Computational Methods for Image-Based Spatial Transcriptomics

AXEL ANDERSSON
Why does cancer develop, spread, grow, and lead to mortality? To answer these questions, one must study the fundamental building blocks of all living organisms — cells. Like a well-calibrated manufacturing unit, cells follow precise instructions by gene expression to initiate the synthesis of proteins, the workforces that drive all living biochemical processes. Recently, researchers have developed techniques for imaging the expression of hundreds of unique genes within tissue samples. This information is extremely valuable for understanding the cellular activities behind cancer-related diseases. These methods, collectively known as image-based spatial transcriptomics (IST) techniques, use fluorescence microscopy to combinatorically label mRNA species (corresponding to expressed genes) in tissue samples. Here, automatic image analysis is required to locate fluorescence signals and decode the combinatorial code. This process results in large quantities of points, marking the location of expressed genes. These new data formats pose several challenges regarding visualization and automated analysis.

This thesis presents several computational methods and applications related to data generated from IST methods. Key contributions include: (i) A decoding method that jointly optimizes the detection and decoding of signals, particularly beneficial in scenarios with low signal-to-noise ratios or densely packed signals; (ii) a computational method for automatically delineating regions with similar gene compositions — efficient, interactive, and scalable for exploring patterns across different scales; (iii) a software enabling interactive visualization of millions of gene markers atop Terapixel-sized images (TissUUmaps); (iv) a tool utilizing signed-graph partitioning for the automatic identification of cells, independent of the complementary nuclear stain; (v) A fast and analytical expression for a score that quantifies co-localization between spatial points (such as located genes); (vi) a demonstration that gene expression markers can train deep-learning models to classify tissue morphology.

In the final contribution (vii), an IST technique features in a clinical study to spatially map the molecular diversity within tumors from patients with colorectal liver metastases, specifically those exhibiting a desmoplastic growth pattern. The study unveils novel molecular patterns characterizing cellular diversity in the transitional region between healthy liver tissue and the tumor. While a direct answer to the initial questions remains elusive, this study sheds illuminating insights into the growth dynamics of colorectal cancer liver metastases, bringing us closer to understanding the journey from development to mortality in cancer.

Keywords: Spatial omics, Spatial transcriptomics, Spatial biology, In Situ Sequencing, Visualization, Spatial statistics, Fluorescence microscopy

Axel Andersson, Department of Information Technology, Division Vi3, Box 337, Uppsala University, SE-75105 Uppsala, Sweden.

© Axel Andersson 2024

ISSN 1651-6214
ISBN 978-91-513-2052-6
URN urn:nbn:se:uu:diva-524009 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-524009)
List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


Reprints were made with permission from the publishers.
List of related papers

In addition to the papers included in this thesis, the author has also contributed to the following publications:


Contents

1 Introduction ................................................................................................ 11
  1.1 Spatial omics .................................................................................. 11

2 Image-based spatial transcriptomics ........................................................ 14
  2.1 In situ sequencing ........................................................................... 14
  2.2 The image data ............................................................................... 15
  2.3 The pipeline .................................................................................... 17

3 Restoration ................................................................................................. 19
  3.1 Removing the baseline ................................................................... 20
  3.2 Scaling the images ......................................................................... 22
  3.3 Color demixing ............................................................................... 22
  3.4 Deconvolving the point-spread function ....................................... 23

4 Registration ................................................................................................ 26
  4.1 Phase correlation ............................................................................ 26
  4.2 Stitching .......................................................................................... 27
  4.3 More degrees of freedom ............................................................... 27

5 Decoding .................................................................................................... 29
  5.1 Signal localization .......................................................................... 29
  5.2 Signal identification ....................................................................... 29
  5.3 Joint localization and identification .............................................. 30
  5.4 Quality control ................................................................................ 32

6 Spatial analysis .......................................................................................... 33
  6.1 Expression-based region identification ......................................... 33
  6.2 Visualization ................................................................................... 36
  6.3 Quantifying spatial patterns ........................................................... 36

