



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *25th International Conference on Pattern Recognition (ICPR)*.

Citation for the original published paper:

**Partel, G., Wählby, C. (2020)**

**Graph-based image decoding for multiplexed in situ RNA detection**

**In:**

**N.B. When citing this work, cite the original published paper.**

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-419165>

# Graph-based image decoding for multiplexed in situ RNA detection

Gabriele Partel  
SciLifeLab, and  
Dept. of Information Technology  
Uppsala University  
Uppsala, Sweden  
Email: gabriele.partel@it.uu.se

Carolina Wählby  
SciLifeLab, and  
Dept. of Information Technology  
Uppsala University  
Uppsala, Sweden  
Email: carolina.wahlby@it.uu.se

**Abstract**—Image-based multiplexed in situ RNA detection makes it possible to map the spatial gene expression of hundreds to thousands of genes in parallel, and thus discern at the same time a large numbers of different cell types to better understand tissue development, heterogeneity, and disease. Fluorescent signals are detected over multiple fluorescent channels and imaging rounds and decoded in order to identify RNA molecules in their morphological context. Here we present a graph-based decoding approach that models the decoding process as a network flow problem jointly optimizing observation likelihoods and distances of signal detections, thus achieving robustness with respect to noise and spatial jitter of the fluorescent signals. We evaluated our method on synthetic data generated at different experimental conditions, and on real data of in situ RNA sequencing, comparing results with respect to alternative and gold standard image decoding pipelines.

## I. INTRODUCTION

Recent technological advances in molecular profiling of RNA transcripts in situ paved the way to unprecedented exploration of cells and tissues, uncovering the gene expression heterogeneity and preserving the spatial morphological context of targeted molecules. The advent of these new technologies allowed a variety of novel studies and applications comprising unsupervised molecular classification of tissue compartments [1], generation of spatial atlases of tissues and entire organs [2]–[4], and the discovery of spatio-temporal regulation of gene expression networks in relation to development and disease [5], [6]. The majority of these technologies rely on different image-based RNA sequencing techniques [3], [4], [7], [8] where RNA transcripts are targeted with fluorescent probes and imaged in multiple channels and rounds. All of these techniques produce a large amount of conceptually similar image datasets that need to be processed and analyzed in order to localize and map the gene expression into the tissue samples. Image processing of multiplexed in situ RNA detection data consists of decoding fluorescent signals across fluorescent channels and imaging rounds, in order to identify what gene transcript is present in a given spatial position. Each targeted gene is typically identified by a unique sequence of fluorescent signals called barcode (Figure 1).

The decoding process presents several challenges due to image quality, transcript density and sequencing artifacts.

Signals typically appear as diffraction limited fluorescent spots of 3-5 pixels diameter, whose detection efficiency depends on image signal-to-noise ratio and tissue autofluorescence. Transcript density is another limiting factor for image decoding, restraining the ability to optically discern overlapping fluorescent molecules and track them between sequencing rounds. At the same time, high signal density and large area coverage is desirable as the purpose of the technique is to explore spatial heterogeneity with high multiplexity. Moreover, sequences of signals could be affected by small spatial jitter that shifts their locations by a few pixels between imaging rounds, and thus complicates the decoding process.

Here we present a graph-based approach for image decoding of multiplexed in situ RNA detection that globally optimizes signal candidate associations. In the proposed approach, the image decoding problem is modeled as a maximum a posterior probability (MAP) estimation problem. Where, given a set of signal candidate detections of fluorescent signals across sequencing channels and rounds (Figure 1), non-overlapping signal traces of candidate detections are modeled as disjoint flow paths in a cost-flow network where flow costs represent observation likelihoods and transition probabilities. Solving the cost-flow network with the maximum flow of minimum cost algorithm finds the globally optimal set of non-overlapping signal traces that represents decoded barcodes of targeted RNA molecules. We compare and validate our method on synthetic data generated at different noise and jitter conditions, with respect to two alternative and commonly adopted decoding strategies presented in [4], [9]. We also validate our approach on real data from in situ RNA sequencing [7], comparing decoded results with respect to the assay’ gold standard image decoding pipeline [9].

## II. ALTERNATIVE IMAGE DECODING STRATEGIES

A variety of approaches have been developed to decode multiplexed in situ RNA detection data and a team effort by CZI (Chan Zuckerberg Initiative) engineers and scientists from the SpaceTx consortium collected them in an open source framework called starFISH (<https://github.com/spacetx/starfish>).

