https://www.perplexity.ai/search/dataset-overview-cell-line-mda-b3z9eT.QTy6gwZWQ.HZqCw?

Processing_features.py

Comprehensive 3D Invasion Analysis Report: MDA-MB-231 Spheroids in Collagen Matrix

Introduction

Your invasion analysis pipeline quantifies invasiveness of MDA-MB-231 triple-negative breast cancer cells invading from 3D spheroids into a collagen matrix. The approach measures both individual cell behavior (morphology, position) and population-level dynamics (distribution patterns, collective migration) to provide a multidimensional assessment of invasion capacity.

Core Measurement Strategy

Fixed Reference Frame Approach

The analysis uses a fixed spheroid center computed from the initial timepoint (t=0) as the reference point for all subsequent measurements. This critical design choice prevents the "floating center" artifact where the reference point would shift with cell migration, masking true invasion dynamics. Each cell's invasion distance is measured as the 3D Euclidean distance from its centroid to this fixed center, after applying proper voxel scaling (1.242 × 1.242 × 6.0 µm).

Nine Complementary Metrics

Your analysis captures invasion through nine distinct but complementary dimensions, organized into five biological categories:

1. Mean Invasion Radius: Overall Penetration Depth

What it measures: The average radial distance of all cells from the spheroid center at each timepoint.

How it's calculated: After applying voxel scaling to convert pixel coordinates to physical units (μm), the 3D Euclidean distance from each cell centroid to the fixed t=0 center is computed as

$$r = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}$$

The mean radius is simply the average of all r values.

Biological meaning: This metric quantifies the overall invasion penetration into the collagen matrix. Increasing mean radius indicates cells are migrating away from the spheroid core into surrounding ECM, while decreasing values suggest retraction or peripheral cell death.

Your data interpretation:

- Test S1 (green): Shows progressive invasion from ~855 to ~915 μm, indicating active outward migration
- Test S2 (red): Maintains small radius (~570 μm), suggesting limited invasion capacity or early-stage spheroid
- Train S1 (blue): Stable around 850-920 μm, representing invasion-contraction equilibrium
- Train S2 (orange): Contracts from ~870 to ~780 μm, possibly indicating peripheral cell death despite high proliferation

Invasiveness ranking: Test S1 > Train S1 ≈ Train S2 > Test S2

2. Leader Cell Invasion Depth: Pioneer Cell Penetration

What it measures: The average distance of the top 10% most invasive cells (leaders) at each timepoint.

How it's calculated: First, the 90th percentile (P90) of all radial distances is computed. Cells with $r \geq \mathrm{P90}$ are classified as "leader cells." The mean leader radius is the average distance of only these leader cells. This represents the pioneering subpopulation at the invasion front.

Biological meaning: Leader cells are a functionally distinct subpopulation that drives collective invasion by exploring the microenvironment, degrading ECM through protease secretion (MT1-MMP, cortactin), and creating tracks for follower cells. They exhibit higher glucose uptake, rely on mitochondrial respiration (OXPHOS), and have elevated cytoskeletal contractility compared to followers. The leader's invasion depth quantifies how far this specialized population penetrates beyond the main cell mass.

Your data interpretation:

- Test S1 (green): Consistently highest leader depth (~1720-1770 μm), indicating the most aggressive and stable leader population
- Test S2 (red): Highly variable (1135-1530 μm), suggesting unstable leader-follower dynamics and frequent position switching
- Train S1/S2: Moderate, stable leader penetration (~1400-1460 μm)

Invasiveness ranking: Test S1 (aggressive leaders) > Train S1/S2 (moderate) > Test S2 (unstable)

3. Cell Dispersion Index: Invasion Heterogeneity

What it measures: The ratio of maximum invasion radius to mean radius, quantifying how much leader cells extend beyond the average population.

How it's calculated:

$$\text{Dispersion Index} = \frac{\text{max radius}}{\text{mean radius}}$$

Biological meaning: This dimensionless metric captures the heterogeneity of invasion patterns. Low dispersion (~1.0-1.5) indicates uniform, spherical expansion where all

cells migrate similarly (collective invasion). High dispersion (>2.5) indicates a few leader cells extend far beyond a compact core, characteristic of leader-follower dynamics. The dispersion index reflects the degree of cell phenotypic plasticity and the strength of leader-follower stratification.

Your data interpretation:

- Test S2 (red): Highest dispersion (~2.8-3.2), revealing a paradox—small mean radius but extreme heterogeneity. This indicates a compact core with few far-reaching leaders, suggesting strong leader-follower distinction
- Test S1 (green): Moderate dispersion (~2.0-2.2), balanced heterogeneity
- Train S1/S2: Low dispersion (~1.7-1.8), indicating uniform, collective migration

Pattern types:

- High dispersion = Leader-follower invasion (individual pioneers)
- Low dispersion = Collective invasion (cohesive migration)

4. Nearest Neighbor Distance: Cell-Cell Spacing

What it measures: The average distance from each cell to its closest neighboring cell, quantifying spatial clustering versus dispersal.

