大腸がん検体における空間トランスクリプトームおよびH&E染色画像のマルチモーダル情報に基づく免疫状態の空間的不均一性の解析

プロジェクト: GAC25019_MID06_crc_spatial

Abstract

本解析では、大腸がん検体における上皮細胞集団の多様性と免疫微小環境の空間的特徴を明らかにするため、空間トランスクリプトームと隣接切片のH&E染色画像を用いたマルチモーダル解析を実施した。具体的には、遺伝子発現情報に基づく細胞種アノテーションおよびcopy number variation（CNV）の推定結果を、H&E染色画像に基づく組織形態学的所見と統合し、各細胞集団の分子的特徴と空間分布を評価した。

組織上に対応付けられた4,366スポットのうち、上皮細胞に分類されたスポットから、KLK6/CEACAM6高発現、EREG高発現、および杯細胞・分泌型の3つの主要なサブクラスターが同定された。これらのサブクラスター間ではRNA由来CNVの推定結果が大きく異なり、CNV陽性判定率はKLK6/CEACAM6高発現サブクラスターで85.7%であったのに対し、杯細胞・分泌型サブクラスターでは1.6%であった。以上から、同一検体内の上皮細胞集団には、遺伝子発現特性のみならず、RNA発現情報から推定されるCNVパターンにも顕著な不均一性が存在することが示された。

また、各上皮サブクラスターは異なる空間分布を示し、それぞれが局所的に濃縮する領域が認められた。H&E染色画像との対応を評価したところ、杯細胞・分泌型サブクラスターの濃縮領域は、淡明な空胞を伴う陰窩様腺構造が認められる組織辺縁部に位置していた。一方、KLK6/CEACAM6高発現およびEREG高発現サブクラスターの濃縮領域では、いずれも多様な腺管構造に加え、小型類円形核を有する細胞や間質成分の混在が認められた。これらの結果は、杯細胞・分泌型サブクラスターの転写特性が特徴的な腺構造と対応する一方、類似した腺管形態を呈する領域においても、異なる分子的特徴を有する上皮細胞集団が存在することを示している。

免疫微小環境については、CD163/MSR1高発現を特徴とする骨髄系細胞群に着目し、マクロファージ受容体関連遺伝子群の発現スコアと上皮細胞からの空間的距離との関係を評価した。その結果、同スコアは全上皮細胞を参照集団とした最近傍距離と弱い正の相関を示し、骨髄系細胞のマクロファージ受容体関連転写状態が、上皮細胞との空間的位置関係に応じて変化する可能性が示された。

以上より、本検体では、遺伝子発現特性およびRNA由来CNVの異なる複数の上皮細胞集団が組織内に局所的に偏在し、その一部が特徴的な組織形態と対応することが示された。さらに、骨髄系細胞の転写状態と上皮細胞への近接性との関連も認められた。

Results and Discussion

空間トランスクリプトーム情報に基づく細胞アノテーション

最適なクラスタリング条件の探索

主要な細胞集団を適切に識別しつつ、安定したクラスタリング条件を選定するため、複数のresolutionでクラスタリングを実施し、その安定性とクラスター分離度を比較した。クラスタリング結果の安定性はadjusted Rand index（ARI）、クラスター間の分離度は主成分分析（PCA）空間におけるシルエット係数を用いて評価した。

PCAでは50主成分を算出し、このうち第1–16主成分（累積寄与率75.9%）を後続解析に使用した。これらの主成分に基づき、k = 20としてshared nearest neighbor（SNN）グラフを構築し、Louvain法によるクラスタリングを実施した。複数のresolutionを比較した結果、最終的に採用したresolution = 0.6付近では、resolution = 0.4との間のARIが0.994と高く、クラスタリング結果が安定していることが確認された（Fig. 1a）。また、resolution = 0.6におけるPCA空間のシルエット係数は0.246であった（Fig. 1b）。

一方、数値指標のみではより高いシルエット係数を示す低resolution条件も認められたが、この条件ではT細胞と骨髄系細胞が同一クラスターに統合され、主要な細胞種を十分に識別できなかった。これに対し、resolution = 0.6ではT細胞と骨髄系細胞を独立したクラスターとして分離できるとともに、血管内皮細胞および収縮性間質細胞の集団も個別に保持された。

以上より、ARIおよびシルエット係数による数値的評価に加え、主要な細胞種を生物学的に妥当な単位で識別できることを総合的に考慮し、第1–16主成分を用いてk = 20のSNNグラフを構築し、Louvain法、resolution = 0.6でクラスタリングする条件を最終条件として採用した。

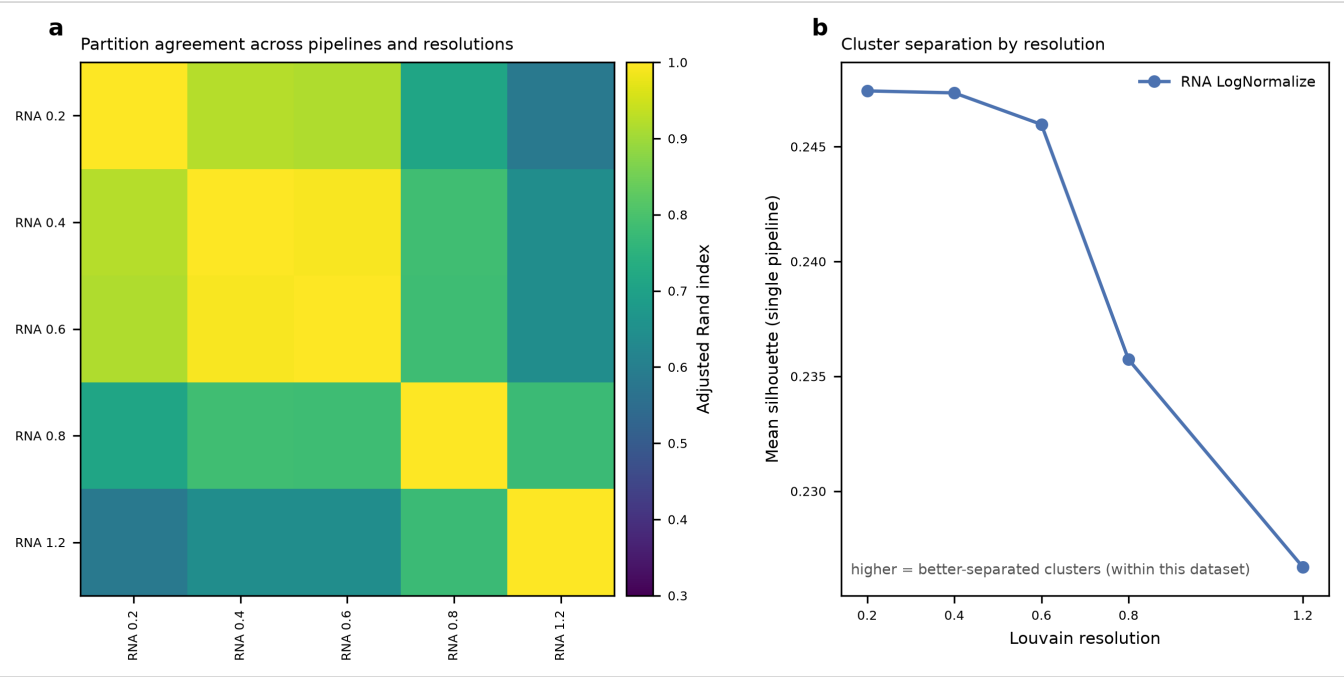


Fig. 1. クラスタリング条件の安定性とクラスター分離度の評価。

(a) 異なるresolutionで得られたクラスタリング結果間のadjusted Rand index（ARI）。

(b) 各resolutionにおけるPCA空間上の平均シルエット係数。

UMAPによるトランスクリプトーム分布の可視化

空間トランスクリプトーム解析により同定された細胞集団の分布を可視化するため、前節で決定したクラスタリング結果を固定した上で、UMAPのハイパーパラメータを比較し、可視化条件を最適化した。まず、初期条件として（metric = cosine、n_neighbors = 30、min_dist = 0.3）でUMAPによる可視化を実施し、最終的なクラスターについて、UMAP空間上で各細胞集団の分布を確認した（Fig. 2a）。続いて、クラスター内の凝集性、クラスター間の分離、および全体の分布構造を評価し、75通りのUMAP条件を比較した（Fig. 3b）。

その結果、metric = euclidean、n_pcs = 12、n_neighbors = 50、min_dist = 0.01をハイパーパラメータとして採用した（Fig. 3a）。注意する点として、これらのパラメータはUMAPによる低次元表現を最適化するためのものであり、前節で決定したクラスタリング結果自体は変更していない。

以上より、細胞集団の定義とUMAPによる可視化を独立して最適化することで、既存のクラスター構造を維持しながら、各細胞集団の関係を視覚的に把握しやすいUMAPによる可視化条件を決定した。以降の解析では、これらの空間上で可視化された各クラスターを対象に、細胞種アノテーションを実施した。

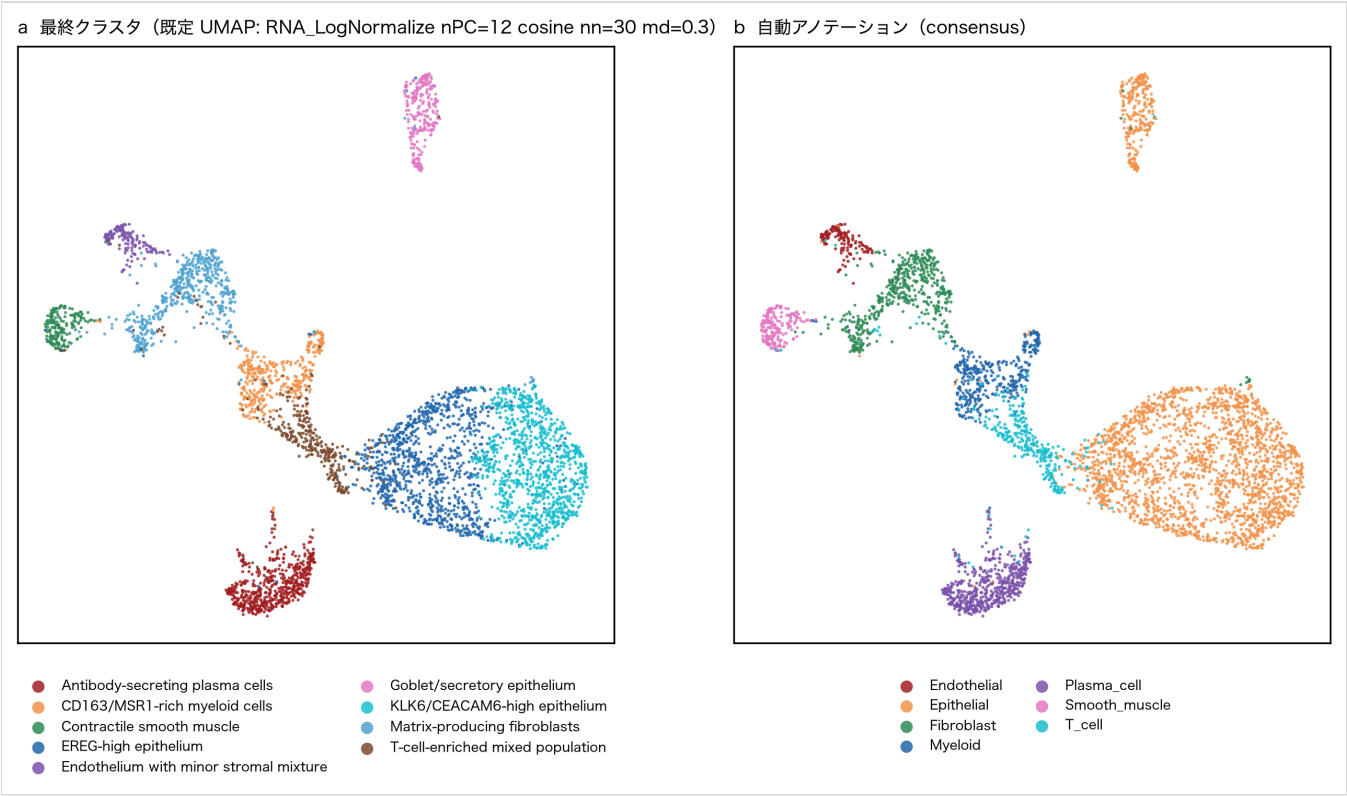


Fig 2. トランスクリプトーム情報に基づく細胞アノテーション (a) 初期UMAP条件（metric = cosine、n_neighbors = 30、min_dist = 0.3）における最終的な細胞集団。(b) 複数のアノテーション手法を統合したコンセンサス細胞種。

