changed to the 25th and 75th percentiles, respectively. Compared with the baseline 20th/80th percentile definition, these settings produced agreement rates of 87.9% and 89.0%, with Cohen's κ values of 0.777 and 0.826, respectively. These results indicate substantial stability under moderate threshold perturbations.

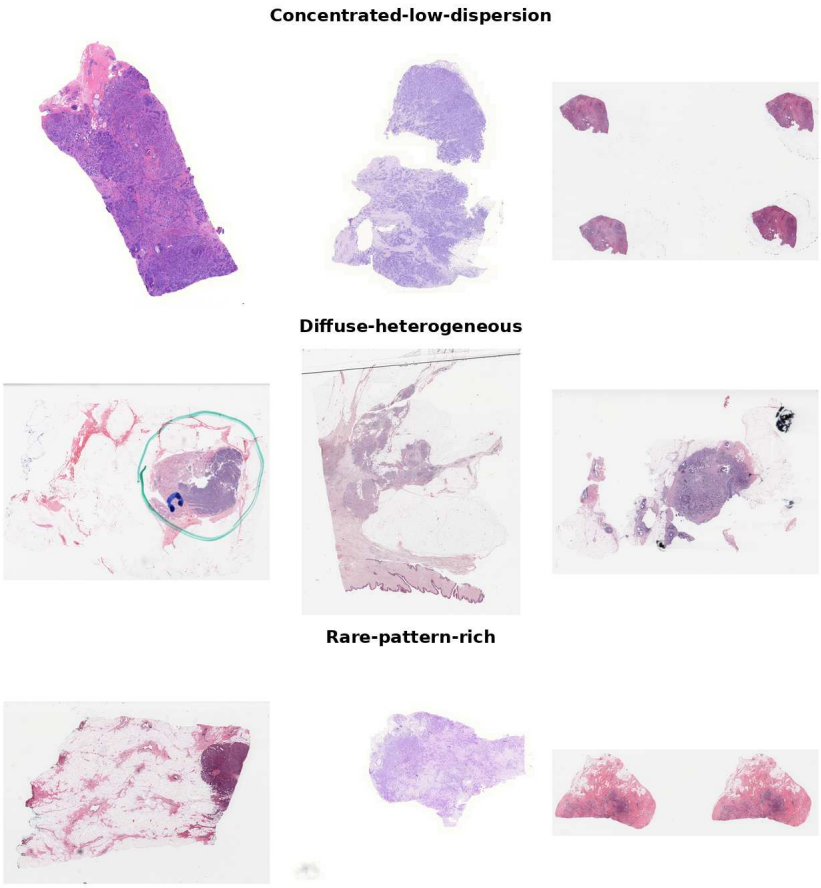


Fig. 3 Representative examples of the three organizational phenotypes. The first row shows concentrated low-dispersion slides, the second row shows diffuse heterogeneous slides, and the third row shows rare-pattern-rich slides. Examples were selected using the metrics defining each phenotype and were used solely for qualitative illustration. Visible markings, when present, originate from the original histopathological slides.

4.5 Extreme Case Studies

For qualitative illustration, three representative slides were selected for each phenotype using the phenotype-specific ranking criteria described in Section 3.7. Figure 3 presents low-resolution thumbnails of the selected cases. The selected cases illustrate visually distinct appearances associated with the three representation-space phenotype definitions. These examples are provided for qualitative illustration and do not constitute independent pathological validation.

Table 5 Stability of heterogeneity metrics across region sizes relative to the 1024 × 1024 baseline in the 173-patient cohort.