A typical workflow for decoding barcodes from multiplexed in situ RNA detection images consists in a first alignment

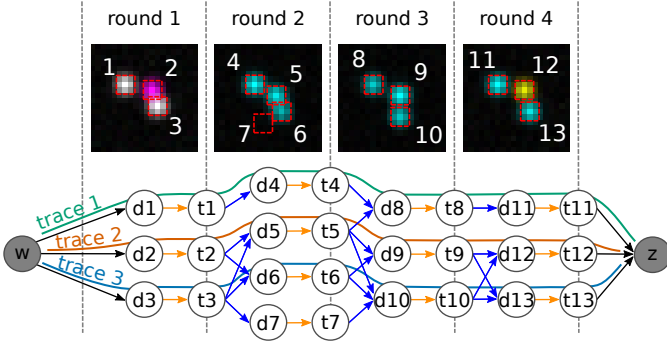


Fig. 1. Graph-based decoding of multiplexed in situ RNA detections. Upper panel shows four imaging rounds in four fluorescent channels representing bases of the genetic code (i.e. A, C, G, T), and displayed as composite images with gray (i.e. A), magenta (i.e. C), cyan (i.e. G), yellow (i.e. T) pseudo-colors. Signal detections in each imaging round are numerated and highlighted with red bounding boxes. Lower panel shows a relative cost-flow network of the given signal candidate detections, with decoded signal traces that represent three decoded barcodes of RNA molecules expressed by a total of two different genes. Specifically,  $\tau_1 : \{s_1, s_4, s_8, s_{11}\}$  (i.e. barcode AGGG),  $\tau_2 : \{s_2, s_5, s_9, s_{12}\}$  (i.e. barcode CCGT),  $\tau_3 : \{s_3, s_6, s_{10}, s_{13}\}$  (i.e. barcode AGGG). Note that signal candidate detection  $s_7$  is a false detection with low quality, thus not selected to be part of any trace during the optimization process.

step of imaging rounds and fluorescent channels to a common reference, followed by a preprocessing step for background attenuation and spot enhancement. Afterwards, two possible decoding strategies, based on previously published image analysis pipelines [4], [9] have been re-implemented in starFISH. The first approach (exact match, EM) initially detects fluorescent signals on a reference image (e.g. a reference fluorescent stain channel of all target barcodes, or an artificial reference image produced by the maximum intensity projection across all four fluorescent channels of each imaging round). Successively, fluorescent intensities are extracted at the exact same locations from the single fluorescent channels of each imaging round. Thereafter, barcode sequences of RNA molecules are decoded by selecting the fluorescent channel with highest intensity from each imaging round. The second approach (nearest neighbor search, NNS) instead consists of detecting fluorescent signals in all fluorescent channels and imaging rounds, and a complementary matching strategy based on a nearest neighbor search within a maximum distance for association of signal candidate detections across rounds.

### III. GRAPH-BASED DECODING

We propose a graph-based decoding approach based on global data association for multi-object tracking using network flow [10] for matching signal candidate detections nominated across fluorescent channels and sequencing rounds. Given an input set of fluorescent signal detections, we define barcode decoding as a MAP estimation problem. Where, the optimal set of non-overlapping signal traces, representing decoded barcode sequences of RNA molecules, is defined as disjoint flow paths in a cost-flow network. Signal observation likelihood (i.e. signal quality) and matching probabilities are modeled as

flow costs. The decoding is then solved by the maximum flow minimum cost algorithm.

#### A. MAP formulation

Given a set of signal candidate detections  $\chi = \{s_i\}$ , such that each candidate detection is defined as  $s_i = (x_i, c_i, r_i, I_i, q_i)$ , where  $x_i$  is its location in the image,  $c_i$  and  $r_i$  are respectively the fluorescent channel and imaging round it belongs to,  $I_i$  is the intensity value of the fluorescent signal, and  $q_i$  is an associated quality to the signal detection. A possible signal trace is defined as an ordered list of signal candidate detections, i.e.  $\tau_k = \{s_{k_1}, s_{k_2}, \dots, s_{k_n}\}$ , where  $s_{k_r} \in \chi : r \in \{1, 2, \dots, n\}$ , and  $n$  is the number of imaging rounds. And decoded barcodes are defined as a set of non-overlapping signal traces, i.e.  $T = \{\tau_k\}$ , such that  $\forall \tau_k, \tau_l \in T : \tau_k \cap \tau_l = \emptyset$ , where  $k \neq l$ .

Thus, the barcode decoding problem is solved by maximizing the posterior probability of  $T$  given the input set of signal candidate detections  $\chi$ . Assuming conditional independence of likelihood probabilities:

$$\hat{T} = \underset{T}{\operatorname{argmax}} P(T|\chi) \quad (1)$$

$$= \underset{T}{\operatorname{argmax}} P(\chi|T)P(T) \quad (2)$$

$$= \underset{T}{\operatorname{argmax}} \prod_i P(s_i|T) \prod_{\tau_k \in T} P(\tau_k). \quad (3)$$

Where  $P(s_i|T) = q_i$  is the observation likelihood of signal detection  $s_i$ , defined as its quality. And,

$$\begin{aligned} P(\tau_k) &= P(\{s_{k_1}, s_{k_2}, \dots, s_{k_n}\}) \\ &= P(s_{k_2}|s_{k_1})P(s_{k_3}|s_{k_2}) \dots P(s_{k_n}|s_{k_{n-1}}), \end{aligned}$$

where  $P(s_{k_{i+1}}|s_{k_i})$  is the matching probability of signals candidate detections  $s_{k_{i+1}}, s_{k_i} \in \tau_k$ .