7 The end of the pipeline .............................................................................. 39
  7.1 Summary of publications and contributions ................................. 39

8 Summary in Swedish ................................................................................. 47

9 Acknowledgement ..................................................................................... 50

References ........................................................................................................ 52
1. Introduction

The name Robert Hooke often evokes painful memories of high school physics lessons centered on Hooke’s Law and the dynamics of springs and elastic objects. Yet, Robert Hooke’s legacy extends far beyond this. As a 17th-century English physicist, Hooke was one of the pioneers in the microscopic exploration of the natural world. In 1665, Hooke published his seminal work, Micrographia, a detailed account of his microscopic observations [1]. This publication, filled with illustrations of insects, fungi, and plant life, became a cornerstone of biology and biological imaging [2]. It is celebrated for its artistic merit and for introducing the biological term cell. This book ignited an ever-long interest in microscopy and spurred advancements in lens technology, thereby revolutionizing our understanding of the microscopic world.

The legacy of Hooke’s Micrographia persists through roughly three centuries, with the volume still being printed today. However, the field of biological imaging has undergone a radical transformation. It encompasses an array of sophisticated instruments designed to elucidate biological processes and structures across multiple scales — from the molecular scale to the organismic.

Among the many microscopy imaging techniques, fluorescence imaging stands out as a cornerstone. This technique transcends the mere visualization of cells, enabling the detailed observation of organelles and biomolecules within cells and tissues. By employing fluorescent dyes, researchers can illuminate cellular and molecular components, thereby providing a window into the cellular machinery at work. Traditional fluorescence microscopy has paved the way for today’s advanced studies of the spatial organization of biomolecules.

1.1 Spatial omics

Spatial omics — the study of the spatial distribution of biomolecules — has emerged as a rapidly growing field of study, especially over recent years. This is rather unsurprising, given that the spatial arrangement of biomolecules within an organism is not just a matter of physical structure but is central to understanding how biological processes unfold in relationship to health and disease. The positioning of cells within tissues, for example, is a key determinant of their functionality and their role in the onset of various diseases.
Going back almost a decade, techniques such as immunostaining and single-molecule in situ hybridization (smFISH) were revolutionary, offering the first glimpse into the spatial organization of biomolecules [3, 4]. Yet, these methods were limited by spectral overlap, a physical phenomenon where the emitted light from one fluorescent dye interferes with another’s, restricting their multiplexing capabilities. ¹ This challenge spurred the development of modern spatial omics techniques, encompassing a diverse array of techniques today. Among them, we find proteomics methods, where the goal, as hinted by the name, is to study the spatial organization of proteins. These techniques can be categorized into two subcategories:

**Multiplex antibody staining techniques:** These proteomics techniques are derived from classical immunofluorescent staining. Here, fluorescently labeled antibodies are used to visualize proteins and other biomolecules within tissue samples. Repeated rounds of staining and signal removal mitigate the problem of spectral overlap, thereby increasing multiplexity. There are several variations of this technique, where some variations involve using customized antibodies for faster binding [5] or variations on how fluorescent signal is removed between the repeated rounds of imaging [6–8].

**Imaging mass cytometry techniques:** Like the previous, imaging mass cytometry techniques rely on antibodies to detect different biomolecules. The multiplexing limitations induced by the spectral overlap are mitigated by tagging the antibodies with metal isotopes instead of fluorophores. During imaging, a laser sweeps across the tissue, pixel by pixel, vaporizing the tissue. Mass spectrometry is thereafter used to discriminate the vaporized masses, making it possible to classify the antibodies and the corresponding biomolecules they have bound to [9–11]. This technique allows for higher multiplexity than the previous but suffers from poor spatial resolution.