Metric	512 vs. 1024	2048 vs. 1024
Feature variance	**0.964**	**0.944**
Mean cosine distance	**0.968**	**0.948**
Maximum embedding-norm proportion	**0.908**	**0.953**
Minimum cluster proportion	0.393	0.107
Phenotype agreement (%)	60.1	45.1
Cohen's κ	0.318	0.286

4.6 Region-Size Ablation

The sensitivity of the proposed measures to region granularity was evaluated using region sizes of 512 × 512, 1024 × 1024, and 2048 × 2048 pixels. The 1024 × 1024 configuration was treated as the baseline. As reported in Table 5, feature variance, mean pairwise cosine distance, and maximum embedding-norm proportion showed strong rank correlations across region sizes, with Spearman coefficients ranging from 0.908 to 0.968. These results indicate that the global dispersion and concentration measures were largely preserved when region granularity changed. In contrast, the minimum cluster proportion was more sensitive to region size. Its correlation with the baseline was 0.393 for 512×512 regions and 0.107 for 2048×2048 regions. Phenotype agreement was moderate, reaching 60.1% for the smaller-region configuration and 45.1% for the larger-region configuration, with corresponding Cohen's κ values of 0.318 and 0.286. Thus, the dispersion and embedding-norm concentration measures exhibited strong rank stability across region sizes, whereas clustering-based rarity and phenotype assignments depended more strongly on the selected spatial scale.

4.7 Encoder Sensitivity Analysis

To evaluate sensitivity to the selected feature extractor, the proposed framework was additionally applied to pathology-specific UNI representations [18] on a pilot subset of 30 slides. The subset was selected reproducibly by random sampling from the patient-independent cohort using a fixed random seed of 42. Because this analysis was conducted on a pilot subset, the results should be interpreted as evidence of cross-encoder sensitivity rather than definitive encoder invariance. The same tiles, region definitions, and heterogeneity metrics were used, with only the underlying representation model changed. Spearman correlations between metric values obtained from ResNet-50 and UNI representations are reported in Table 6. Region feature variance and mean pairwise cosine distance exhibited strong cross-encoder correlations ($\rho = 0.868$ and $\rho = 0.769$, respectively). Maximum embedding-norm proportion showed very strong agreement between encoders ($\rho = 0.960$), while the minimum cluster proportion demonstrated more moderate agreement ($\rho = 0.540$). These results indicate that the rankings of the dispersion and embedding-norm concentration measures were substantially preserved across the two representation spaces, whereas

Table 6 Encoder-sensitivity analysis on a 30-slide pilot subset. Spearman correlations compare metrics computed using ResNet-50 and UNI representations.

Metric	Spearman ρ	p-value
Region feature variance	0.868	5.08×10^{-10}
Mean pairwise cosine distance	0.769	7.02×10^{-7}
Maximum embedding-norm proportion	0.960	4.48×10^{-17}
Minimum cluster proportion	0.540	0.0021

the clustering-derived minimum cluster proportion showed only moderate preservation. Although quantitative values differed across representation spaces, the relative ordering of slides remained largely preserved, particularly for the dispersion-oriented measures.

5 Discussion

This work adopts an analysis-first perspective to characterize tissue organization and heterogeneity in breast cancer whole-slide images. By leveraging region-level representations derived from spatially localized tissue regions, this study shows that organizational variation within WSIs can be quantified using explicitly defined representation-space measures independently of specific prediction tasks. The findings from our study indicate that tissue organization cannot be fully described by a single metric. Rather than relying on a single heterogeneity score, the proposed framework characterizes tissue organization through multiple representation-space properties and evaluates their stability across threshold and spatial-scale perturbations. This analysis-oriented perspective differs from conventional prediction-driven computational pathology pipelines, in which organizational characteristics often remain implicit within a supervised objective. The observed correlation structure identified a strongly shared dispersion component together with less strongly associated embedding-norm concentration and clustering-derived prevalence measures, indicating that no single proposed statistic fully summarizes the region-embedding distribution. The region-aware design is central to this analysis. Aggregating neighboring tiles into intermediate-scale regions introduces a spatially localized representation that enables organizational characteristics to be examined independently of downstream predictive objectives. In contrast to patch-level MIL approaches, heterogeneity in our framework is explicitly defined rather than implicitly embedded in attention mechanisms. The primary contribution of this work is not the introduction of a single heterogeneity metric, but the development of a region-aware analysis framework that combines representation-space characterization, organizational phenotypes, threshold robustness, region-size analysis, sampling-property assessment, and pilot cross-encoder evaluation within a unified pipeline. This design enables systematic investigation of tissue organization independently of supervised prediction objectives. We identify recurring heterogeneity phenotypes—concentrated low dispersion, diffuse-heterogeneous, and rare-pattern-rich—and illustrate their qualitative appearance using representative cases. While no