How it's calculated: A pairwise distance matrix (N × N) is computed between all cells at a given timepoint. For each cell, the minimum distance to any other cell is identified. The mean NN distance is the average of these minimum distances.

Biological meaning: Cell-cell spacing reflects the mode of invasion—collective versus individual migration. Low NN distances (<80 μm) indicate cells are tightly clustered, maintaining cell-cell contacts (E-cadherin, N-cadherin) characteristic of collective invasion. High NN distances (>100 μm) suggest cells have lost contact and invade independently. Decreasing NN distance over time indicates cells are aggregating, transitioning from individual to collective migration.

Your data interpretation:

- Train S1 (blue): Shows dramatic transition from highly dispersed (~135 μm) to tightly clustered (~75 μm), revealing a shift from individual to collective invasion mode over the 880-minute time course
- Other samples: Maintain relatively stable NN spacing (70-95 μm), indicating consistent invasion strategies

Migration modes:

- NN < 80 μm = Collective migration (cell-cell contacts maintained)
- NN > 100 μm = Individual migration (cells separated)
- Decreasing NN = Aggregation/collective behavior emerging

5. Cell Density: Proliferation-Migration Balance

What it measures: The number of cells per unit invaded volume, reflecting cell packing and the balance between proliferation and spatial expansion.

How it's calculated:

$$\text{Cell Density} = \frac{\text{cell count}}{\text{invasion volume}}$$

where invasion volume is approximated as a sphere:

$$V = \frac{4}{3}\pi(\text{max radius})^3$$

Biological meaning: Cell density integrates two competing processes—proliferation (increases density) and migration/dispersion (decreases density). High density with moderate invasion suggests proliferation-dominant phenotype, while low density with large invasion indicates migration-dominant behavior. Density also reflects cell packing efficiency and local nutrient availability.

Your data interpretation:

- Train S2 (orange): Highest density (~0.005-0.006 $\times 10^{-3}$ cells/μm^3), indicating tightly packed cells despite moderate invasion—proliferation-dominant
- Test S1 (green): Lowest density (~0.002), indicating dispersive invasion—cells spread through large volume
- Phenotypes:
    - High density + moderate radius = Proliferative invasion (Train S2)
    - Low density + large radius = Dispersive invasion (Test S1)

6. Leader Fraction: Relative Leader Population

What it measures: The proportion of the total cell population classified as leaders (top 10% invaders).

How it's calculated:

$$\text{Leader Fraction} = \frac{\text{number of cells with } r \geq \text{P90}}{\text{total cell count}}$$

Biological meaning: This metric validates the leader definition and detects shifts in the relative size of the pioneer population. By definition, leader fraction should approximate 0.10 (10%). Deviations indicate whether more or fewer cells are pushing into the leading edge.

Your data interpretation: All samples fluctuate around 0.10-0.125, with minor variations (<2%) likely representing statistical noise. The consistency confirms the leader identification method is working correctly across samples and timepoints.

7. Distribution Skewness: Invasion Front Asymmetry

What it measures: The asymmetry of the radial distance distribution, quantifying whether cells are symmetrically distributed or have an elongated tail.

How it's calculated:

$$\text{Skewness} = \frac{E[(r-\mu)^3]}{\sigma^3}$$

Biological meaning: Skewness characterizes the shape of the invasion front:

- Positive skewness (right tail): Few cells extend far beyond the bulk population, creating an asymmetric distribution characteristic of leader-follower invasion
- Near-zero skewness: Symmetric distribution where cells are uniformly distributed around the mean, indicating collective invasion without pronounced leaders
- Negative skewness (left tail): Cells concentrated at the front with trailing tail (rare in invasion)

High positive skewness correlates with tumor heterogeneity, aggressiveness, and poor prognosis in clinical studies.

Your data interpretation:

- Test S2 (red): Highest positive skewness (~1.6-2.3), indicating strong leader-follower distinction with long right tail (distinct pioneer population)
- Test S1 (green): Moderate positive skewness (~0.4-0.7), some leader extension
- Train S1/S2: Near-zero or slightly negative skewness, indicating symmetric, collective invasion without pronounced leaders

Clinical relevance: Tumors with higher skewness show increased invasiveness and complexity.

8. 90th Percentile Radius: Leading Edge Boundary

What it measures: The radial distance threshold that separates the top 10% invading cells (leaders) from the remaining population.

How it's calculated: Radial distances are sorted in ascending order, and the value at position

$0.9\times\mathrm{N}$

$0.9\times N$ is identified as P90.

Biological meaning: P90 defines the spatial boundary of the leading edge. While mean leader radius quantifies the average penetration depth of leaders, P90 specifies where the leading edge begins spatially. This metric is commonly used in invasion quantification to delineate the invasion front from the core.