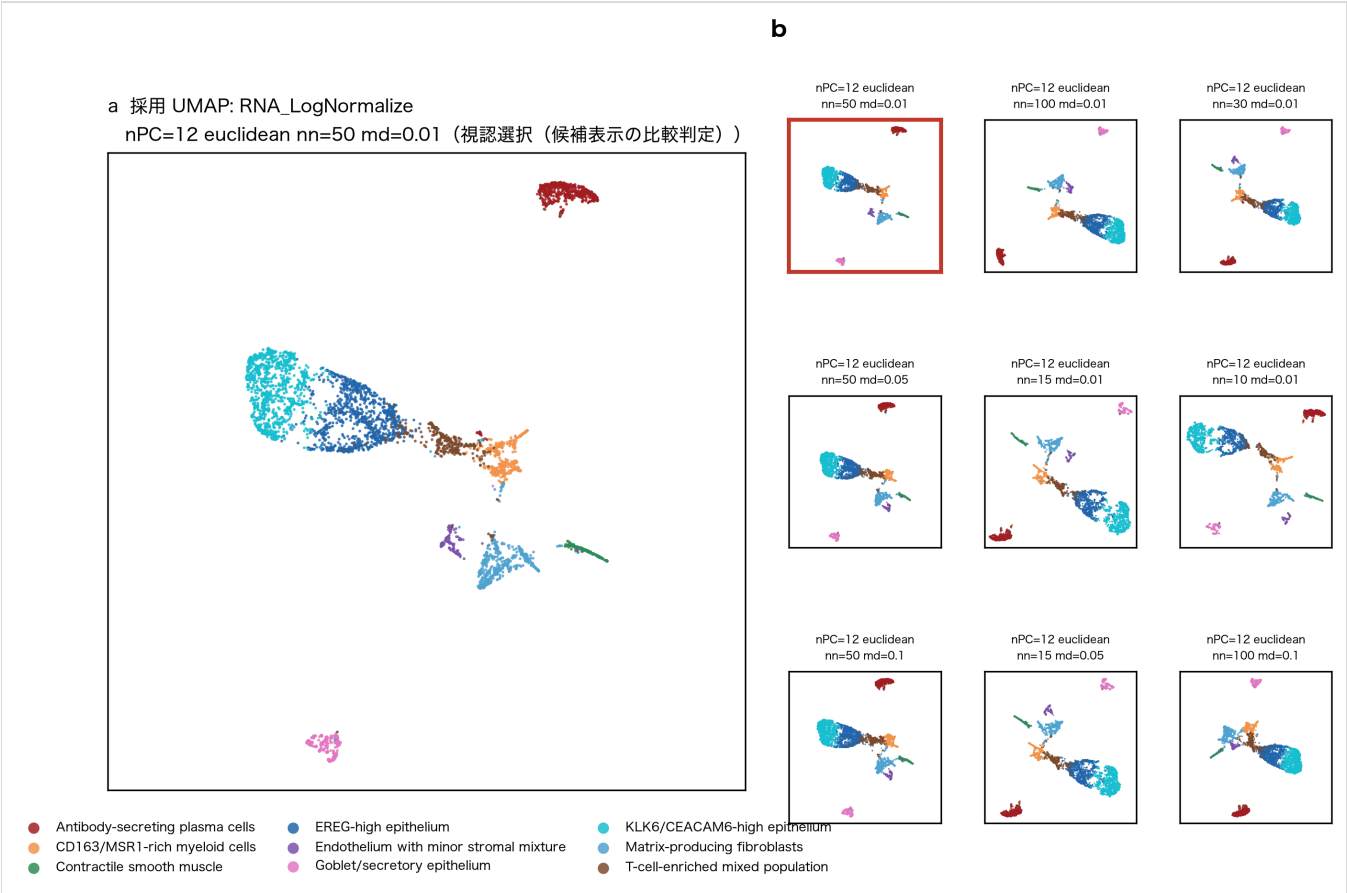


Fig 3. UMAPによるトランスクリプトームプロファイルの可視化 (a) 選定されたUMAP条件。(b) 比較検討された可視化条件の一例。

細胞種アノテーション

各クラスターの細胞種を推定するため、複数の独立したアノテーション手法を組み合わせて解析を実施した。具体的には、CELLxGENE Census由来の大腸組織参照マーカーを用いたAddModuleScoreによるマーカースコアリング、各クラスターのdifferentially expressed genes（DEGs）と参照マーカーとのFisher検定による対応付け、CellTypistの腸管アトラスモデル（Cells_Intestinal_Tract）、および免疫細胞に対するImmune_All_Lowモデルの4手法から細胞種を推定した。各手法から得られた判定を統合し、一致率80%以上および3手法以上で判定が一致したスポットを高信頼アノテーションとした。

その結果、4,387スポットはEpithelial、Plasma cell、Fibroblast、Myeloid、T cell、Smooth muscle、Endothelialの7つの主要な細胞系統に分類され、このうち2,159スポット（49.2%）が高信頼基準を満たした。特に、Plasma cell、Fibroblast、Myeloid、Smooth muscle、およびEndothelialでは、各系統に特徴的なマーカー発現と複数のアノテーション経路が高い一致を示した。一方、T細胞を主体とするクラスターでは、参照モデル間でTreg、central memory T cell、naive/helper T cellなど判定が分かれ、少量の間質系細胞も含まれていた。このため、当該クラスターは特定のT細胞サブタイプに限定せず、「T-cell-enriched mixed population」として扱った。

最終的な解析単位には、前節で採用したresolution = 0.6の9クラスターをそのまま用い、アノテーション結果に基づくクラスター間の統合は行わなかった。上皮系では、同一のEpithelial系統に分類されたスポットが3つのクラスターに分離しており、それぞれ異なるマーカー発現、上皮状態スコア、およびRNA由来CNVパターンを示した。このため、これらをKLK6/CEACAM6-high（n = 1,011）、EREG-high（n = 1,006）、Goblet/secretory（n = 253）の3つの上皮サブクラスターとして保持した。また、Fibroblastの大部分は単一クラスターに集積したが、一部はT細胞および内皮系クラスターに少数混在していたため、これらについてはminor stromal mixtureとして解釈した。

なお、独立したnon-plasma B-cellクラスターは本解析では同定されなかった。B-lineageについてはPlasma cell集団が複数の解析経路から明確に支持された一方、それ以外のB細胞については、現在のクラスタリングおよびアノテーション解像度では独立した集団として確定できなかった。

以上より、複数の独立したアノテーション手法を統合することで主要な細胞系統を同定するとともに、判定が一致しない集団については過度に細分化せず、解析上の不確実性を保持した細胞集団定義を採用した。

トランスクリプトーム情報に基づくcopy number variation解析

上皮サブクラスター間の分子的特徴をさらに評価し、腫瘍性上皮の候補となる集団を探索するため、非上皮細胞を参照集団としてRNA発現情報からcopy number variation（CNV）を推定した。その結果、全4,387スポットのうち1,267スポットがCNV陽性と判定された（Fig. 4a–b）。また、推定されたCNVパターンに基づいてスポットを分類したところ、各CNVパターン群で染色体レベルの平均CNVプロファイルに異なる特徴が認められた（Fig. 4c）。

空間情報との対応が得られた上皮スポットについてサブクラスター別にCNV陽性率を比較したところ、KLK6/CEACAM6-highでは866/1,010スポット（85.7%）、EREG-highでは318/1,000スポット（31.8%）、Goblet/secretoryでは4/248スポット（1.6%）がCNV陽性であった（Table 1）。さらに、CNVスコア、CNV判定、および上皮細胞におけるCNVパターンは、UMAP空間上でも異なる分布を示した（Fig. 4d–f）。以上の結果から、KLK6/CEACAM6-highサブクラスターは、特徴的なマーカー発現に加えて高いCNV陽性率を示す上皮集団として、腫瘍性上皮の候補と位置付けた。

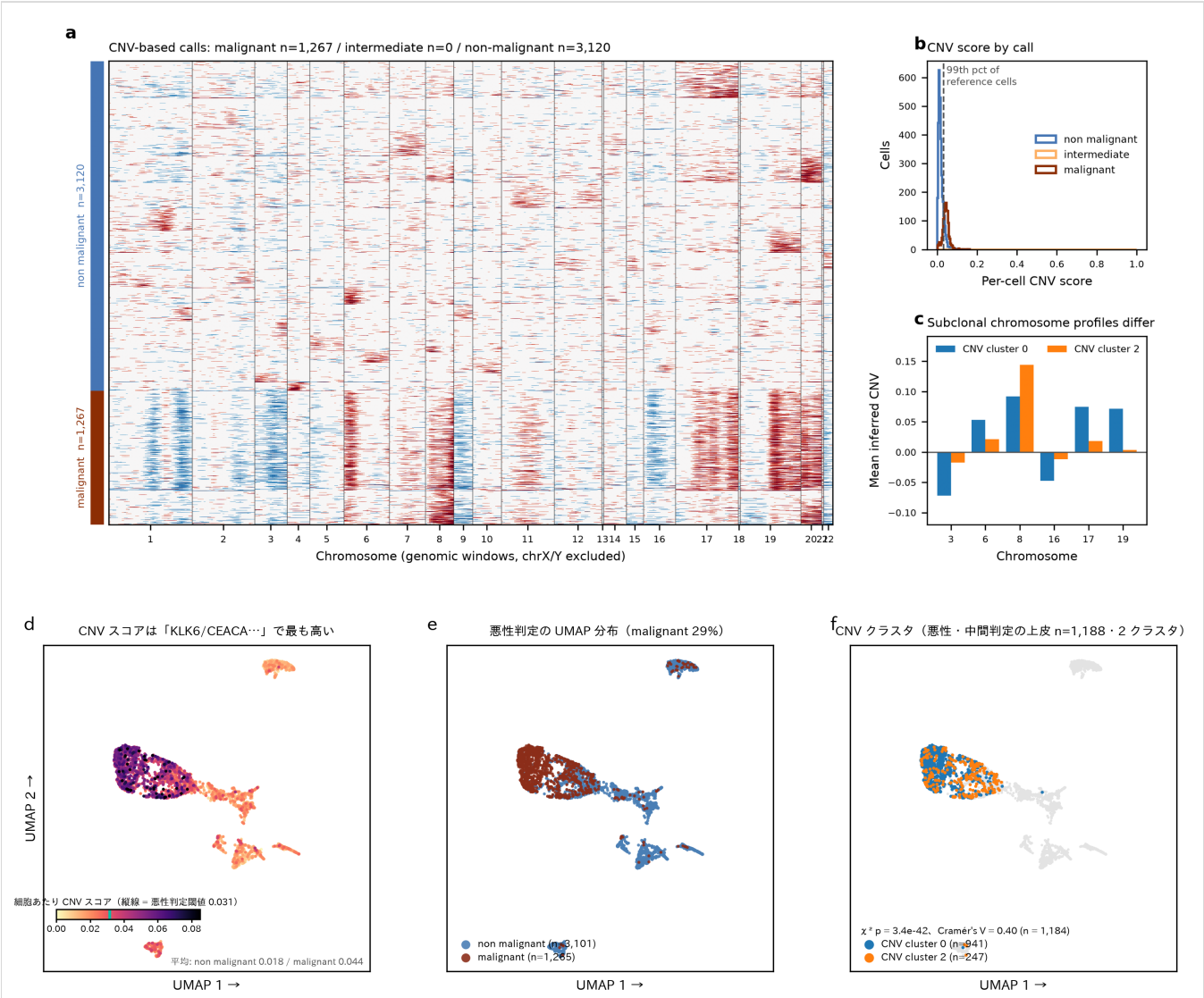


Fig. 4. RNA発現情報に基づくCNV推定と上皮サブクラスターとの関連。 (a) RNA由来CNVプロファイルのヒートマップ。(b) 各スポットのCNVスコア分布。(c) CNVパターン群ごとの代表的な染色体における平均CNVプロファイル。(d) UMAP空間上におけるCNVスコアの分布。(e) UMAP空間上におけるCNV判定の分布。(f) CNV陽性上皮スポットにおけるCNVパターン群の分布。

