

空間トランスクリプトミクスデータにおけるCellposeを用いた細胞セグメンテーションの改善

大阪大学大学院医学研究科医化学講座 松岡 研

目的	空間トランスクリプトミクス（Xenium, MERSCOPE）データに深層学習セグメンテーション Cellpose（CPSAMモデル）を適用し、従来手法と比較して検出細胞数・精度の向上を検証。スパコン SQUID の GPU を活用し、大規模データ（815GB）の実用的処理を評価した。
内容	MERSCOPE マウス心臓組織データ（815GB）を対象に、SOPA パイプライン+Cellpose v4.0.5 で解析。① sopa convert（zarr変換）→ ② patchify image（パッチ分割）→ ③ segmentation cellpose（DAPI, セル径35μm, GPU）→ ④ resolve + aggregate（境界競合解決・集約）。データ変換は dask 並列化により 68時間 → 約1時間へ短縮。
結果	最終的に 4,372,552個の細胞境界を検出（従来比で大幅増加）。境界競合 8,268,378件を約10分で解決。ただし、1細胞あたりの遺伝子発現カウントが不十分で、下流の1細胞レベル解析データとしては品質不足。データ取得条件・リード深度の最適化との組み合わせが今後の課題。

利用した計算機		その他	
SQUID GPUノード群（大阪大学CMC）		ソフトウェア	SOPA + Cellpose v4.0.5 (CPSAM)
ノード時間	GPU 1基: ~42.5h / CPU 16コア: ~1h	実行環境	Singularityコンテナ
使用メモリ	26.5 GB	言語	Python 3.10.12 / PyTorch 2.7.1+cu126
並列化	dask（16コア並列）	データ形式	zarr（SpatialData準拠）
総処理時間	約46時間		