Another set of spatial omics techniques relies on the spatial detection of RNA molecules. These are collectively known as spatial transcriptomics methods. Nonbiologists, like myself, might wonder what the point of measuring the RNA is. Recall that RNA serves as an intermediary between the DNA code within a cell’s nucleus and the functional proteins that execute cellular processes. Measuring the RNA is essentially mapping the active expression of genes at a given moment of time. Visualizing the RNA offers a snapshot of the genetic program being executed by the cell, which is of great value for furthering our understanding of health and diseases. The RNA-sequencing-based methods can broadly be categorized into three techniques:

**Tissue dissection techniques:** These approaches involve physically separating pieces of tissues, which are analyzed separately using, e.g., single-cell RNA sequencing methods [12, 13].

¹in other words, increasing the number of biomolecules that could be imaged simultaneously
**Sequencing-based techniques**: Microarray printing technology creates 'traps' on glass slides that capture mRNA from cells upon contact with tissue sections. The captured mRNAs are then sequenced. Sequencing data for individual mRNA molecules can then be mapped to the trap’s location, revealing the spatial distribution of different mRNA species [3, 14].

**Image-based spatial transcriptomics**: This final technique is about imaging individual mRNA molecules within tissue sections using fluorescence microscopy. This technique has been central to this thesis and will be described in more detail in the next chapter.
2. Image-based spatial transcriptomics

Image-based spatial transcriptomics (IST) refers to a collection of techniques that can directly image individual mRNA molecules within the tissue. Traditional single-molecule imaging methods are constrained by the finite number of available fluorescent dyes and spectral overlap, limiting the imaging to very few mRNA species. However, modern IST techniques transcend these limitations by employing innovative combinatorial staining strategies, enabling the simultaneous imaging of multitudes of different mRNA species.

Unlike sequencing-based techniques that target the entire transcriptome, IST techniques can only target a few hundred mRNA species. This is not necessarily a disadvantage. Sequencing-based techniques suffer from relatively low mRNA capture rates, so they tend only to pick up the most abundant molecular patterns. In IST, it is possible to target specific mRNA species, making it easier to pick up on the more intricate molecular patterns. The spatial resolution of IST techniques is also considered higher as the molecules are directly targeted in tissue. As of today, there are several methods within IST [3, 15–22]. The technique mostly studied in this thesis is in situ sequencing.

2.1 In situ sequencing

In situ sequencing, first introduced by Ke et al. [23], is a method based on padlock probes enabling the detection of hundreds of mRNA molecules directly in tissue. Unlike other IST techniques, in situ sequencing utilizes a rolling circle amplification [23] to enhance the fluorescent signals. This allows the imaging to be performed with standard fluorescence microscopes. Since its debut, several derivative techniques have emerged, improving detection efficiencies [21, 22]. One of these techniques is an RNA-direct in situ sequencing [22] that we describe in detail in Figure 2.1.

From an image-analysis point-of-view, the different in situ sequencing techniques (as well as IST techniques) tend to share two common traits: Firstly, detecting mRNA molecules is done using high-resolution imaging, resulting in massive amounts of spatially resolved data. Secondly, the different mRNA species are not labeled by a single fluorescent dye but by a combination of fluorescent signals that are either present or absent in different imaging rounds and channels. These shared traits present several challenges in the analysis of the data.
**Figure 2.1.** Overview of the key steps involved in the RNA direct in situ sequencing method. The process begins by hybridizing padlock probes (PLPs) to mRNA molecules within a tissue sample. The probes are then connected (ligated) and amplified using rolling-circle amplification (RCA). The amplified products, known as rolling circle products (RCPs), contain many repeated copies of a unique barcode. This barcode is specific to the mRNA molecule that the padlock probe has attached to. To identify the mRNA species, the barcode is imaged through several rounds of hybridizing L-probes and fluorescently labeled oligonucleotides (DOs) that are imaged using fluorescence wide-field microscopy. In the images below, each color represents a different fluorescent dye. By pinpointing the location of these colored signals and interpreting the sequence of colors across the rounds, the type and position of the mRNA molecules can be determined. This process is done automatically using image analysis and is called decoding. Image adapted from the original illustration by colleague Maria Conde.