H&E染色画像の解析

全スライドにおける形態領域の分類

空間トランスクリプトームの測定領域を含む切片全体の組織構造を把握するため、H&E染色画像から抽出した画像特徴量に基づいて組織タイルを分類した。その結果、3,140タイルが9種類の形態領域に分類され、解析対象となったタイルの総面積は160.42 mm²であった（Fig. 5、Table 3）。

各形態領域は、腺管構造に富む領域、紡錘形細胞を含む間質領域、緻密な好酸性間質、脂肪組織、および複数の組織成分が混在する領域など、それぞれ異なる組織学的特徴を示した。また、これらの形態領域は切片

内で異なる空間分布を示した（Fig. 5b）。各領域の代表的な高解像度H&E画像と画像特徴量を比較したところ、それぞれの形態学的特徴と分類結果との対応が確認された（Fig. 6）。

以上の形態領域分類により、空間トランスクリプトームから同定された細胞集団の局所的な分布を、切片全体の組織構造と対応付けて評価することが可能となった。

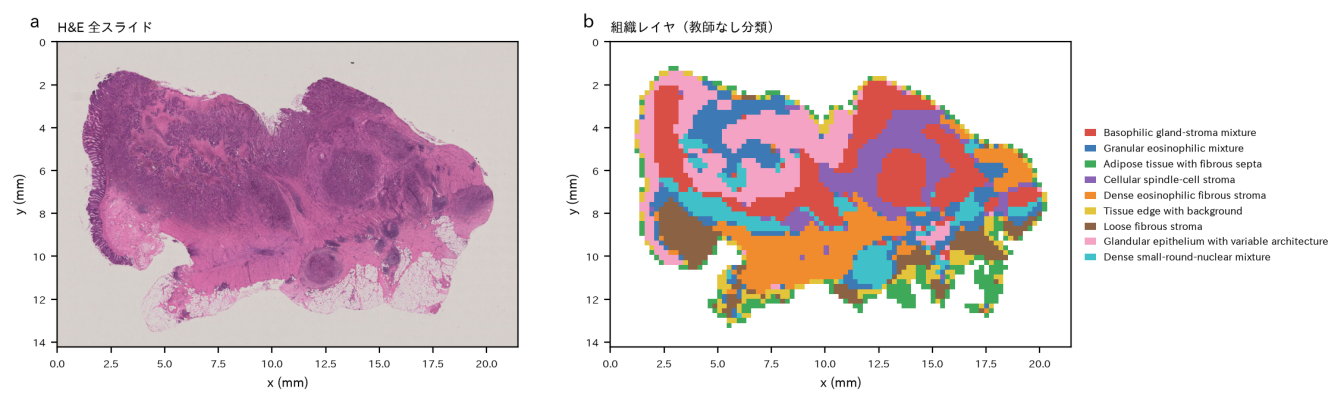


Fig. 5. H&E画像における組織アノテーション (a) 全スライドH&E画像。(b) 組織タイルの教師なし分類に基づく形態領域マップ。

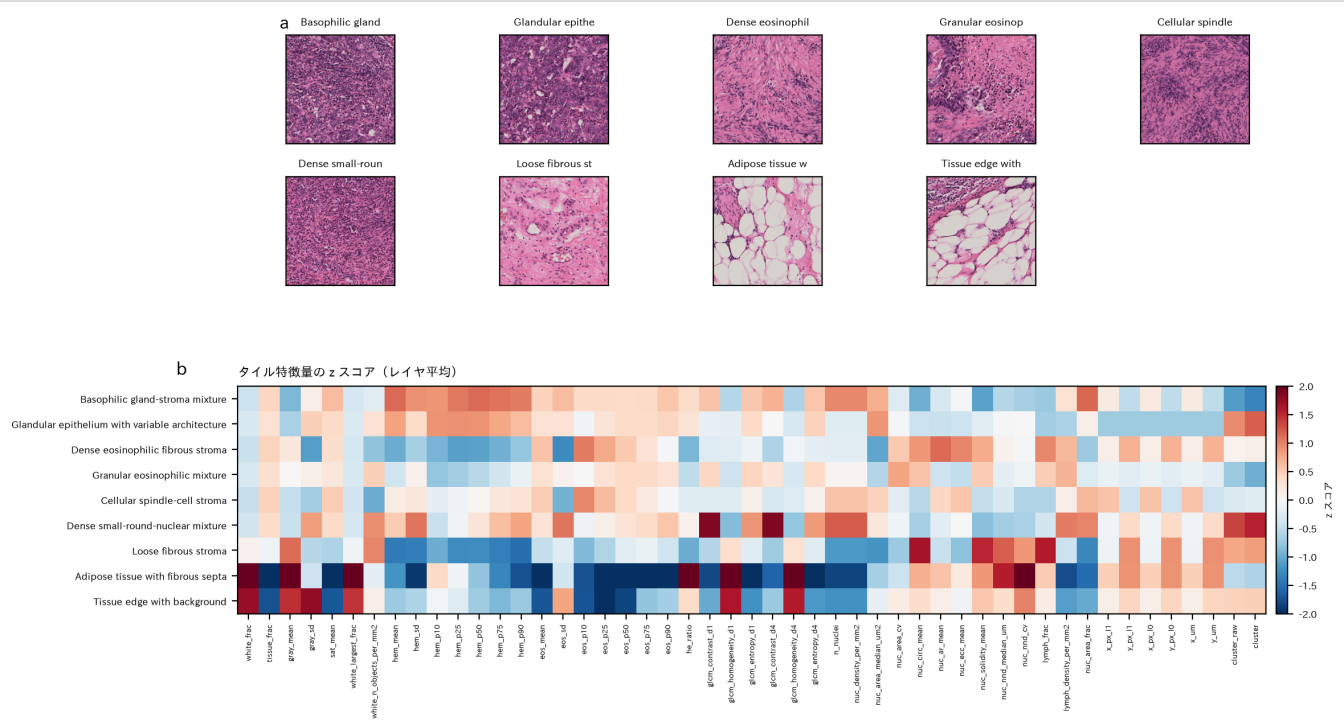


Fig. 6. 各形態領域の代表的な組織像と画像特徴量。 (a) 各形態領域を代表するH&E染色像。各領域の名称はTable 3に示す。(b) 各形態領域における標準化画像特徴量の比較。座標およびクラスターコードは解析管理用の変数として示す。

空間座標とH&E染色画像の位置合わせ

遺伝子発現に基づいて同定した細胞集団の空間分布を組織形態と対応付けるため、高信頼アノテーションが得られた2,159スポットを基準として、空間トランスクリプトームの座標とH&E染色画像の位置合わせを行った。

位置合わせ後、フィッティングに使用したスポットの99.4%がH&E画像上の組織領域内に配置された。採用した座標変換における目的関数値は0.1104であり、y軸反転を含む代替条件では0.0797であった。最終的には、組織形態との対応を含めて位置関係を評価し、追加のy軸反転を行わない座標変換を採用した。この変換により、空間トランスクリプトームの測定領域はH&E染色切片の左側領域に対応した（Fig. 7a–b）。

位置合わせ後の細胞集団と組織形態との対応を評価したところ、Contractile smooth muscle cellsは緻密な好酸性線維性領域に集積していた（Fig. 7c）。また、その他の主要な細胞集団についても、H&E染色画像から分類された組織領域に応じて異なる空間分布が認められた（Fig. 7d）。これらの結果から、空間トランスクリプトームで得られた細胞集団の分布とH&E染色画像上の組織形態との間に、領域レベルでの対応が確認された。

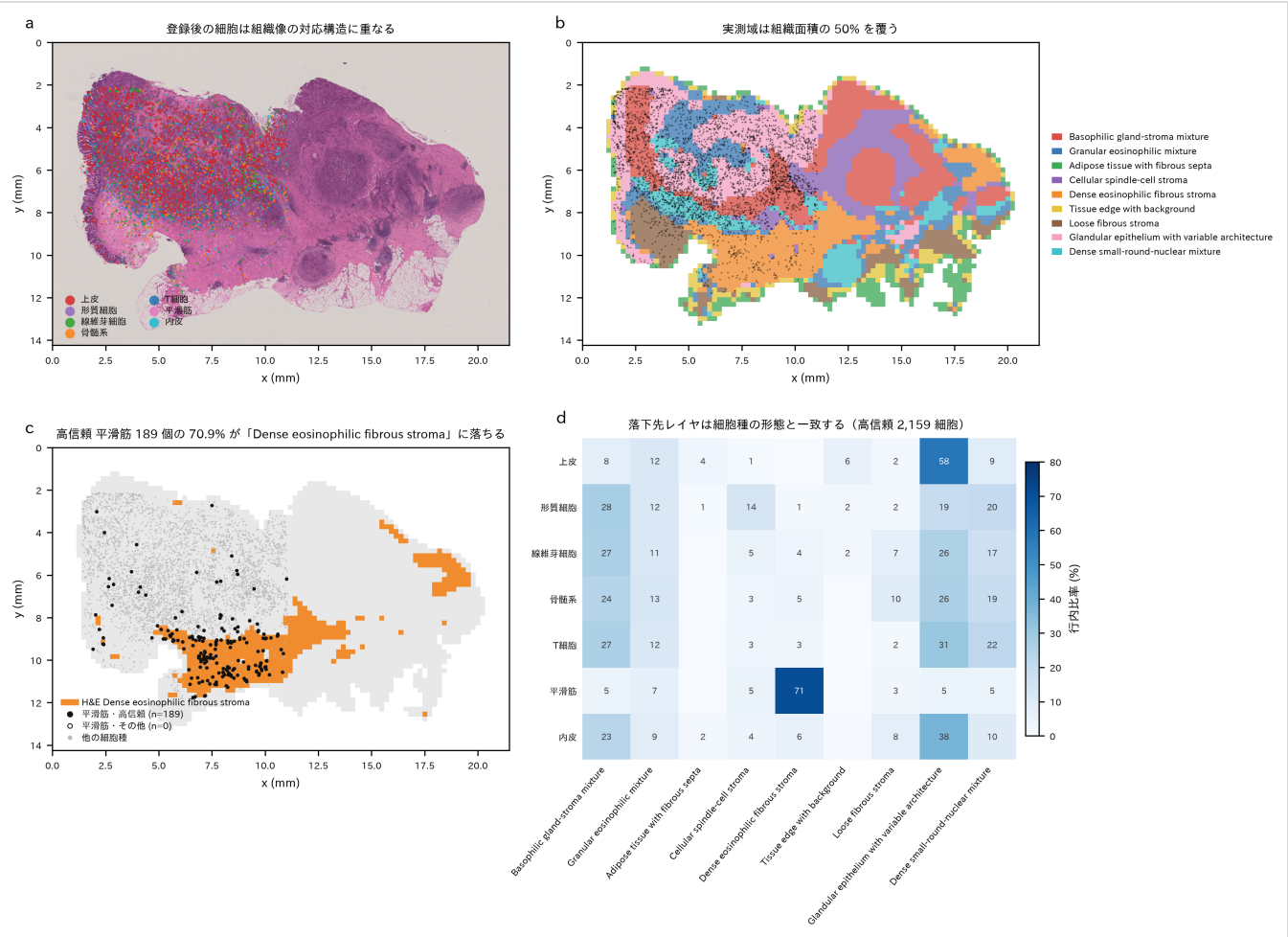


Fig. 7. 空間トランスクリプトーム座標とH&E染色画像の位置合わせ。 (a) H&E染色画像上に投影した組織上の全4,366スポットの細胞集団。(b) 形態領域マップ上における空間トランスクリプトーム測定領域。(c) 高信頼アノテーションの細胞集団と組織領域の組み合わせのうち最も強い集積を示した、Contractile smooth muscleと緻密な好酸性線維性領域の対応（134/189スポット）。(d) H&E染色画像から分類した組織領域ごとの高信頼アノテーション細胞集団の分布。