#### B. Cost-flow network formulation

The respective cost-flow network formulation of the problem consists of a graph  $G(\chi)$  with source  $w$  and sink  $z$  (Figure 1). Where signal candidate detections  $s_i \in \chi$  are represented in the graph as detection nodes  $d_i$ . Each  $d_i$  node is connected to a transition node  $t_i$  by an edge (orange arrows in Figure 1) with cost  $\epsilon_i$  equal to:

$$\epsilon_i = -\log P(s_i|T) = -\log q_i.$$

Possible matching hypotheses among signal candidate detections belonging to different imaging rounds are modeled as edges connecting  $t_i$  nodes of a given round to  $d_j$  nodes of a different round (blue edges in Figure 1). Each edge connecting a pair of  $(t_i, d_j)$  nodes has a cost that depends on the Euclidean distance between the signal candidate detections  $s_i, s_j \in \chi$ , specifically:

$$\epsilon_{ij} = -\log P(s_j|s_i) = -\log \left( \frac{1}{1 + \delta d_{ij}} \right),$$

where  $d_{ij}$  is the euclidean distance between  $s_i$  and  $s_j$ , and  $\delta$  is a parameter that weights the contribution of  $d_{ij}$ . Finally,

signal candidate detections belonging to the first and the last imaging rounds are connected through edges with zero costs respectively to the source  $w$  and the sink  $z$  (black edges in Figure 1). In order to enforce the non-overlapping constraint for signal traces we add the following flows to the graph edges:

$$f(w, d_i) = f_{w,i}, \quad (4)$$

$$f(t_i, z) = f_{i,z}, \quad (5)$$

$$f(t_i, d_j) = f_{i,j}, \quad (6)$$

$$f(d_i, t_i) = f_i, \quad (7)$$

with,

$$f_{w,i} = \begin{cases} 1 & \exists \tau_k \in T, \text{ s.t. } \tau_k \text{ start from } s_i \\ 0 & \text{otherwise} \end{cases}, \quad (8)$$

$$f_{i,z} = \begin{cases} 1 & \exists \tau_k \in T, \text{ s.t. } \tau_k \text{ ends at } s_i \\ 0 & \text{otherwise} \end{cases}, \quad (9)$$

$$f_{i,j} = \begin{cases} 1 & \exists \tau_k \in T, \text{ s.t. } s_j \text{ is right after } s_i \text{ in } \tau_k \\ 0 & \text{otherwise} \end{cases}, \quad (10)$$

$$f_i = \begin{cases} 1 & \exists \tau_k \in T, \text{ s.t. } s_i \in \tau_k \\ 0 & \text{otherwise} \end{cases}. \quad (11)$$

Such that,  $T$  contains non-overlapping signal traces if and only if:

$$f_{w,i} + \sum_j f_{i,j} = f_i = f_{i,z} + \sum_j f_{i,j}, \forall s_i \in \chi. \quad (12)$$

Thanks to equation (12) all the nodes in the graph respect the flow conservation property that allows to find a globally optimal solution of the following objective function through max flow min cost algorithm implementation:

$$\hat{T} = \underset{T}{\operatorname{argmin}} \sum_i -\log P(s_i|T) + \sum_{\tau_k \in T} -\log P(\tau_k) \quad (13)$$

$$= \underset{T}{\operatorname{argmin}} \sum_i \epsilon_i f_i + \sum_{i,j} \epsilon_{i,j} f_{i,j}. \quad (14)$$

### C. Practical implementation

An initial graph is built from a given set of signal candidate detections by searching for connected components among imaging rounds with a maximum connection distance  $d_{th}$  between candidate detection pairs. Successively, the graph is refined by adding connections between nodes belonging to consecutive imaging rounds and within the same connected component that are closer than a maximum distance  $d_{max}$ . The graph is next pruned of all connections between signal candidate detections not belonging to consecutive imaging rounds. Finally, signal candidate detections belonging to the first and last imaging rounds of each connected component are connected respectively to a source and a sink, and the cost-flow networks derived from the graph are solved by iterating a max flow min cost algorithm between sink and source of each connected component.

### D. Quality of decoded sequences

Each decoded signal trace  $\tau_k \in T$  is supplemented with a quality score  $Q_k$  used to assess the reliability of decoded barcodes during post processing analyses. The quality score  $Q_k$  encodes quality information  $q_{k_r}$  and pairwise distances from each signal candidate detection  $s_{k_r}$  of the decoded signal trace  $\tau_k$ , such that:

$$Q_k = \mu_k \cdot \sum_{r=1}^n q_{k_r}, \quad (15)$$

where  $\mu_k$  is a function that depends on the maximum distance between signal candidate detections  $\{s_{k_1}, s_{k_2}, \dots, s_{k_n}\}$  of the decoded signal trace  $\tau_k$ . The function  $\mu_k$  penalizes signal traces composed of signal candidate detections far apart from each other, and is defined as:

$$\mu_k = 1 - \frac{\log(1 + d_{ij_{\max}})}{\sigma}, \quad (16)$$

and clipped to the interval  $[0, 1]$ , where  $d_{ij_{\max}}$  denotes the maximum distance between pairs of signal candidate detections  $\{s_{k_1}, s_{k_2}, \dots, s_{k_n}\}$  of the decoded signal trace  $\tau_k$ , and  $\sigma$  is a parameter weighting the penalty.