clinical association is claimed, these phenotypes provide a structured vocabulary for describing tissue organization and may support future hypothesis-driven investigations. The region-size ablation showed that dispersion and feature-space concentration were stable across spatial scales, whereas the clustering-based rarity measure and phenotype assignments were scale dependent. A pilot encoder-sensitivity analysis using pathology-specific UNI representations further demonstrated that the major metric rankings remained substantially preserved across representation spaces. The strongest agreement was observed for maximum embedding-norm proportion and region feature variance, whereas clustering-derived rarity showed lower stability. These findings suggest that the principal organizational characteristics identified by the framework are not exclusively tied to a specific pretrained encoder. Consequently, the proposed phenotypes should be interpreted relative to the selected region granularity rather than as scale-invariant categories. The analysis of slide sampling properties showed that the proposed measures were not strongly associated with the number of retained tiles. However, feature variance and mean pairwise cosine distance exhibited moderate positive associations with the number of constructed regions, while the minimum cluster proportion showed a weaker association. The maximum embedding-norm proportion was strongly negatively associated with region count ($\rho = -0.968$). Accordingly, this measure should be interpreted as a region-count-dependent feature-space concentration statistic rather than a scale-independent estimate of biological pattern dominance. Although its relative slide ordering was largely preserved across region sizes, this rank stability does not eliminate the underlying dependence on region count.

Several limitations should be acknowledged. The primary analysis was conducted on a single TCGA-BRCA cohort using a fixed grid-based region-construction strategy. The encoder-sensitivity analysis was limited to a pilot subset of 30 slides, and the organizational phenotypes were not independently evaluated by pathologists or validated against clinical endpoints. In addition, the maximum embedding-norm proportion was strongly dependent on region count. The minimum cluster proportion was sensitive to both region granularity and clustering resolution, with additional analyses using alternative values of k producing different slide rankings and category assignments. Consequently, this measure should be interpreted as a clustering-derived prevalence descriptor rather than an intrinsic biological characteristic. Consequently, the reported phenotypes should be interpreted as exploratory representation-space descriptors rather than clinically or biologically validated categories.

Overall, this study demonstrates that region-aware, analysis-first approaches offer a principled and interpretable alternative to prediction-driven pipelines for studying tissue heterogeneity in WSIs, particularly in settings with limited or uneven clinical annotation.

6 Conclusion

This work presents an analysis-first, region-aware framework for characterizing tissue organization and heterogeneity in breast cancer whole-slide images. By constructing spatially localized regions and defining unsupervised heterogeneity metrics from

region-level representations, we show that complementary representation-space characteristics of tissue organization can be quantified beyond patch-level appearance. The proposed framework captures complementary aspects of tissue organization through measures of dispersion, embedding-norm concentration, and clustering-based rarity, and enables the identification of recurring organizational phenotypes. Correlation analysis, robustness experiments, and qualitative case studies showed that these metrics provide transparent representation-space descriptions of tissue organization. In addition, a region-size ablation demonstrated that dispersion and embedding-norm concentration measures remain largely stable across spatial scales, while clustering-based rarity and phenotype assignments are more sensitive to region granularity. The sampling-property analysis clarified the dependence of the maximum embedding-norm proportion on region count, while the pilot encoder-sensitivity analysis showed that the main metric rankings were moderately to strongly preserved across ResNet-50 and UNI representation spaces. Overall, the proposed metric framework enables a principled and interpretable characterization of tissue organization without relying on supervised learning objectives or dense annotations. Future work will investigate validation using independently annotated pathology datasets, alternative region construction strategies, and broader applications of analysis-first computational pathology for studying tissue heterogeneity.

Acknowledgments

The results published here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

Ethics statement

This study used publicly available, de-identified data from The Cancer Genome Atlas (TCGA)-BRCA project. No new human participants were recruited, and no identifiable patient information was accessed. Therefore, this study did not require additional ethical approval or informed consent from the authors' institution.