マルチモーダル情報に基づく細胞アノテーションの決定

最終的な細胞集団を定義し、後続の空間的不均一性解析の基盤とするため、遺伝子発現に基づく細胞種アノテーション、RNA由来CNV、H&E染色画像上の空間分布、マーカー遺伝子の発現、および参照マーカースコアを統合して各クラスターの特徴を評価した（Fig. 2b、Fig. 4、Fig. 8）。

位置合わせ後の組織領域には4,366スポットが含まれ、最終的に9種類の細胞集団に分類された（Table 1）。上皮系では、KLK6/CEACAM6-high epithelium（n = 1,010）、EREG-high epithelium（n = 1,000）、およびGoblet/secretory epithelium（n = 248）の3つのサブクラスターが同定された。これらはマーカー遺伝子の発現に加え、RNA由来CNVの陽性率およびCNVパターンにも異なる特徴を示したことから、それぞれ独立した上皮サブクラスターとして保持した。

免疫系では、Antibody-secreting plasma cells（n = 565）、CD163/MSR1高発現を特徴とするCD163/MSR1-rich myeloid cells（n = 374）、およびT-cell-enriched mixed population（n =

332）の3集団が同定された（Fig. 8c）。このうちT-cell-enriched mixed populationでは、複数のアノテーション手法間でT細胞サブタイプの判定が一致せず、少数の間質系スポットも含まれていた。そのため、特定のT細胞サブタイプには限定せず、アノテーションの不確実性を保持した集団として扱った。

間質・血管系では、細胞外基質関連遺伝子の発現を特徴とするMatrix-producing fibroblasts（n = 509）、収縮関連遺伝子の発現を特徴とするContractile smooth muscle cells（n = 189）、およびEndothelium with minor stromal mixture（n = 139）の3集団が同定された（Fig. 8c）。一部のクラスターには少数の異なる細胞系統が混在していたが、主要なマーカー発現およびアノテーション結果に基づいて各集団を定義した。

これら9集団はH&E染色画像上でも異なる空間分布を示し（Fig. 8b）、組織内における局所的な細胞組成の違いが認められた。以上より、主要な細胞系統を識別するとともに、上皮細胞集団の分子的多様性およびアノテーションの不確実性を保持した9種類の細胞集団を、以降の空間解析における最終的な解析単位として採用した。

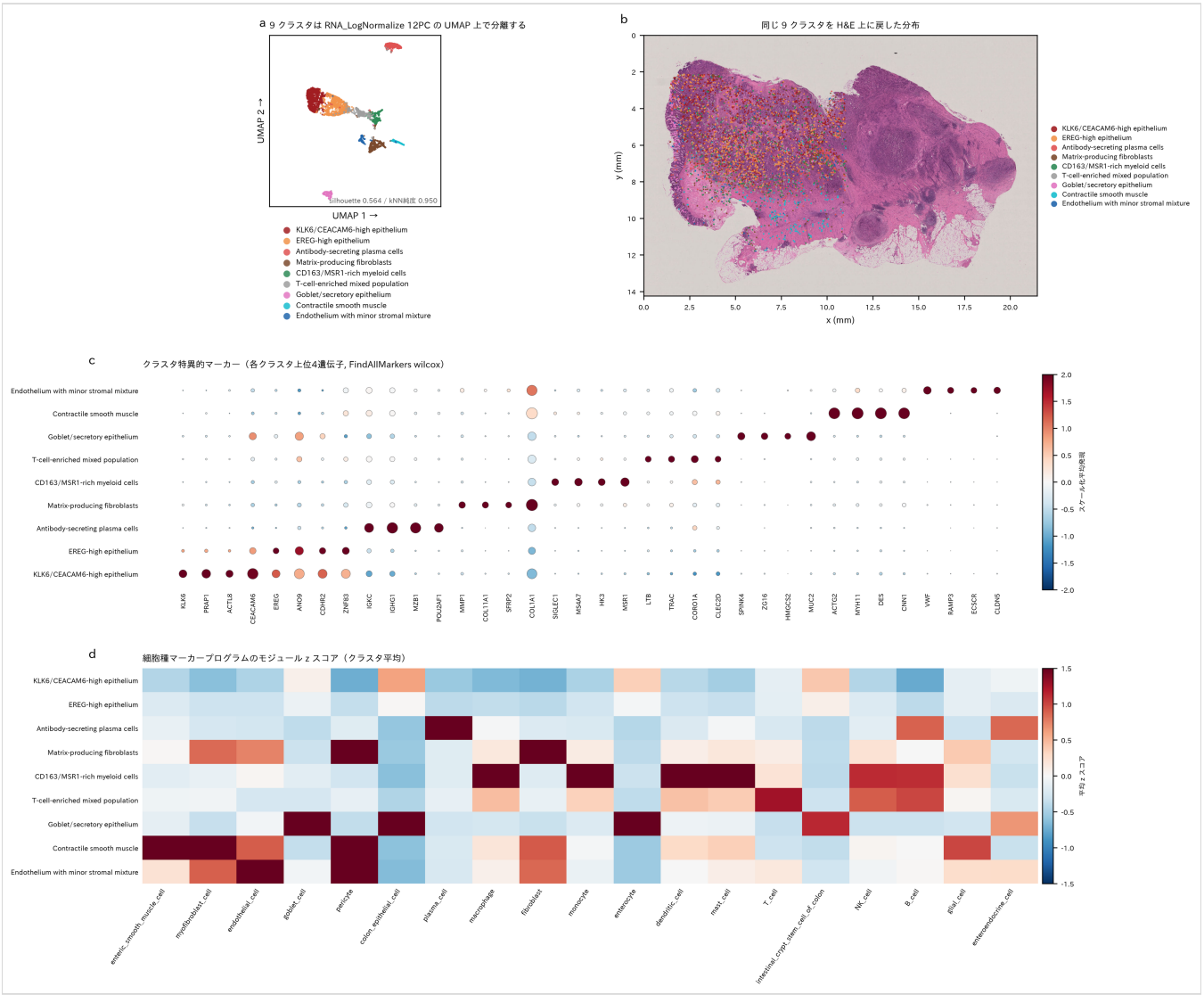


Fig. 8. マルチモーダル情報に基づく細胞アノテーション (a) 最終埋め込み。(b) H&E投射。(c) マーカー発現割合およびスケーリング後平均発現量。(d) 平均参照マーカースコア。

上皮細胞群の転写状態の比較

上皮サブクラスターによる分類に加えて、上皮細胞内の機能的な転写特性を評価するため、遺伝子セットスコアに基づく相対的な転写状態の分類を行った。全2,270上皮スポットのうち、Hypoxia_responseに1,042スポット、Cyclingに259スポット、EMT_likeに232スポット、Secretory_gobletに204スポット、Stem_cryptに147スポット、Absorptiveに78スポットが分類された（Table 8）。また、複数の状態スコアが同程度であった266スポットはmixed、明確な状態を示さなかった42スポットはunassignedとして分類した。これらの分類は、上皮細胞集団内における相対的な遺伝子発現状態を示すものであり、前節で定義した上皮サブクラスターとは独立した機能的状態として評価した。

各転写状態の空間分布を解析したところ、組織内で局所的な濃縮を示す領域が認められた（Fig. 9）。Hypoxia_responseが濃縮した代表領域（Fig. 9aの枠）では、上皮121スポット中72スポットがHypoxia_responseに分類されたのに対し、上皮全体に近い組成を示す比較領域では174スポット中82スポットであった。この結果から、Hypoxia_responseに関連する転写状態が組織内で一様に分布するのではなく、局所的に偏在する可能性が示唆された。

一方、同領域のH&E染色画像を目視で確認したところ（画像は本報告書に掲載していない）、顆粒状の好酸性物質はHypoxia_response濃縮領域よりも比較領域で顕著に認められた。このため、少なくとも本解析で評価した形態学的特徴については、Hypoxia_responseの局所的な濃縮を支持する明確な対応は認められなかった。

以上より、上皮細胞集団にはサブクラスター分類とは独立した機能的な転写状態の不均一性が存在し、Hypoxia_responseをはじめとする一部の状態は組織内で局所的な偏在を示した。一方、これらの転写状態とH&E染色画像上の形態学的特徴との対応は限定的であった。なお、状態間の差次的発現解析は状態分類に使用した発現データを用いているため、各状態の分子的特徴を記述するための解析として解釈した。

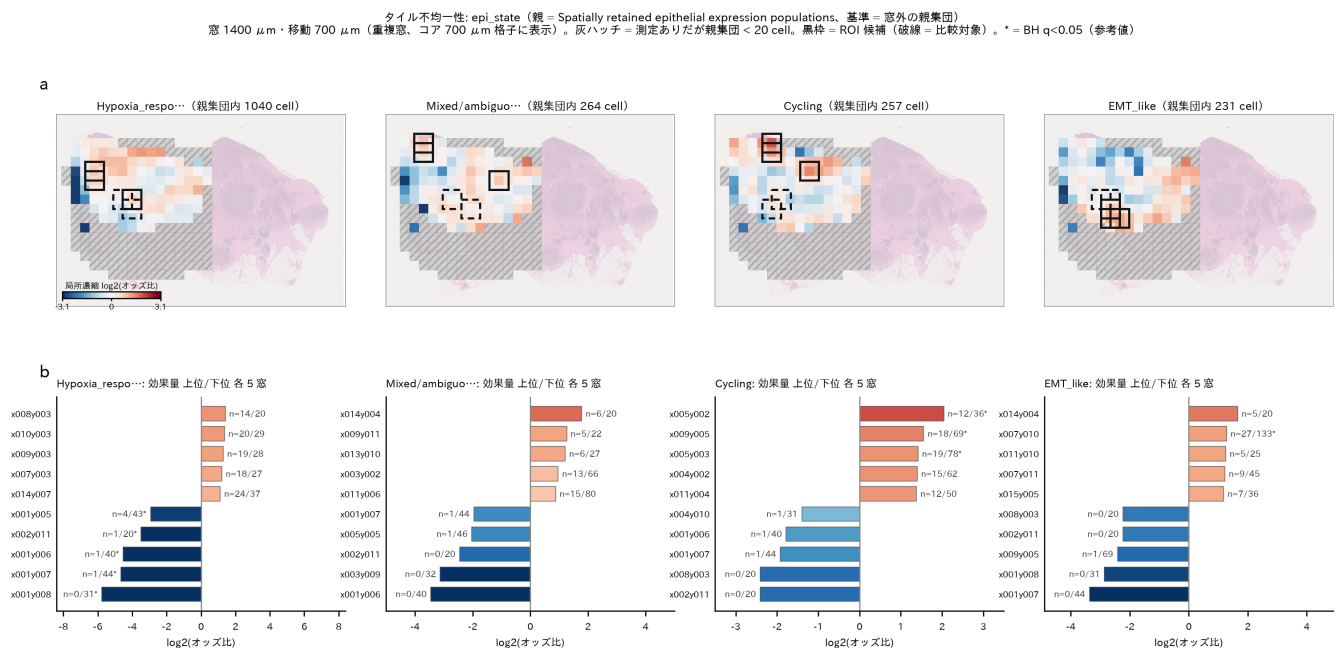


Fig. 9. 上皮細胞における転写状態の空間的不均一性。 (a) H&E染色画像上における主要な上皮転写状態の局所的な濃縮領域。(b) 各転写状態について効果量の上位・下位各5領域の濃縮および減少。各領域における対象状態のスポット数と上皮スポット総数を併せて示す。

上皮サブクラスターの局所的な空間的不均一性

遺伝子発現によって区別された上皮サブクラスターと局所的な組織形態との対応を評価するため、1,400 μm四方の解析領域を用いて各上皮サブクラスターの空間的エンリッチメントを解析し、代表的なエンリッチ領域と比較領域についてH&E染色画像を評価した。