We propose two alternative options for defining the quality  $q_i$  of each signal candidate detection  $s_i \in \chi$ . In the first case we use a signal/noise classifier (e.g as a convolutional neural network) trained on a subset of signal candidate detections to predict for each candidate the probability  $p_i$  of being signal or noise, and we define the signal quality as:

$$q_i = p_i.$$

Then, if multiple candidates  $\{s_l\} \in \chi$  have been detected in the same location of a candidate detection  $s_{k_r} \in \tau_k$ , such that:

$$(r_{k_r} = r_l) \wedge (x_{k_r} = x_l) \wedge (c_{k_r} \neq c_l) \wedge (p_l > 0.5),$$

their intensities and probability prediction are combined, such as:

$$q_{k_r} = \frac{I_{k_r} p_{k_r}}{I_{k_r} p_{k_r} + \max\{I_l p_l\}}. \quad (17)$$

Otherwise the quality of each signal candidate detection  $s_{k_r}$  of the decoded signal trace  $\tau_k$  is equal to its probability prediction:

$$q_{k_r} = q_i = p_{k_r}. \quad (18)$$

If a training set is not available and the quality of signal candidate detections can not be assessed by a signal/noise classifier, then fluorescent intensities can be exploited for assessing the quality. Specifically, the fluorescent intensity of the signal candidate detection can be divided by the  $l_2$  norm of the vector of fluorescent intensities extracted from all the channels in the same signal location and imaging round,  $J(x_i, r_i)$ :

$$q_{k_r} = q_i = \frac{I_i}{\|J(x_i, r_i)\|_2}. \quad (19)$$

#### IV. EVALUATIONS

We evaluated the graph based decoding approach on both synthetic data, and real in situ sequencing data [7], comparing with alternative and gold standard decoding approaches [4], [9].

##### A. Quantitative evaluation on synthetic data

We used the starFISH synthetic data generation framework to evaluate the graph-based decoding (GRAPH) using fluorescent intensities as observation likelihoods (eq. 19), and two other alternative image decoding strategies (EM and NNS, described in paragraph II). We generated synthetic data with three increasing background noise conditions (Figure 2). And, additionally, we incorporated simulation of spatial jitter of fluorescent signals between imaging rounds. Thus, we further evaluated the different decoding strategies on three different spatial jitter conditions at fixed background noise (Figure 2). All the simulations were run on image datasets generated as in typical in situ sequencing experiments with four fluorescent channels and five imaging rounds, for a total of 1024 targeted barcodes sequences. A Gaussian kernel with 1.8, 1.2, 1.2 standard deviations for  $z$ ,  $y$ ,  $x$  axes was used for modelling the point spread function. For each of the noise and jitter condition we generated 50x50 px image datasets at five increasing density level of RNA molecules (i.e. 10, 20, 40, 70, 100 RNA molecules per image dataset) in 2D and 3D with 10 depth levels. We then detected signal candidate detections in each generated image dataset with the h-maxima transform using three different threshold levels to simulate several false positive - true positive signal detection rates. The same set of signal candidate detections derived by applying the h-maxima transform to all image channels and rounds of each datasets was given as input to the NNS and GRAPH decoding approaches. For the EM approach, signal candidate detections were instead extracted from the maximum intensity projections of the fluorescent image channels belonging to the first imaging round of each dataset, using the same h-maxima thresholds.

Next, we estimated optimal parameter settings for the three evaluated decoding approaches (i.e. maximum distance for signal candidate detections association between imaging round for NNS decoding strategy, and  $d_{th}$  for the GRAPH decoding approach with  $d_{max}$  set to  $d_{th} + 2$ ). We ran a grid search between probable parameter intervals to find the best parameter setting that optimized the f1 score of decoded sequences for each simulated condition. Specifically, for each decoding result we optimized the best possible match between decoded sequences and ground truth sequences, within euclidean distance equal to 2, through the Hungarian algorithm. And we computed f1 scores considering perfect matches as true positives, ground truth sequences without a matching decoded sequence as false negatives, and all the remaining decoded sequences as false positives.

We ran the grid search parameter estimation on 100 synthetic image datasets for each experimental condition discussed previously, thus selecting the median values of esti-

mated parameter settings for each experimental condition. Finally, we compared each decoding strategy on an independent test set of 100 image stacks for each experimental condition using the selected parameter settings (showing results in section V).

##### B. Quantitative evaluation on in situ sequencing data

We evaluated an image analysis pipeline (described in section IV-C) with an implementation of the graph-based decoding approach (ISS-GRAPH) with respect to the gold standard in situ sequencing image analysis pipeline implementation (ISS-EM) [9] based on the exact match decoding algorithm on in situ sequencing mouse brain data. As in many microscopy applications, ground-truth data for methods evaluation is difficult to obtain unless one has access to alternative detection methods. Hence, we compared decoding results obtained at two different image resolutions – thus pushing the signal size and signal-to-noise ratio of detections at the lower resolution samples. Specifically, we performed and compared ISS-GRAPH and ISS-EM analyses of approximately 1.79mm x 1.46mm mouse brain section imaged with 40x and 20x magnification objective to assess precision and recall of image decoding strategies. We first decoded the 40x image dataset in 3D and 2D with the ISS-GRAPH pipeline and 3D analysis was chosen as reference for the validation of 20x decoding analyses due to higher recall and better discriminatory power between true targeted sequences and false positives (Figure 3a).