Disclosure of Interests

The author declares that there are no competing interests relevant to the content of this article.

Funding

This research received no external funding.

Data Availability Statement

The TCGA-BRCA whole-slide images analyzed in this study are available from The Cancer Imaging Archive under the applicable access conditions. The derived slide-level metrics, patient-independent slide manifest, phenotype assignments, representative-case list, and analysis code are available from the corresponding author upon reasonable request.

References

[1] Gecer, B., Aksoy, S., Mercan, E., Shapiro, L.G., Weaver, D.L., Elmore, J.G.: Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks. Pattern Recognition **84**, 345–356 (2018) https://doi.org/10.1016/j.patcog.2018.07.022

[2] Kleczek, P., Jaworek-Korjakowska, J., Gorgon, M.: A novel method for tissue segmentation in high-resolution h&e-stained histopathological whole-slide images. Computerized Medical Imaging and Graphics **79**, 101686 (2020) https://doi.org/10.1016/j.compmedimag.2019.101686

[3] Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A.L., Razavian, N., Tsirigos, A.: Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nature medicine **24**(10), 1559–1567 (2018) https://doi.org/10.1038/s41591-018-0177-5

[4] Kather, J.N., Pearson, A.T., Halama, N., Jäger, D., Krause, J., Loosen, S.H., Marx, A., Boor, P., Tacke, F., Neumann, U.P., *et al.*: Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nature medicine **25**(7), 1054–1056 (2019) https://doi.org/10.1038/s41591-019-0462-y

[5] Navikas, V., Kowal, J., Rodriguez, D., Rivest, F., Brajkovic, S., Cassano, M., Dupouy, D.: Semi-automated approaches for interrogating spatial heterogeneity of tissue samples. Scientific Reports **14**(1), 5025 (2024)

[6] Yang, P., Jin, L., Liao, J., Jin, K., Shao, X., Li, C., Qian, J., Cheng, J., Yu, D., Guo, R., et al.: Revealing spatial multimodal heterogeneity in tissues with spatrio. Cell Genomics **3**(12) (2023)

[7] Zhao, T., Chiang, Z.D., Morriss, J.W., LaFave, L.M., Murray, E.M., Del Priore, I., Meli, K., Lareau, C.A., Nadaf, N.M., Li, J., *et al.*: Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. Nature **601**(7891), 85–91 (2022)

[8] Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature medicine **25**(8), 1301–1309 (2019) https://doi.org/10.1038/s41591-019-0508-1

[9] Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: Dy, J., Krause, A. (eds.) Proceedings of the 35th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 80, pp. 2127–2136 (2018). PMLR. https://proceedings.mlr.press/v80/ilse18a.html

[10] Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021) https://doi.org/10.1038/s41551-020-00682-w

[11] Guan, Y., Zhang, J., Tian, K., Yang, S., Dong, P., Xiang, J., Yang, W., Huang, J., Zhang, Y., Han, X.: Node-aligned graph convolutional network for whole-slide image representation and classification. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 18791–18801 (2022). IEEE

[12] Lingle, W., Erickson, B.J., Zuley, M.L., Jarosz, R., Bonaccio, E., Filippini, J., Net, J.M., Levi, L., Morris, E.A., Figler, G.G., et al.: The cancer genome atlas breast invasive carcinoma collection (tcga-brca) (2016) https://doi.org/10.7937/dx22-8j71

[13] Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B.: Histopathological image analysis: A review. IEEE reviews in biomedical engineering **2**, 147–171 (2009)

[14] Liu, L., Liu, Z., Xie, J., Qiao, H., Chang, J.: Transformers in pathological image analysis: A survey. Engineering Applications of Artificial Intelligence **156**, 111114 (2025) https://doi.org/10.1016/j.engappai.2025.111114

[15] Otsu, N.: A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics **9**(1), 62–66 (1979)

[16] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)

[17] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009). IEEE

[18] Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., *et al.*: Towards a general-purpose foundation model for computational pathology. Nature medicine **30**(3), 850–862 (2024)