Goblet/secretory epitheliumは、3つの上皮サブクラスターの中で特に明瞭な局所的濃縮を示した。代表的な濃縮領域では、Goblet/secretory epitheliumが上皮44スポット中41スポット（93.2%）を占め、領域外の上皮スポットを基準とした$\log_2$ odds ratioは6.84であった。一方、上皮全体に近い組成を示す比較領域では、115スポット中13スポット（11.3%）であり、$\log_2$ odds ratioは0.09であった（Tables 9–10）。Goblet/secretory epitheliumの濃縮領域は組織辺縁部に位置し、H&E染色画像では淡明な空胞を伴う陰窩様腺構造が認められた。これに対し、比較領域では不規則な腺管構造と細胞性間質の混在が認められた（Fig. 10a–b, e）。

KLK6/CEACAM6-high epitheliumおよびEREG-high epitheliumについても、それぞれ局所的に濃縮する領域が認められた。代表的な濃縮領域では、KLK6/CEACAM6-high epitheliumが上皮111スポット中80スポット、EREG-high epitheliumが45スポット中36スポットを占めた（Fig. 10c–d）。一方、両集団の濃縮領域では、H&E染色画像上で多様な腺管構造に加え、小型類円形核を有する細胞や間質成分の混在が認められ、Goblet/secretory epitheliumでみられたような明瞭な形態学的対応は認められなかった。各上皮サブクラスターの濃縮領域および比較領域について、H&E染色像、空間ラベル、および上皮細胞集団内の組成をFig. 10に示し、測定領域全体における局所的な濃縮および減少パターンをFig. 11に示した。

以上より、Goblet/secretory epitheliumでは、分泌型の転写特性と特徴的な腺形態との領域レベルでの対応が認められた。一方、KLK6/CEACAM6-high epitheliumとEREG-high epitheliumは、それぞれ異なる局所的な空間分布を示したものの（Fig. 11a）、H&E染色画像上では共通して多様な腺管構造を呈していた。これらの結果から、組織形態のみでは区別しにくい腺管領域においても、遺伝子発現に基づく分子的に異なる上皮サブクラスターが局所的に存在することが示された。

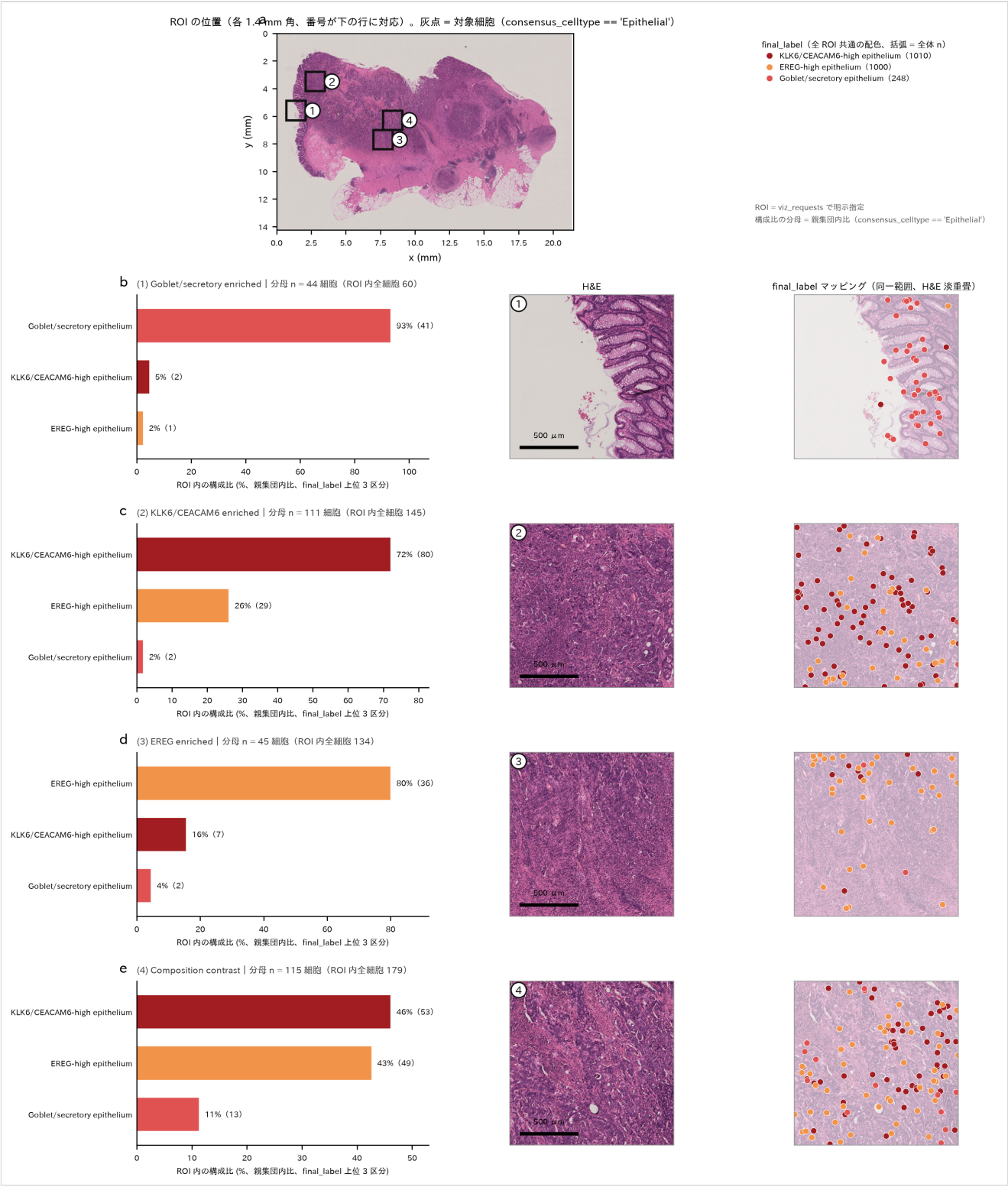


Fig. 10. 上皮サブクラスターの局所的enrichmentと組織形態との対応。 (a) H&E染色画像上における解析領域の位置。(b) Goblet/secretory epitheliumのenrichment領域。(c) KLK6/CEACAM6-high epitheliumのenrichment領域。(d) EREG-high epitheliumのenrichment領域。(e) 上皮サブクラスターの構成比が全体に近い比較領域。各領域について、上皮スポットに占める各サブクラスターの割合、H&E染色像、および各スポットの上皮サブクラスターを示す。

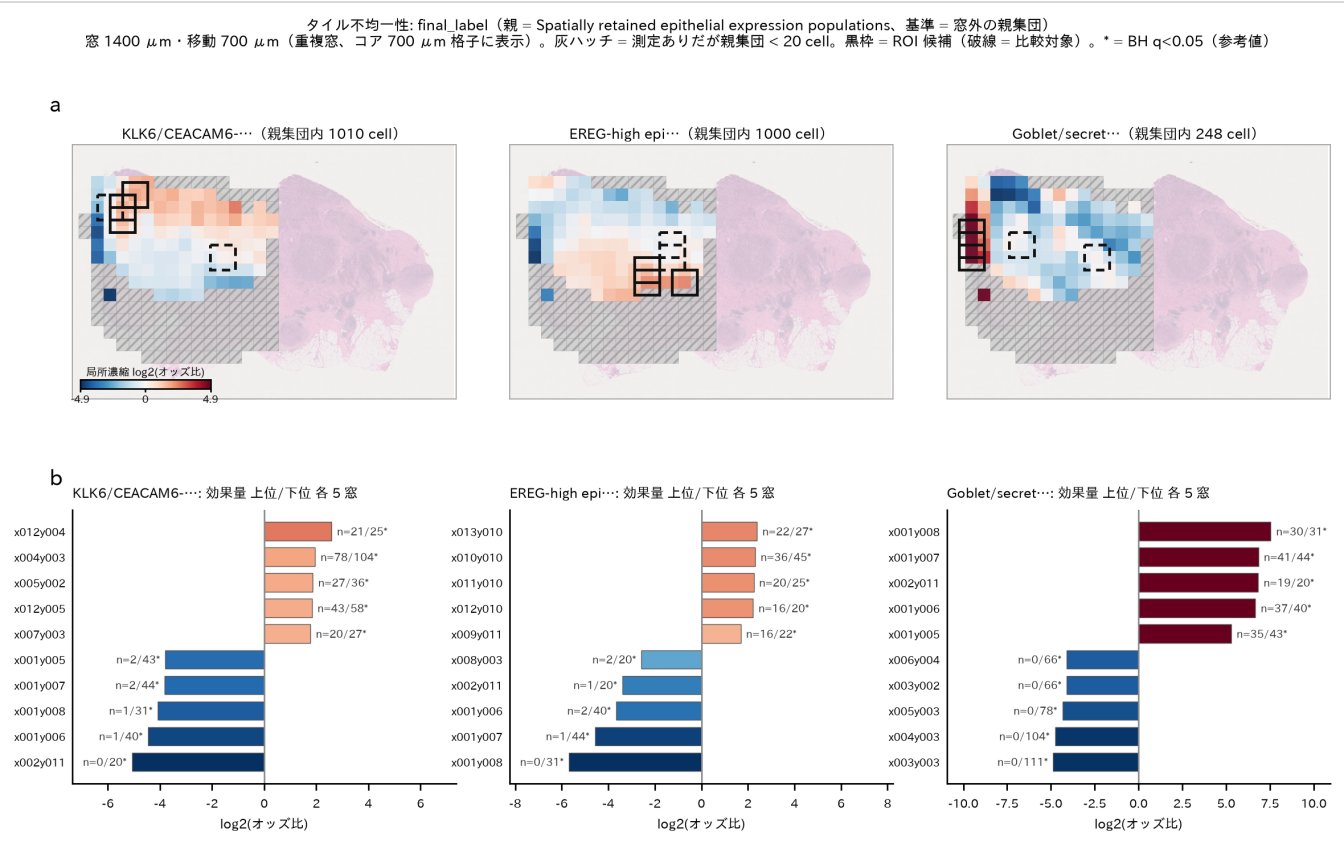


Fig. 11. 上皮サブクラスターの局所的な空間enrichmentパターン。 (a) H&E染色画像上における各上皮サブクラスターの局所的なenrichmentおよびdepletion領域。(b) 効果量の上位・下位各5解析領域における上皮サブクラスターのenrichmentおよびdepletion。対象サブクラスターのスポット数および上皮スポット総数を併記した。enrichmentスコアは、各解析領域の外側に存在する上皮スポットを比較対象として算出した。

免疫集団の空間分布と近傍組成

免疫微小環境の局所的な構成と組織形態との対応を評価するため、Antibody-secreting plasma cells、CD163/MSR1-rich myeloid cells、およびT-cell-enriched mixed populationを対象として、各集団の空間分布および局所的な濃縮を解析した。これら3つの免疫関連集団は測定領域内で異なる空間分布を示した（Fig. 13a）。さらに、半径250 μm以内のスポット組成に基づいて局所的な免疫環境を評価したところ、1近傍あたりのスポット数の中央値は17であった。近傍半径を150 μmおよび400 μmに変更した場合、250 μmで得られた近傍分類とのARIはそれぞれ0.386および0.410であり、局所的な免疫組成の分類が評価する空間スケールに依存することが示された。

次に、各免疫集団の濃縮領域と組織形態との対応を評価した（Fig. 12–13）。Antibody-secreting plasma cellsの代表的な濃縮領域では、免疫関連スポット103スポット中75スポットが同集団に分類され、領域外の免疫関連スポットを基準としたlog₂ odds ratioは1.87であった。CD163/MSR1-rich myeloid cellsの濃縮領域では35スポット中24スポット（log₂ odds ratio = 2.43）、T-cell-enriched mixed populationの濃縮領域では37スポット中24スポット（log₂ odds ratio = 2.50）がそれぞれ対象

集団に分類された。一方、共通の比較領域では、それぞれ45/95、25/95、24/95スポットであり（Fig. 12e）、$\log_2$ odds ratioは0.18、−0.22、−0.00であった。