We then ran analyses on the 20x image dataset (after aligning fluorescent channels and imaging rounds). First, we applied ISS-EM and ISS-GRAPH on a maximum intensity projection (MIP) of the image stacks. And second, we also applied image decoding in 3D, only applicable for the ISS-GRAPH pipeline.

Afterwards, for evaluating each of the 20x decoding result, we spatially aligned decoded RNA detections to the 40x reference by applying an affine transformation estimated with SIFT [11] feature based landmark registration in Fiji [12] between maximum intensity projected reference stain images of the 20x and 40x first imaging rounds. We then performed a second alignment based on locally affine point cloud registration of decoded barcodes. Where decoded sequences are divided into 1120x1376 px patches based on their spatial location and an affine transformation is estimated and applied for each patch using iterative-closest point algorithm with matches further than 5 pixels away excluded [2]. Aligned decoded barcodes were then paired with their nearest neighbor in the reference within a maximum euclidean distance of 3. For each pair of barcodes, we evaluated the number of mismatches in their decoded sequences. Due to the presence of a consistent amount of false positive sequences in the reference (caused by computational and biological noise) whose majority present one base mismatch with respect to one of targeted barcode (Figure 3a), we exploit the fact that targeted barcode sequences are encoded with one base redundancy allowing to detect one base errors. Thus, in order for the evaluations to be less

affected by errors in the reference, we considered as perfect matches (i.e. 0 mismatches) decoded sequences that are true targeted barcodes but present one base mismatch with the paired reference sequence.

### C. Image analysis pipeline for graph-based decoding (ISS-GRAPH)

Images from all fluorescent channels and imaging rounds are first normalized independently. The intensities values are scaled between the background level and the signal level estimated respectively as the mode and the 98th percentile of the image. Successively, fluorescent signals are enhanced with top-hat filtering and signal candidates are extracted from each fluorescent channel and imaging round applying the h-maxima transform. A merging step is then performed to aggregate signal candidates from the same imaging round but detected at the same location in multiple fluorescent channels. This is a necessary step to filter spurious signal candidate detections caused by fluorescent bleed-through between channels. Qualities of signal candidate detections are then assessed as probability predictions (eq. 17, 18) of a convolutional neural network trained in-house on a subset of 5x5 px windows centered on annotated candidate detections from multiple in situ sequencing experiments. Signal candidate detections together with signal qualities are then modeled in a cost-flow network as described previously and barcode sequences are finally decoded together with sequence quality scores.

## V. RESULTS AND DISCUSSION

Results of quantitative evaluation on synthetic data for each experimental condition are presented in Figure 2 in terms of precision, recall and f1 scores for 2D and 3D analyses.

All three evaluated decoding approaches (EM, NNS, GRAPH) show a general decrease in performance in terms of f1 score at increasing levels of signal density, noise, and spatial jitter. A critical factor that greatly impacts the precision and recall of decoded sequences is the false positive - true positive and miss-detection rates of input signal candidate detections given as input to the evaluated decoding approaches. The proposed GRAPH decoding approach shows better robustness in relation to noisy detections as compared to NNS both in terms of recall and precision since it can exploit quality observations for signal candidate detections other than spatial distances when assessing signal candidate associations between imaging rounds. Instead, methods like NNS and GRAPH that base their decoding strategies on signal associations between fluorescent channels and imaging rounds are more prone to have lower recall respect to the EM method at the same miss-detection rate, despite of being able to better resolve barcodes of overlapping RNA molecules. Indeed, EM does not require (as NNS and GRAPH) detection of all fluorescent signals of a given barcode, but instead a single detection of the fluorescent signal in the reference image is sufficient to decode a barcode sequence. On the other hand, NNS can partially compensate for miss-detections exploiting association of signal detections to multiple barcodes

(increasing recall), since it does not impose a constraint of non-overlapping barcode sequences. This could show an advantage when decoding barcode sequences of spatially close RNA molecules presenting overlapping fluorescent signals with a shared peak and consequently generating a single candidate detection. But at the same time, this could also have a negative impact on the results increasing the number of false decoded sequences, arising from the erroneous association of the same fluorescent signals to multiple barcode sequences, and consequently decreasing precision. Instead, the proposed GRAPH decoding approach models barcode sequences as non overlapping signal traces preventing a single signal candidates detection to be included in multiple barcode sequences during the decoding process.

When the effect of spatial jitter is also evaluated, EM shows considerably lower f1 scores for experimental conditions where jitter is greater than 2 as compared to GRAPH and NNS, noticeable in 2D and even more in 3D. Specifically, when the shift of fluorescent signal peaks caused by the spatial jitter is smaller than their point spread function, EM could still decode the right sequences looking only at the intensity values of the different fluorescent channels.