### 2.2 The image data

The data generated from an *in situ* sequencing experiment can be characterized by a large multi-dimensional image stack. During the experiment, a tissue sample is repeatedly imaged over $n_r$ number of rounds. In each round, there are several types of fluorophores whose signals are captured in $n_c$ number of channels. The entire sample is usually too large to capture all at once, so
Figure 2.2. Visualization of the data produced from an in situ sequencing experiment. The left panel (A) shows the full image of the first staining round and first channel. Here, the field of views has been stitched together into a large composite image. In practice, the image data is captured in smaller fields of view, as is shown by the central figure (B). Usually, the data is captured with 4 to 6 rounds using 5 fluorescent channels and at various focus levels, resulting in a substantial amount of image data per field of view, as depicted in the right figure (C).

the imaging is done in $n_f$ different smaller fields of view. Each field of view consists of $n = n_h n_w n_d$ number of pixels, where $n_h$, $n_w$, and $n_d$ respectively refer to the height, width, and number of focus levels used for each tile. The image data can therefore be represented as an image stack $X \in \mathbb{R}_{+}^{n_f \times n_r \times n_c}$. The notation is summarized in Table 2.1, and Figure 2.2 shows a graphical example of the image data.

Consider a typical ISS experiment on a 1 cm by 1 cm tissue section using a pixel size of 320 nm, with 6 rounds, 5 channels, and 20 focus levels. This would generate $6 \cdot 10^{11}$ number of pixels. When saved in uint16 format, the data size reaches roughly a terabyte.

The size of the data presents challenges in handling it effectively. Most standard computers cannot fit one terabyte of data in their working memory. Recall that mRNA species are determined through various combinations of signals, present or absent, across different rounds and channels, see Figure 2.1. This means several images must be examined simultaneously to find where and of what type the mRNA molecules are. To manage this complexity, it is necessary to simplify the data into a less dense and more interpretable format. This simplification involves an extensive image-processing pipeline, see Figure 2.3, where the raw in situ sequencing data is broken down into a set of spatial points, pinpointing the location of the different species of mRNA within the tissue.
Table 2.1. The image data produced from a typical in situ experiment (values can vary) can be characterized as a multi-dimensional image stack $X$ with dimensions for fields of views, staining rounds, fluorescent channels, and spatial dimensions.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value(s)</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{fr\times ci}$</td>
<td>$\mathbb{R}_+$</td>
<td>A pixel in the image stack.</td>
</tr>
<tr>
<td>$f$</td>
<td>${1, \ldots, n_f}$</td>
<td>Index for different field of views.</td>
</tr>
<tr>
<td>$r$</td>
<td>${1, \ldots, n_r}$</td>
<td>Index for staining rounds.</td>
</tr>
<tr>
<td>$c$</td>
<td>${1, \ldots, n_c}$</td>
<td>Index for fluorescent channels.</td>
</tr>
<tr>
<td>$i$</td>
<td>${1, \ldots, n_hn_wn_d}$</td>
<td>Pixel index.</td>
</tr>
<tr>
<td>$n_f$</td>
<td>245</td>
<td>Number of imaged field of views.</td>
</tr>
<tr>
<td>$n_r$</td>
<td>6</td>
<td>Number of staining rounds.</td>
</tr>
<tr>
<td>$n_c$</td>
<td>5</td>
<td>Number of channels.</td>
</tr>
<tr>
<td>$n_h$</td>
<td>2000</td>
<td>Field of view width.</td>
</tr>
<tr>
<td>$n_w$</td>
<td>2000</td>
<td>Field of view height.</td>
</tr>
<tr>
<td>$n_d$</td>
<td>20</td>
<td>Field of view depth (number of focus levels).</td>
</tr>
<tr>
<td>$n$</td>
<td>$n_hn_wn_d$</td>
<td>Number of pixels.</td>
</tr>
</tbody>
</table>

2.3 The pipeline

In the previous sections, we discussed the generation of raw in situ sequencing data, denoted as $X$, and introduced some notations. Following the generation of the data, an extensive pipeline is used to transform this data into a format that is easier to work with. This format is usually a set of spatial points marking the location of the different mRNA species within the tissues. However, obtaining points is not the real end of the pipeline. Usually, the researcher wants to analyze and explore the molecular diversity within the tissue. The real end of the pipeline is, of course, a groundbreaking "scientific discovery".