H&E染色画像との対応を評価したところ、Antibody-secreting plasma cellsの濃縮領域は、細胞成分に富む紡錘形間質を特徴としていた。一方、CD163/MSR1-rich myeloid cellsおよびT-cell-enriched mixed populationの濃縮領域では、小型核の混在を伴う顆粒状の好酸性組織が認められた。比較領域においても小型核は豊富に認められたものの、免疫集団の組成は免疫関連スポット全体に近い構成を示した。各免疫集団の濃縮領域と比較領域について、H&E染色像、空間ラベル、および免疫関連集団内の組成をFig. 12に示し、測定領域全体における局所的な濃縮および減少パターンをFig. 13に示した。

以上より、Antibody-secreting plasma cells、CD163/MSR1-rich myeloid cells、およびT-cell-enriched mixed populationは、組織内でそれぞれ異なる局所的な濃縮パターンを形成していた。一方、小型核を伴う組織形態は複数の免疫集団の濃縮領域で共通して認められ、H&E染色像のみから各免疫集団を明確に識別することは困難であった。したがって、本検体における免疫微小環境の空間的不均一性は、H&E染色画像による組織形態と遺伝子発現に基づく細胞種アノテーションを統合して評価することが有用であることが示された。

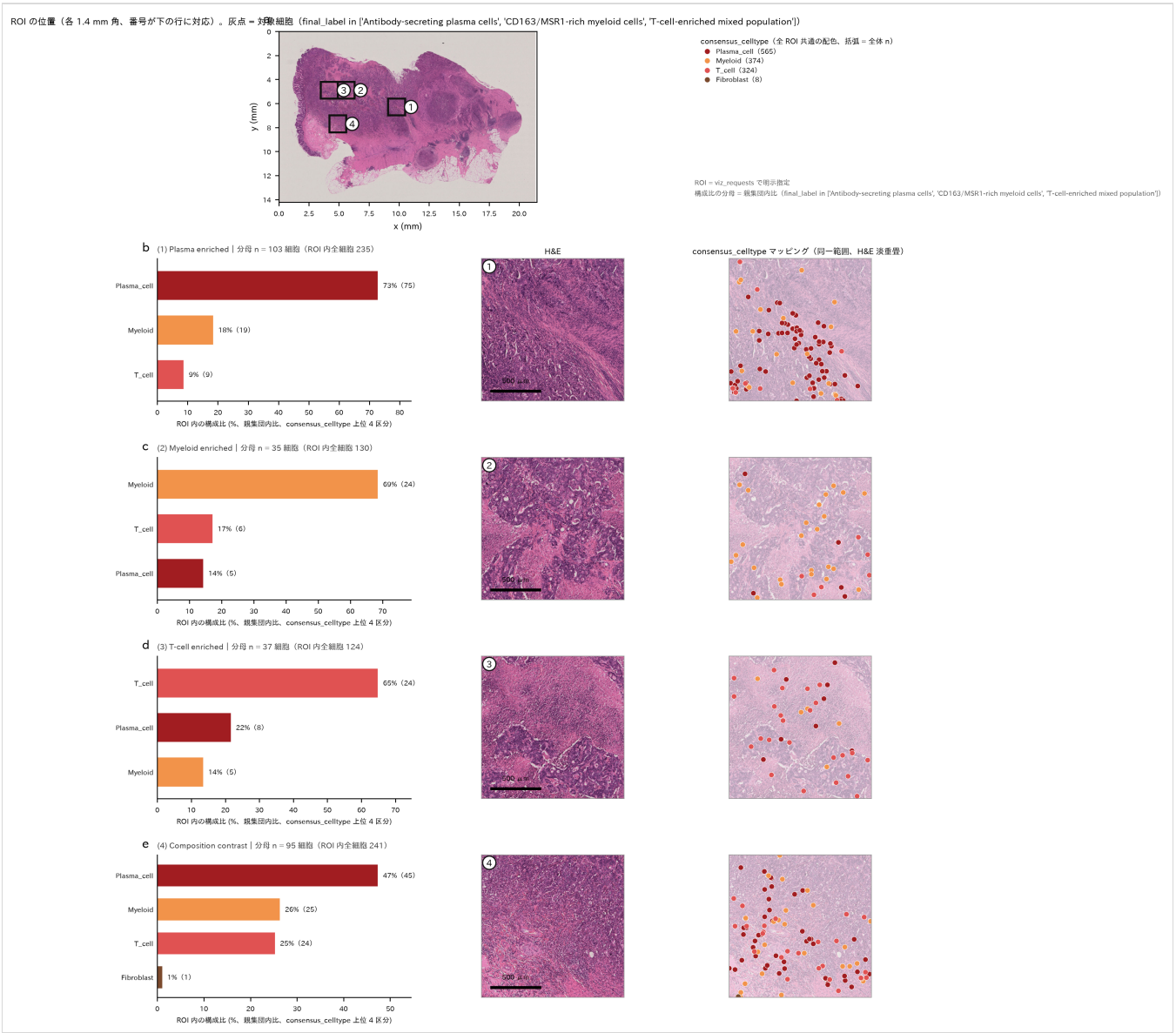


Fig. 12. 免疫細胞集団の局所的エンリッチメントと組織形態との対応。 (a) H&E染色画像上における解析領域の位置。(b) Antibody-secreting plasma cellsの濃縮領域。(c) CD163/MSR1-rich myeloid cellsの濃縮領域。(d) T-cell-enriched mixed populationの濃縮領域。(e) 免疫集団組成の比較領域。各領域について、免疫関連集団内の組成、H&E染色像、および空間ラベルを示す。図中の集団はコンセンサス細胞種（Plasma cell、Myeloid、T cell、Fibroblast）で集計した。

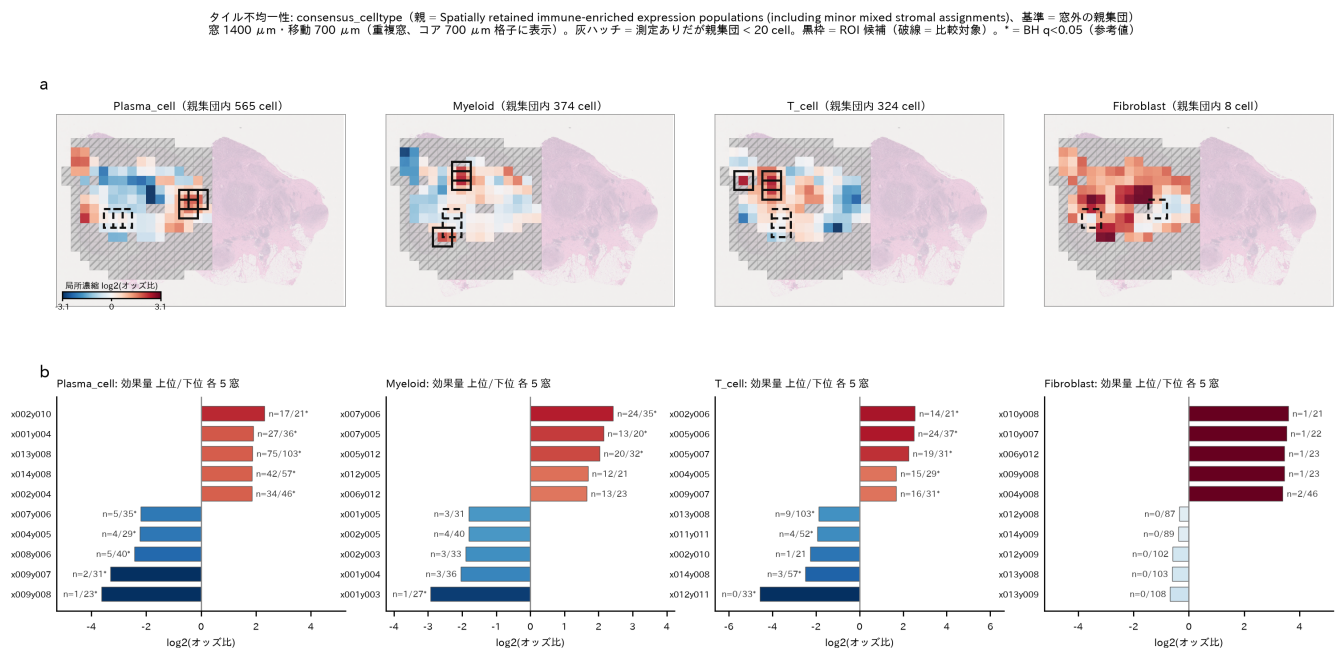


Fig. 13. 免疫細胞集団の局所的な空間濃縮パターン。 (a) H&E染色画像上における各免疫細胞集団の局所的な濃縮および減少領域。(b) 各解析領域における免疫細胞集団の濃縮・減少効果。各領域について、対象集団のスポット数および免疫関連スポット総数を併せて示す。図中の集団はコンセンサス細胞種で集計しており、Fibroblastは親集団内のスポット数が8と少ないため参考表示である。

上皮細胞への近接性と骨髄系細胞の転写プログラムの関連性解析

上皮細胞への近接性と免疫微小環境との関係を評価するため、各スポットから最も近い上皮スポットまでの距離を算出し、距離帯ごとの細胞集団の組成および骨髄系細胞の転写プログラムを解析した。主解析では、アノテーションされた全上皮スポットを距離算出の比較対象とし、各距離帯における細胞集団の構成を比較した（Fig. 14–15、Table 4）。さらに、比較対象を腫瘍性上皮の候補として位置付けたKLK6/CEACAM6-high epitheliumに限定した解析も行った（Fig. 16）。

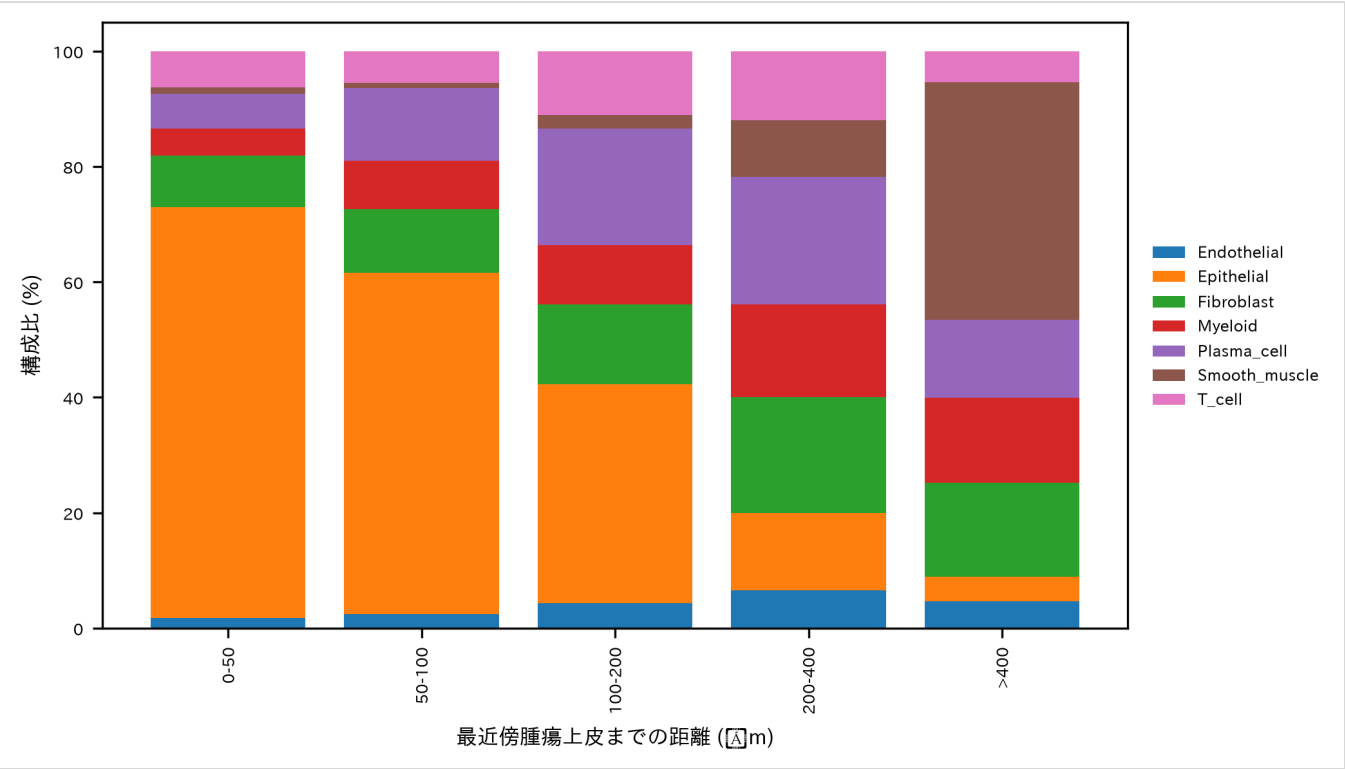


Fig. 14. 全上皮スポットからの距離に応じた細胞集団の構成。 各スポットから最も近い上皮スポットまでの距離に基づき、距離帯ごとの細胞集団の割合を示す。上皮スポット自身については、最も近接する別の上皮スポットまでの距離を用いた。したがって、比較対象となる上皮スポット自身の距離を0としたFig. 15とは、上皮スポットに対する距離の定義が異なる。