Performing analyses in 3D rather than in 2D generally does not improve results in terms of f1 score. Despite the fact that overlapping fluorescent signals could be better discerned in 3D, the additional dimension introduces a higher order of complexity in the decoding process that could lead to lower recall and precision of the results in dense and noisy experimental conditions. The proposed GRAPH decoding approach preserves high precision in most of the evaluated experimental conditions as compared to the other two alternative decoding strategies, showing that jointly optimizing quality and distance information of signal candidate detections can achieve a better robustness to noise.

Results from quantitative evaluation of in situ sequencing data are presented in Figure 3 with respect to quality of the decoded barcode sequences. The proposed ISS-GRAPH pipeline allows to resolve and decode barcodes in 2D and in 3D. Where, 2D analysis of maximum projected images of 3D focal planes are usually preferred for faster processing time. The proposed ISS-GRAPH pipeline achieves higher recall of decoded barcode sequences without losing precision with respect to the gold standard ISS-EM pipeline (Figure 3b,c), decoding a total of 59236 and 58873 true positives for 3D and 2D analysis, compared to 42283 true positives for the ISS pipeline.

## VI. CONCLUSION

We presented an alternative approach for decoding barcode sequences for multiplexed in situ RNA detections based on global data association for multi-object tracking using network flow. We evaluated the proposed approach on generated synthetic data at different experimental conditions in relation to signal density, noise level and spatial jitter, and we compared it with two alternative commonly used decoding strategies. We also evaluated an image analysis pipeline based on the