Your data interpretation:

- Test S1 (green): Highest P90 (~1595-1640 μm), confirming extensive and stable leading edge
- Test S2 (red): Highly variable P90 (710-1280 μm), indicating unstable leading edge position
- The P90 values closely parallel the mean leader radius trends, validating consistency between metrics

9. Total Cell Count: Proliferation-Death Dynamics

What it measures: The total number of segmented cells at each timepoint, reflecting the balance between cell proliferation and cell death/loss.

How it's calculated: Direct count of all segmented 3D cell objects in the image volume at each timepoint.

Biological meaning: Cell count dynamics reveal the proliferation versus migration trade-off:

- Increasing count: Proliferation exceeds cell death and migration out of the field of view
- Decreasing count: Cell death or migration exceeds proliferation
- Stable count: Equilibrium between proliferation, death, and migration

Leader cells are typically in G1 phase during active invasion and less proliferative, while follower cells may proliferate more. High cell count with limited invasion suggests proliferation-dominant strategy, while low count with extensive invasion indicates migration-dominant behavior.

Your data interpretation:

- Train S2 (orange): Highest count (66-84 cells), indicating proliferation-dominant phenotype despite radius contraction
- Test S2 (red): Lowest count (30-50 cells), suggesting migration-dominant or stressed phenotype
- Test S1/Train S1: Moderate counts (38-53 cells) with fluctuations indicating proliferation-death balance

Integrated Biological Insights

Sample Phenotype Classification

Combining all nine metrics reveals four distinct invasion phenotypes:

1. Test S1 (Green) - Aggressive Dispersive Invasion

- Large invasion radius (~900 μm)
- Highest, most stable leader depth (~1750 μm)
- Moderate dispersion and skewness
- Low cell density
- Interpretation: Cells prioritize migration over proliferation, with aggressive leader cells creating stable invasion tracks. Dispersive strategy with strong leader-follower coordination.

2. Test S2 (Red) - Heterogeneous Core-Leader Invasion

- Small invasion radius (~570 μm)
- High dispersion (~3.0) and skewness (~2.0)
- Variable leader depth and P90

- Low proliferation (fewest cells)
- Interpretation: Compact core with few far-reaching leaders. Strongest leader-follower distinction. Possibly early-stage spheroid or stressed phenotype with limited resources.

3. Train S1 (Blue) - Transitioning Collective Invasion

- Moderate invasion radius (~870 μm)
- Dramatic decrease in NN distance (135→75 μm)
- Low dispersion (~1.7) and near-zero skewness
- Interpretation: Cells transitioning from individual dispersed migration to collective cohesive invasion. Emerging cell-cell contacts and coordinated movement.

4. Train S2 (Orange) - Proliferative Retraction

- Contracting radius (870→780 μm)
- Highest cell density and cell count
- Low dispersion
- Interpretation: Proliferation-dominant phenotype with possible peripheral cell death causing inward contraction despite high cell numbers. Atypical invasion pattern.

Key Biological Mechanisms

Leader-Follower Dynamics: Leader cells exhibit higher energetic demands (glucose uptake, OXPHOS), create ECM tracks through protease activity, and dynamically switch positions when energy depletes. Your metrics (leader depth, dispersion, skewness) capture the strength and stability of this stratification.

Collective vs. Individual Invasion: Nearest neighbor distance reveals the invasion mode—collective invasion maintains cell-cell contacts (<80 μm spacing), while individual invasion involves separated cells (>100 μm). Train S1's decreasing NN indicates collective behavior emergence.

Proliferation-Migration Trade-off: Cell density and count reveal whether cells prioritize proliferation (high density, high count) or migration (low density, large radius). Test S1 shows migration-dominant, Train S2 shows proliferation-dominant strategies.

Invasion Heterogeneity: Dispersion index and skewness quantify phenotypic heterogeneity—whether all cells behave similarly (uniform invasion) or distinct subpopulations exist (leader-follower). Test S2 shows highest heterogeneity.

Quantifying Invasiveness: Summary Recommendations

To rank invasiveness across your samples, integrate multiple metrics:

Primary Metrics:

1. Mean invasion radius: Overall penetration depth
2. Leader invasion depth: Pioneer cell aggressiveness
3. Dispersion index: Invasion pattern heterogeneity

Secondary Metrics:
4. Nearest neighbor distance: Invasion mode (collective vs. individual)
5. Cell density: Proliferation-migration balance
6. Skewness: Leader-follower distinction

Overall Invasiveness Ranking:

- Most invasive: Test S1 (aggressive dispersive invasion)
- Moderate: Train S1 (collective invasion), Train S2 (proliferative, atypical)
- Least invasive: Test S2 (limited penetration, unstable)

These metrics provide a comprehensive, multidimensional quantification of invasion that captures spatial patterns, temporal dynamics, cell-cell interactions, and phenotypic heterogeneity—essential for understanding the complex biology of cancer cell invasion in 3D matrices