全上皮スポットからの距離を用いた解析では、CD163/MSR1-rich myeloid cellsにおけるmacrophage receptor-associated programのスコアと上皮からの距離との間に弱い正の相関が認められた（Spearmanのρ = 0.185、補正後p = 0.0057；Table 7、Fig. 15b）。同プログラムを構成する遺伝子は、骨髄系スポットの77.3%で少なくとも1遺伝子が検出された。さらに、スコアの算出方法による影響を評価した感度分析においても同方向の関連が認められ、生カウントの合計値ではρ = 0.223、補正後p = 0.00025、構成遺伝子が検出されたスポットのみに限定した解析ではρ = 0.228、未補正p = 0.00009であった（Table 12）。一方、interferon-response programのスコアと上皮からの距離には明確な関連は認められなかった（ρ = 0.002、補正後p = 0.976；Fig. 15c）。なお、Fig. 15b–cおよびFig. 16b–cに表示したq値は各図に示したプログラムの範囲で補正した値であり、補正の対象とした検定の範囲が異なるため本文およびTable 7の補正後p値とは一致しない。

比較対象をKLK6/CEACAM6-high epitheliumに限定した場合、距離帯ごとの細胞集団の構成には変化が認められたものの、macrophage receptor-associated programは上皮からの距離と正の関係を維持した（Fig. 16b）。一方、interferon-response programは距離に応じた明瞭な変化を示さなかった（Fig. 16c）。これらの解析では、距離帯ごとの細胞集団の構成に加え、比較対象とした上皮スポットの空間分布、各スポットから上皮までの距離、および骨髄系細胞における各転写プログラムの空間分布を可視化した（Fig. 15–16）。

以上より、CD163/MSR1-rich myeloid cellsにおけるmacrophage receptor-associated programは、上皮細胞との空間的位置関係に応じて変化することが示された。この関連は複数のスコア算出方法でも同方向に認められた一方、相関の強さ自体は小さかった。また、interferon-response programでは同様の距離依存性は認められず、骨髄系細胞と上皮との空間的関係が転写プログラムによって異なることが示された。

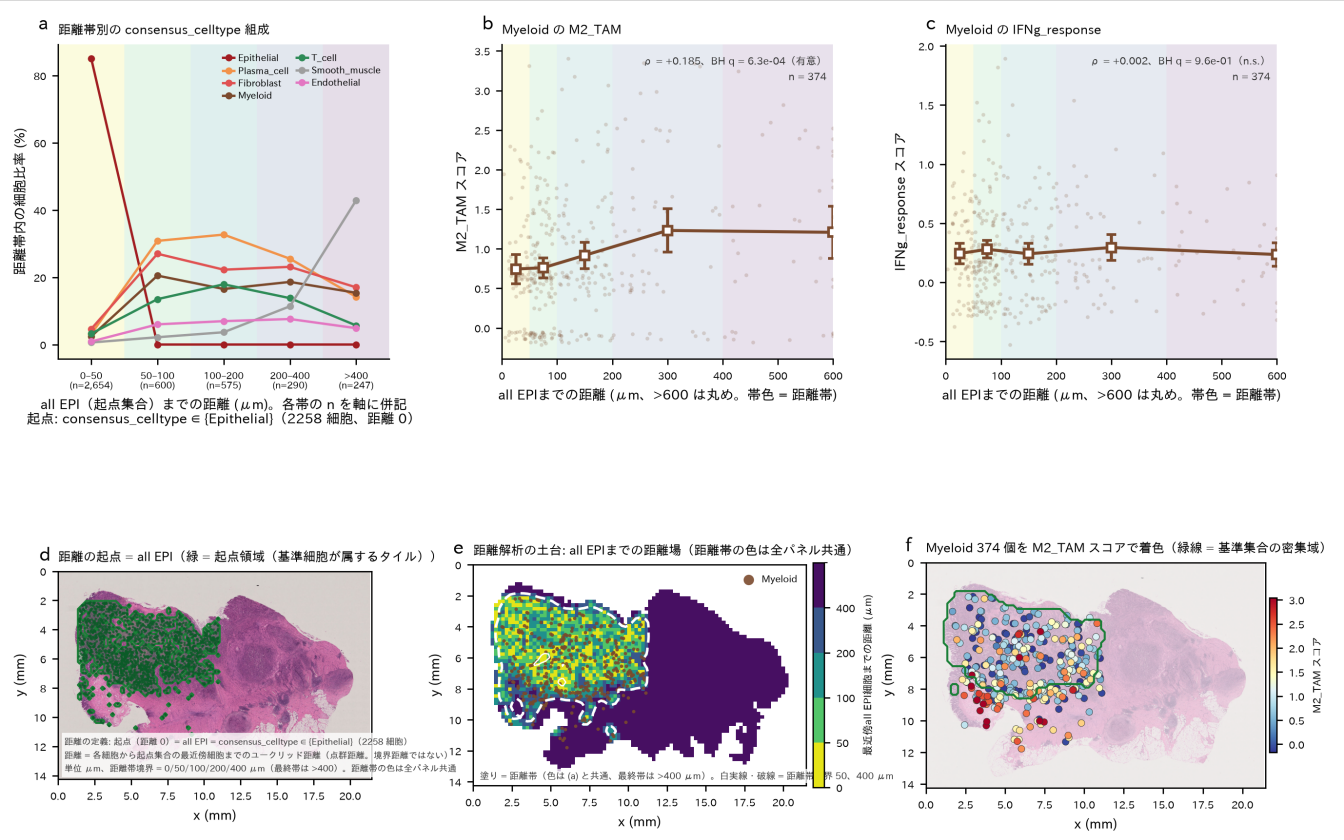


Fig. 15. 全上皮スポットからの距離に沿った細胞組成と骨髄系細胞の転写プログラム。 (a) 距離帯ごとの細胞集団の構成およびスポット数。(b) CD163/MSR1-rich myeloid cellsにおけるmacrophage receptor-associated programのスコアと上皮からの距離との関係。(c) Interferon-response programのスコアと上皮からの距離との関係。(d) H&E染色画像上における比較対象の上皮スポットの分布。(e) 各スポットから最も近い比較対象の上皮スポットまでの距離分布。(f) Macrophage receptor-associated programスコアの空間分布。比較対象となる上皮スポット自身には距離0を割り当てた。図中のM2_TAMはmacrophage receptor-associated program、IFNg_responseはinterferon-response programを示す。

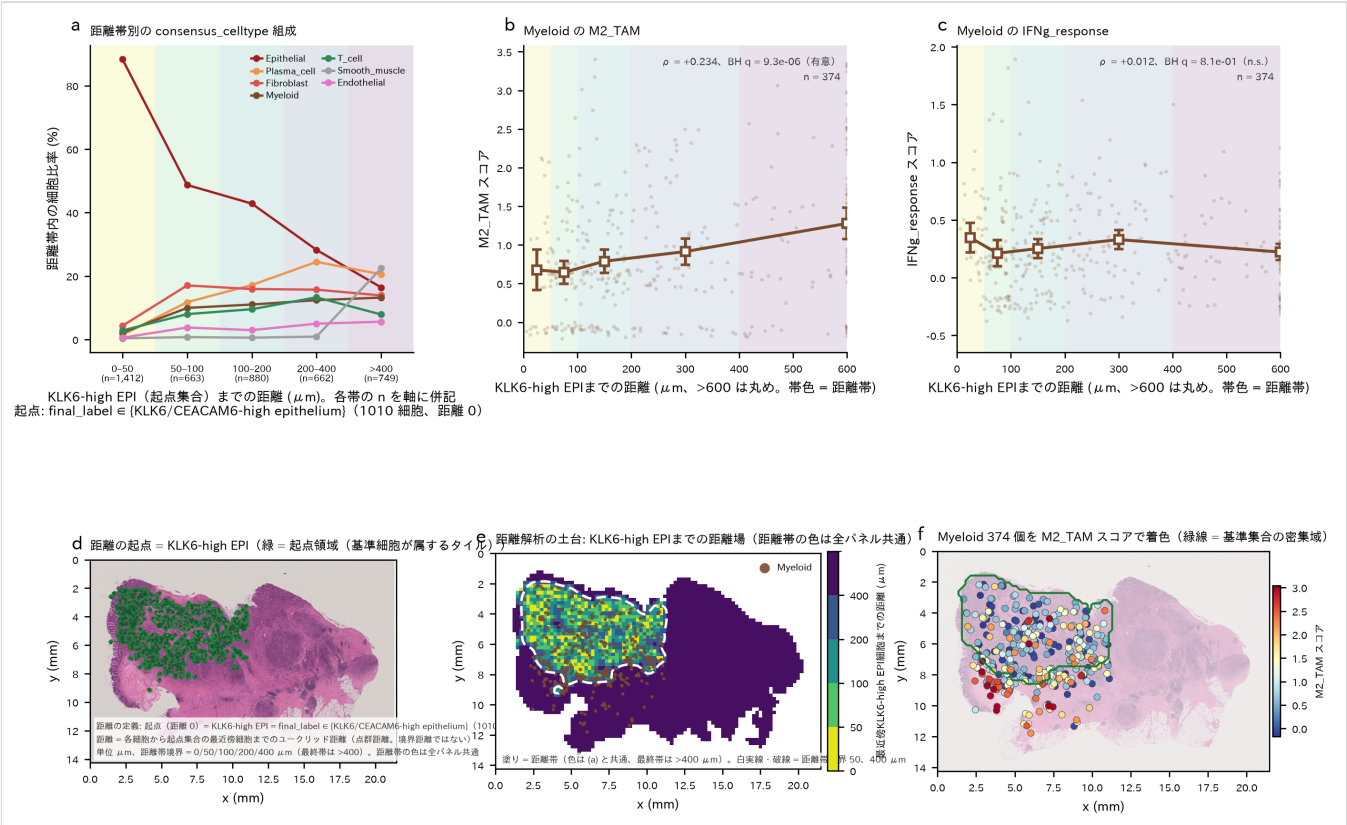


Fig. 16. KLK6/CEACAM6-high epitheliumからの距離に沿った細胞組成と骨髄系細胞の転写プログラム。 (a) 距離帯ごとの細胞集団の構成およびスポット数。(b) CD163/MSR1-rich myeloid cellsにおけるmacrophage receptor-associated programのスコアとKLK6/CEACAM6-high epitheliumからの距離との関係。(c) Interferon-response programのスコアと距離との関係。(d) H&E染色画像上におけるKLK6/CEACAM6-high epitheliumの分布。(e) 各スポットから最も近いKLK6/CEACAM6-high epitheliumまでの距離分布。(f) Macrophage receptor-associated programスコアの空間分布。KLK6/CEACAM6-high epitheliumに分類されたスポット自身には距離0を割り当てた。図中のM2_TAMはmacrophage receptor-associated program、IFNg_responseはinterferon-response programを示す。

