Qeios

Peer Review

Review of: "Non-Negative Blind Source Separation and the Identifiability of Cell-Type Deconvolution in Spatial Transcriptomics: A Mathematical Review"

Asier Antoranz[1]

1. Imaging and Pathology, KU Leuven, Leuven, Belgium

Walinjkar presents a very good review that brings a useful mathematical perspective to cell-type deconvolution in spatial transcriptomics. Its main contribution is to place reference completeness, signature conditioning, and source identifiability at the centre of the discussion. The distinction between structural identifiability and statistical estimation is particularly valuable, as is the recognition that some populations may be recoverable while others remain confounded.

My main concern is the connection between these mathematical conditions and the construction or estimation of the signatures used in practice. In the supervised setting, the analysis assumes a fixed reference matrix, but its representativeness depends on gene selection, annotation resolution, within-population variability, and measurement-platform differences. In semi-supervised and unsupervised settings, the signatures are partly or entirely inferred, but the question remains whether one signature per component adequately represents the biological variation present. Developing these connections would strengthen the review and its practical recommendations.

1. Representativeness of reference signatures

The manuscript discusses missing cell types in Section 3.1 and similarity between cell-type signatures in Section 3.2. However, these are different from heterogeneity within a reference population. A cell-type signature is usually an average across cells that may include several states or subclusters. If their relative abundance differs between the reference and the query, the average signature can change even when the broad cell-type annotation remains appropriate.

doi.org/10.32388/X175S8

The consequences are important. A reference can include all relevant cell types and have well-separated average signatures, yet still produce biased fractions because those averages do not represent the query tissue. Count noise around a fixed mean does not fully describe this situation.

I suggest adding a discussion of uncertainty and variation in the signature matrix itself. This should distinguish variability around a representative mean from systematic changes in that mean across donors, conditions, or cell states. Existing approaches provide useful examples: BLADE (Andrade Barbosa, B., et al., 2021) explicitly models variable cell-type expression, while DestVI (Lopez, R., et al., 2022) models continuous within-cell-type states in spatial data.

2. Reference resolution and reporting resolution

Section 3.4 already recognises that broad lineages can be identifiable while related subtypes remain confounded. Section 4 also mentions relabelling to improve conditioning. This discussion could be extended by distinguishing the resolution used to construct the reference from the resolution of the final output.

In single-cell analysis, expression-distinct clusters are often annotated and subsequently merged into common cell types. Averaging those clusters before deconvolution fixes their internal composition according to the reference. Retaining them as separate components allows their contributions to vary in the query, while their estimated fractions can still be summed into the same cell-type annotation afterward.

This does not require every cluster fraction to be individually identifiable. Different solutions may exchange abundance between clusters of the same cell type while preserving that cell type's total.

I recommend discussing this distinction and assessing identifiability at the level of the intended output. Redeconve (Zhou, Z., et al., 2023), already cited in the manuscript, is relevant to the use of finer reference representations. MuSiC (Wang, X., et al., 2019) provides an example of hierarchical estimation with stage-specific gene selection. Neither approach removes reference assumptions, but both show that reference construction and estimation resolution deserve more attention than a simple choice between fine and coarse labels.

3. Gene selection and the interpretation of conditioning

Section 3.3 discusses anchor genes and acknowledges that strict specificity is uncommon. However, the practical discussion does not sufficiently connect gene selection to the proposed conditioning diagnostic.

A gene can be significantly differentially expressed without providing reliable discrimination in a new mixture. Its usefulness depends on the expression contrast relative to within-population variability, consistency across donors, measurement noise, and platform compatibility. The panel must also

distinguish the relevant populations jointly. Selecting many genes that follow the same variable programme may give that programme disproportionate influence in the objective.

I suggest explaining that the condition number must be evaluated on the actual gene set and scaling used for inference. It should also be clear that favourable conditioning of average signatures does not establish that those signatures are stable. MuSiC's use of cross-subject consistency is a relevant example of an existing method addressing this issue.

4. Measurement scale, platform differences, and coefficient interpretation

The manuscript briefly acknowledges batch drift and capture limitations, but these are not sufficiently integrated into the main model or the practical recommendations.

Reference and query datasets often come from different technologies. Differences in capture efficiency, sequencing depth, and gene-specific sensitivity can change measured expression without changing cell composition. These effects can bias fractions even when the reference is complete and well-conditioned.

I recommend making the measurement assumptions explicit: whether the model uses raw counts or normalised linear-scale expression, how spot-level depth is handled, and how reference and query signatures are brought onto compatible scales. RCTD (Cable, D.M., et al., 2022) is particularly relevant because it explicitly models platform effects and incorporates total UMI counts. These features are substantive parts of the method, rather than minor differences around a common regression objective.

5. Clarification of the mathematical statements

Several statements would benefit from more precise wording.

Reference completeness and rank. Section 3.1 correctly identifies missing populations as a source of error, but completeness is not equivalent to rank. An incomplete reference can still have independent signatures. An omitted population may also be represented by a combination of included signatures, so its absence need not produce an obvious residual. Section 4 should therefore present residual analysis as a useful diagnostic, rather than a sufficient check of completeness.

Identifiability versus instability. Section 3.5 makes this distinction clearly, but some other passages describe similar signatures as structurally non-identifiable. Nearly collinear signatures can still give a unique noiseless solution while being highly unstable under realistic noise. Maintaining this distinction throughout would improve the argument.

Declarations

Potential competing interests: No potential competing interests to declare.