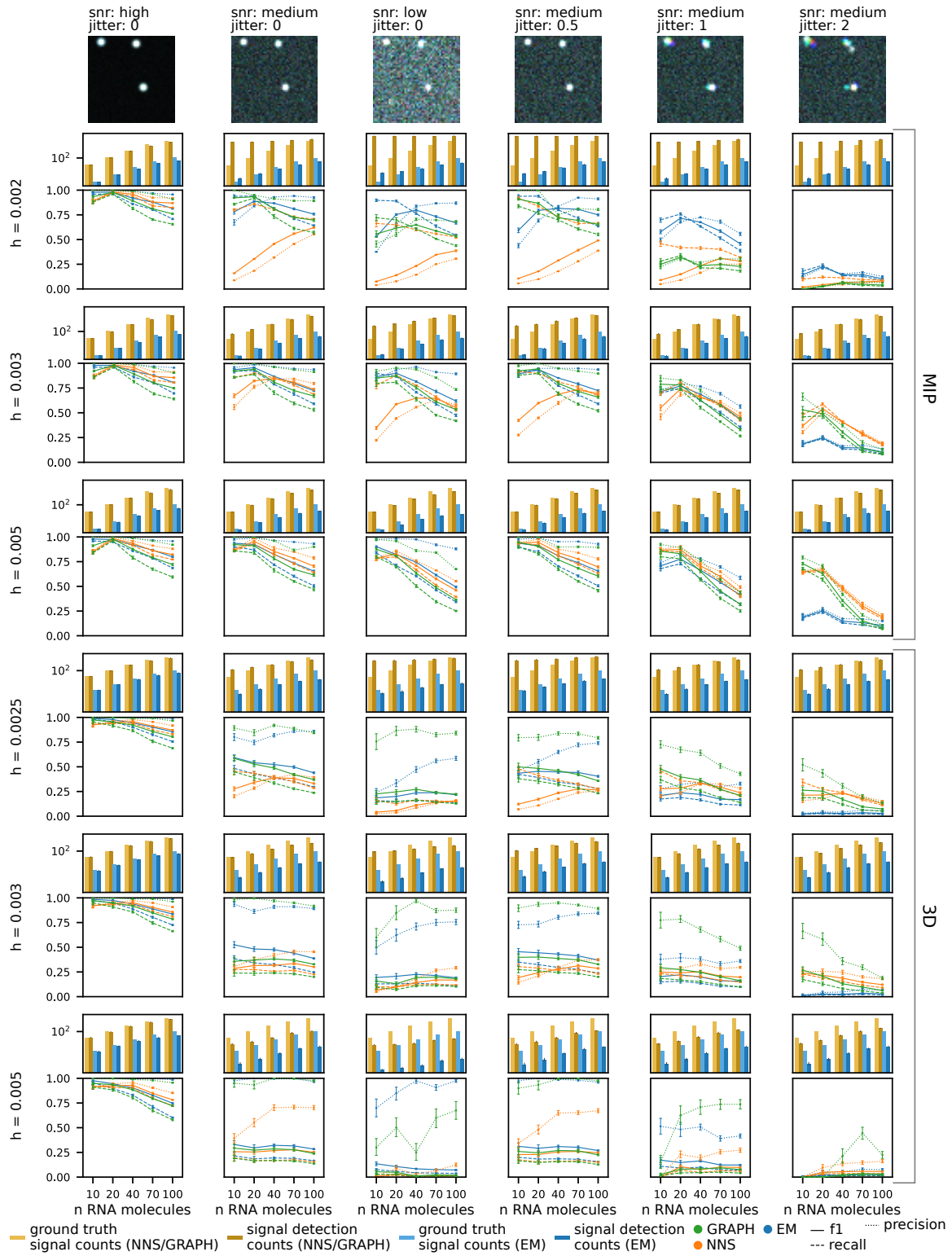


Fig. 2. Quantitative evaluation on synthetic data generated at different experimental conditions. Each column shows results of evaluated decoding approaches at different signal-to-noise ratios (snr) and spatial jitters (jitter). An example dataset of each experimental condition with three RNA molecules is shown at the top of each column as composite images of maximum projected fluorescent channels of five imaging rounds (with red, green, blue, gray, cyan pseudo-colors). Rows show 2D (MIP) and 3D decoding results at different false positive - true positive and miss-detection rates given by setting different threshold levels ( $h$  values) for the signal candidate detection algorithm ( $h$ -maxima transform). For a given threshold level and experimental condition, signal detections counts together with ground truth signal counts are shown on top of decoding results from the three evaluated approaches (EM, NNS, GRAPH). Note that both the NNS and the GRAPH methods match candidate detections across channels and imaging rounds, and thus share the same input set of signals. The EM method, on the other hand, detects signal candidates in the reference image, and therefore has a lower number of ground truth signals. Each decoding result is presented in terms of recall, precision and f1 score (bottom) with 95% confidence intervals at increasing number of RNA molecules.