Materials and Methods

検体および解析入力データ

本解析では、大腸がん検体由来の空間トランスクリプトームデータと、隣接切片のH&E全スライド画像を用いた。空間トランスクリプトームデータは4,387スポット、18,129遺伝子からなるraw count matrixおよびマイクロメートル単位の空間座標で構成され、H&E画像にはHamamatsu NDPIスキャンから変換されたOME-TIFFを使用した。H&E画像のピクセルサイズは0.441462 µmであった。

RNAデータについて、総カウント数、検出遺伝子数、およびミトコンドリアRNA比率を再計算した。検出遺伝子数200以上かつミトコンドリアRNA比率20%未満をQC基準とし、4,363スポットが基準を満たした。全スポットはQC情報を保持したまま解析対象とし、高信頼アノテーションや画像位置合わせに用いるスポットの選定にQC情報を使用した。

H&E染色画像の読み込み

H&E画像はOME-TIFFのピラミッド構造を利用し、解析対象領域をタイル単位で読み込んだ。画像メタデータからピクセルサイズおよび各ピラミッド階層の情報を取得し、スライド左上を原点、右方向をx軸正方向、下方向をy軸正方向として物理座標を定義した。組織領域の検出には低解像度画像を、形態特徴量の抽出にはより高解像度の画像階層を使用した。

クラスタリング、細胞種アノテーション、およびUMAPによる可視化

Raw countをSeurat [1] に読み込み、scale factor 10,000でLogNormalizeを行った後、最大2,000個の高変動遺伝子を選択し、スケーリングおよび主成分分析（PCA）を実施した。PCAで算出した主成分のうち、第1–16主成分を後続解析に使用し、累積寄与率は75.9%であった。これらの主成分に基づき、k = 20でshared nearest neighbor（SNN）グラフを構築し、resolution = 0.2、0.4、0.6、0.8、1.2でLouvainクラスタリングを実施した。クラスタリングの安定性はadjusted Rand index（ARI）、PCA空間のシルエット係数、および共クラスタリング頻度を用いて評価した。主要な細胞系統の分離とクラスタリングの安定性を総合的に評価し、resolution = 0.6を最終条件として採用した。 細胞種アノテーションには、公開大腸組織参照マーカーを用いたモジュールスコアリング、クラスター特異的differentially expressed genes（DEGs）の参照マーカー濃縮解析、CellTypistの腸管参照モデル、および免疫参照モデルを用いた [2]。各手法による判定を統合し、一致率80%以上、QC基準通過、および3手法以上で判定が一致したスポットを高信頼アノテーションとした。最終的に2,159スポットが高信頼基準を満たした。UMAP [3] はクラスタリング結果を固定したまま可視化条件のみを最適化した。距離指標、n_neighbors、min_distなどを変化させた75条件を比較し、シルエット係数、近傍ラベル純度、および

Davies–Bouldin indexを用いて評価した。最終的にmetric = euclidean、n_pcs = 12、n_neighbors = 50、min_dist = 0.01を採用した。

RNA発現に基づくCNV推定

RNA発現情報からcopy number variation（CNV）を推定するため、GRCh38対応のGENCODE release 47 [4] から遺伝子座標を取得し、infercnvpy 0.6.1 [5] を用いて解析した。遺伝子座標および発現量の基準を満たした遺伝子は性染色体の除外前で5,945個であり、性染色体を除外したうえで解析した。CNVシグナルは100遺伝子の移動窓、step size = 10、dynamic threshold = 1.5を用いて推定した。

高信頼かつQC基準を満たしたimmune、fibroblast、endothelial、およびsmooth muscle由来の1,905スポットを非上皮参照群とし、残る2,482スポットを評価対象とした。推定CNVプロファイルをPCAで次元圧縮した後、k = 15の近傍グラフとLeiden clustering（resolution = 1.0）を用いてCNVパターンを分類した。

各スポットのCNV scoreは推定CNVシグナルの絶対値の平均として算出した。CNVパターン群の判定には、CNV scoreの中央値が非上皮参照群の99パーセンタイルを上回ること、および少なくとも2染色体で平均CNVシグナルが経験的ノイズ閾値を上回ることを用いた。両条件を満たす群をCNV陽性、いずれか一方のみを満たす群をintermediateとして解析した。RNA由来CNVとH&E形態領域との関連はχ^2検定により探索的に評価した。RNA由来CNVはDNAレベルのCNVを直接測定するものではないため、腫瘍性や遺伝的サブクローンを確定する指標としては使用しなかった。

H&E画像に基づく形態領域分類およびROI形態計測

H&E全スライド画像から組織領域を抽出し、256 × 256 pixelのタイルに分割して形態特徴量を算出した。組織被覆率10%以上の3,140タイルを解析対象とした。特徴量にはHED色分離によるhematoxylinおよびeosin強度、空白領域率、テクスチャ特徴、およびwatershed法に基づく核オブジェクト特徴量を含めた [6]。

標準化した特徴量にPCAを適用し、累積寄与率95.58%を説明する10主成分を使用した。形態領域数は6–12の範囲で比較し、シルエット係数およびGaussian mixture modelのBayesian information criterion（BIC）を基に9領域を採用した。最終分類にはk-means clusteringを用い、空間的なラベルの孤立を抑えるため3 × 3近傍による多数決フィルターを適用した。各形態領域の名称は、代表的なH&E画像および画像特徴量に基づく記述的な形態学的名称として付与した。

形態領域間の組織学的特徴を定量化するため、各領域から最大8か所、計72か所の高解像度region of interest（ROI）を抽出した。各ROIは1,024 × 1,024 native pixelsとし、hematoxylin強度およびdistance-transform watershed法 [6] を用いて核オブジェクトを抽出した。面積8–400 μm²のオブジェクトを解析対象とし、核密度、面積、軸比、および円形度を算出した。小型円形オブジェクトは面積40 μm²以下かつ円形度0.85以上と定義した。これらの指標は自動セグメンテーションに基づく形態計測値として扱い、病理学的な細胞数そのものとは区別した。

空間トランスクリプトームとH&E画像の位置合わせ

空間トランスクリプトーム座標と隣接切片のH&E画像を対応付けるため、高信頼アノテーションが得られた2,159スポットを用いて位置合わせを行った。位置合わせは、H&E組織領域上に配置されるスポットの割合と、細胞種アノテーションおよびH&E形態領域間のnormalized mutual information（NMI）を組み合わせた目的関数を最大化することで最適化した。回転角度を2°刻みで探索するとともに、y軸反転の有無および平行移動を比較し、上位候補をPowell法で最適化した。主解析では物理スケールを固定した。

採用した座標変換を全スポットに適用した結果、4,366スポットが分類済みH&E組織領域上に位置した。フィッティングに用いたスポットの組織カバレッジは99.44%、NMIは0.1110であった。採用した変換では追加のy軸反転を行わなかった。

H&E画像と空間トランスクリプトームの統合解析

上皮および免疫細胞の転写プログラム 細胞種アノテーションとは独立して機能的な転写状態を評価するため、既報に基づいて選定した遺伝子セット [7], [8], [9] についてモジュールスコアを算出した。上皮2,270スポットについては、分化、cycling、hypoxia response、mesenchymal-like stateなどの遺伝子プログラムを評価した。各プログラムには少なくとも3つの検出可能な構成遺伝子を必要とし、最大スコアが0以下の場合をunassigned、最上位2プログラム間のスコア差が0.05未満の場合をmixedとした。状態間のDEG解析にはWilcoxon検定を用いたが、状態定義と同一の発現データを使用しているため、分子的特徴を記述する探索的解析として扱った。

細胞タイプごとの局所的な空間エンリッチメント解析 上皮サブクラスター、上皮転写状態、および免疫細胞集団の局所的不均一性は、1,400 µm四方の解析領域を700 µm間隔で移動させて評価した。各解析領域について、対象集団が親集団内に占める割合を算出し、領域内と領域外の構成比をFisherの正確確率検定および$\log_2$ odds ratioにより比較した。$\log_2$ odds ratioには0.5のcontinuity correctionを適用し、多重検定にはBenjamini–Hochberg法を用いた。

統計解析には親集団が20スポット以上含まれる領域を用い、濃縮領域の候補には対象集団が10スポット以上かつ正の濃縮を示すことを求めた。比較領域には$|\log_2 \text{odds ratio}| \leq 0.25$の領域を用いた。解析領域は互いに重複するため、p値は独立した反復検定としてではなく、切片内における探索的な空間エンリッチメントの指標として評価した。

上皮細胞への距離と骨髄系細胞の転写プログラムスコアの解析

各スポットから指定した上皮スポットまでの最近傍Euclidean distanceをマイクロメートル単位で算出し、距離帯ごとの細胞組成および骨髄系細胞の転写プログラムを評価した。主解析ではアノテーションされた全上皮スポットを比較対象とし、追加解析ではKLK6/CEACAM6-high epitheliumを比較対象とした。KLK6/CEACAM6-high epitheliumは発現マーカーおよびRNA由来CNVに基づく腫瘍性上皮候補として扱い、いずれの解析も病理学的に定義された腫瘍境界からの距離とは区別した。

骨髄系細胞におけるmacrophage receptor-associated program [10] およびinterferon-response program [9] と上皮からの距離との関連はSpearman相関を用いて評価した。感度分析として、モジュールスコア、生カウントによるプログラム構成遺伝子の合計値、およびプログラム構成遺伝子が検出されたスポットに限定した解析を比較した。多重検定の補正にはBenjamini–Hochberg法を用い、モジュールスコアと生カウント合計の検定群それぞれに個別に適用した。構成遺伝子が検出されたスポットに限定した感度解析のp値は補正していない。

また、局所的な細胞組成を評価するため、各スポットを中心とした半径250 μm以内の細胞組成を算出し、k-means clusteringにより代表的な近傍組成を分類した。空間スケールへの依存性を評価するため、150 μmおよび400 μmの半径でも同様の解析を行い、分類結果をARIで比較した。

ソフトウェアと再現性

トランスクリプトーム解析にはR 4.4.3、Seurat 5.5.1 [1] およびSeuratObject 5.4.0を使用した。Python解析にはScanpy 1.11.5、CellTypist 1.7.1 [2]、infercnvpy 0.6.1 [5]、scikit-image 0.26.0 [6]、scikit-learn 1.9.1、およびumap-learn 0.5.12 [3] を使用した。主要なクラスタリング、画像解析、転写状態解析、およびUMAP解析には固定乱数シードを設定し、同一条件下で解析結果を再現できるようにした。報告した数値は最終解析テーブルおよび実行メタデータと照合し、図表と解析結果の整合性を確認した。

References

1. Hao Y, Stuart T, Kowalski MH, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nature Biotechnology. 2024;42:293–304. doi:10.1038/s41587-023-01767-y.
2. Domínguez Conde C, Xu C, Jarvis LB, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science. 2022;376:eabl5197. doi:10.1126/science.abl5197.
3. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426. 一次プレプリント.
4. GENCODE Consortium. Human release 47, GRCh38.p14. 公式リリース記録.
5. Sturm G and the infercnvpy contributors. infercnvpy: inference of copy-number variation from single-cell RNA sequencing. Version 0.6.1. 公式推定ドキュメント.
6. van der Walt S, Schönberger JL, Nunez-Iglesias J, et al. scikit-image: image processing in Python. PeerJ. 2014;2:e453. doi:10.7717/peerj.453.
7. Wang Y, Song W, Wang J, et al. Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. Journal of Experimental Medicine. 2020;217:e20191130. doi:10.1084/jem.20191130.
8. Masuda K, Kornberg A, Miller J, et al. Multiplexed single-cell analysis reveals prognostic and nonprognostic T cell types in human colorectal cancer. JCI Insight. 2022;7:e154646. doi:10.1172/jci.insight.154646.
9. Liberzon A, Birger C, Thorvaldsdóttir H, et al. The Molecular Signatures Database hallmark gene set collection. Cell Systems. 2015;1:417–425. doi:10.1016/j.cels.2015.12.004.
10. Qi J, Sun H, Zhang Y, et al. Single-cell and spatial analysis reveal interaction of FAP+ fibroblasts and SPP1+ macrophages in colorectal cancer. Nature Communications. 2022;13:1742. doi:10.1038/s41467-022-29366-6.