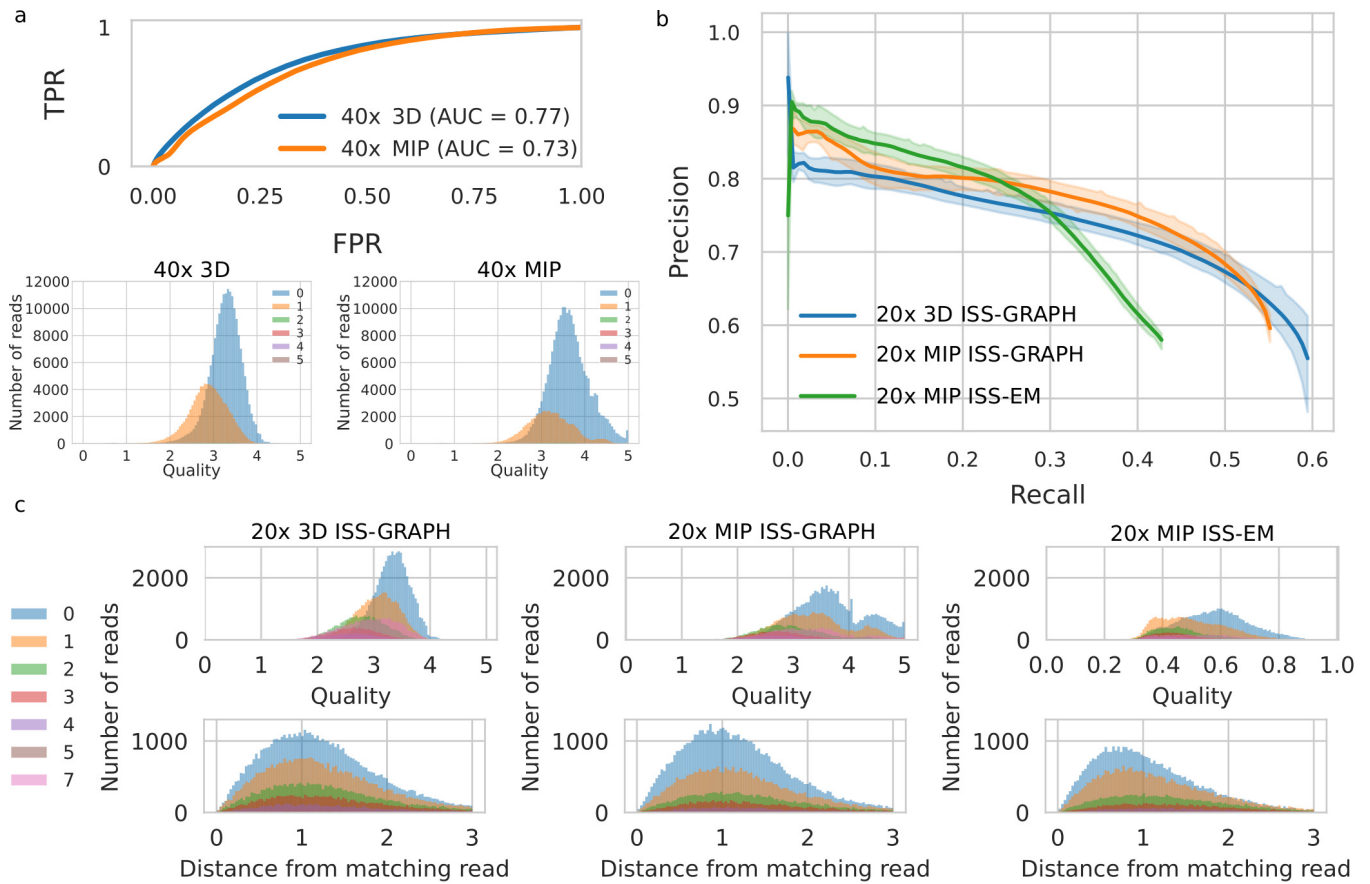


Fig. 3. Quantitative evaluation on in situ sequencing data. a) Receiver operating characteristic of the proposed graph-based decoding pipeline (ISS-GRAPH) for 3D and MIP (maximum intensity projection) analysis of 40x data using list of targeted barcode sequences to evaluate false positive rate (FPR) and true positive rate (TPR) (upper panel). And histograms showing number of decoded sequences (i.e. reads) versus quality for each distribution of decoded sequences with number of mismatch respect to a target sequence from 0 to the number of sequencing rounds (lower panel). b) Precision versus recall for 20x decoding results of 3D and MIP analyses of ISS-GRAPH, and MIP results of the gold standard in situ sequencing pipeline (ISS-EM) [9]. Precision and recall are evaluated on four equally sized parts of the image dataset at decreasing values of quality threshold, and mean and standard deviation are plotted for each of the three analyses. c) Pair of histograms showing for each 20x analysis the distribution of read counts with respect to read quality and to the distance from the reference matching read. Colors shown in legend represent number of mismatches with respect to the reference matching read. Specifically, zero is equal to perfect match, one to five represent the number of mismatches between the two sequences, and seven represents decoded false positive reads without a matching reference read.

proposed graph-based decoding approach on real data from in situ RNA sequencing and compared to the gold standard image analysis pipeline of the assay.

The proposed graph-based decoding is robust to noise and spatial jitter improving precision and preserving recall of decoded results by jointly optimizing quality and distances of signal candidate detections. Overlapping fluorescent signals are the main reason for miss-detections that greatly affects recall of decoding results due to the non-overlapping constraint for signal traces in the network flow model. Therefore, extending the proposed decoding method for modelling occlusions of signal candidate detections (as formulated in [10] for person tracking) could greatly benefit the decoding and defines a promising direction for improvement.

#### ACKNOWLEDGMENT

We would like to thank Gustav Björdal, Pierre Flener, Fred Hamprecht and the StarFISH team for useful input on the

algorithm development, and Markus Hilscher, Mats Nilsson and Anna Klemm for help with methods evaluation. This project was funded by the ERC-CoG 682810 to C. Wählby.

#### REFERENCES

- [1] G. Partel, M. M. Hilscher, G. Milli, L. Solorzano, A. H. Klemm, M. Nilsson, and C. Wählby, "Identification of spatial compartments in tissue from in situ sequencing data," *bioRxiv*, p. 765842, 2019.
- [2] X. Qian, K. D. Harris, T. Hauling, D. Nicoloutsopoulos, A. B. Muñoz-Manchado, N. Skene, J. Hjerling-Leffler, and M. Nilsson, "Probabilistic cell typing enables fine mapping of closely related cell types in situ," *Nature methods*, vol. 17, no. 1, pp. 101–106, 2020.
- [3] J. R. Moffitt, D. Bambah-Mukku, S. W. Eichhorn, E. Vaughn, K. Shekhar, J. D. Perez, N. D. Rubinstein, J. Hao, A. Regev, C. Dulac *et al.*, "Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region," *Science*, vol. 362, no. 6416, p. eaau5324, 2018.
- [4] S. Shah, E. Lubeck, W. Zhou, and L. Cai, "In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus," *Neuron*, vol. 92, no. 2, pp. 342–357, 2016.



- [5] W.-T. Chen, A. Lu, K. Craessaerts, B. Pavie, C. S. Frigerio, R. Mancuso, X. Qian, J. Lalakova, M. Kuhnemund, I. Voytyuk *et al.*, "Spatial and temporal transcriptomics reveal microglia-astroglia crosstalk in the amyloid- $\beta$  plaque cell niche of alzheimer's disease," *bioRxiv*, p. 719930, 2019.
- [6] R. Liu, M. Mignardi, R. Jones, M. Enge, S. K. Kim, S. R. Quake, and J. Zou, "Modeling spatial correlation of transcripts with application to developing pancreas," *Scientific reports*, vol. 9, no. 1, pp. 1–8, 2019.
- [7] R. Ke, M. Mignardi, A. Pacureanu, J. Svedlund, J. Botling, C. Wählby, and M. Nilsson, "In situ sequencing for rna analysis in preserved tissue and cells," *Nature methods*, vol. 10, no. 9, p. 857, 2013.
- [8] X. Wang, W. E. Allen, M. A. Wright, E. L. Sylwestrak, N. Samusik, S. Vesuna, K. Evans, C. Liu, C. Ramakrishnan, J. Liu *et al.*, "Three-dimensional intact-tissue sequencing of single-cell transcriptional states," *Science*, vol. 361, no. 6400, p. eaat5691, 2018.
- [9] A. Pacureanu, R. Ke, M. Mignardi, M. Nilsson, and C. Wählby, "Image based in situ sequencing for rna analysis in tissue," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2014, pp. 286–289.
- [10] L. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.
- [11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [12] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid *et al.*, "Fiji: an open-source platform for biological-image analysis," *Nature methods*, vol. 9, no. 7, pp. 676–682, 2012.