We broadly divide the pipeline into three segments that we will go through in the coming chapters. The first segment, Restoration (Chapter 3), is dedicated to removing different types of stochastic and structural artifacts observed in the data. The second segment, Registration (Chapter 4), is dedicated to correcting the data for misalignment errors. The third segment, Decoding (Chapter 5), involves finding the location of the signals in the image data and decoding the corresponding combinatorial code. The output from this step is the set of spatial points for the various mRNA species, as is shown in Figure 2.3. The final segment (Chapter 6) is dedicated to the confirmatory and exploratory analysis of the points generated in the previous step.

1... which in reality is characterized by a deep sigh, a moment of perplexity, and a murmur of "that's funny..."
Figure 2.3. Image processing pipeline used in an IST experiment, starting from the raw fluorescence data and ending, ideally, with a scientific revelation.
3. Restoration

Let us begin with the first segment of our pipeline: image restoration. The fluorescence images we capture contain both stochastic and structural artifacts related to different physical factors inherent to the imaging. The process of compensating for these artifacts is broadly known as image restoration. Before diving into various restoration techniques, let us break down observed artifacts:

(i) **Photon noise**: Recall that we are working with fluorescence images. The fluorophores are excited once illuminated with light in a particular range. The fluorophores emit photons, which are captured by a camera, upon leaving the excited states. The discrete nature of photons emitted by the fluorophores results in variations in signal intensity. The number of photons captured by the microscope’s sensors is stochastic and follows a Poisson distribution [24–26].

(ii) **Read-out noise**: Noise is also introduced by the electronics in the microscope. Read-out noise arises from the electronic circuitry involved in reading the signal from a camera sensor. This includes the noise that the amplifier generates, the analog-to-digital conversion process, and the signal transmission to the storage medium. Read-out noise is often modeled as Gaussian white noise [25].

(iii) **Thermal noise**: Charge carriers within the camera’s sensors can be randomly excited due to the inherent thermal energy present in the material, resulting in noise [25].

(iv) **Spectral biases**: Each fluorophore has different spectral properties, such as different excitation and emission wavelengths. This will introduce variability in the signal strength recorded for the different fluorophores [27].

(v) **Camera settings**: Variations in camera settings, such as gain and exposure time, can also contribute to differences in signal intensity between imaging rounds and channels.

(vi) **Diffraction of light**: The signals should, ideally, appear as point sources in the images. However, due to diffraction and optical limitations in the instruments, the signals will appear "spread out". The shape of this spreading is modeled by a point-spread function (PSF) [28, 29].

(vii) **Channel crosstalk**: The light emitted from a fluorophore can be captured in several channels [30].
Optical aberrations: Imperfections in the microscope’s optics, such as the lenses, can result in distorted images [31].

Variability due to rolling circle amplification: The size of the rolling circle products might vary between molecules. Consequently, the signals can change slightly in appearance, especially if the products split into groups of signals [32].

Formally modeling all types of variations would result in a complicated model. We usually bundle (i) to (v) together and say that signal intensities in the respective rounds and channels are scaled by a multiplicative value that depends on the round and channel and shifted by a baseline. The baseline is assumed to be low frequency, i.e., slowly changing. The signal spreading (vi) and channel crosstalk (vii) can modeled and compensated using two relatively similar concepts called deconvolution and spectral demixing. In our work, we have not explicitly worked with addressing (viii). However, other authors have encountered aberrations such as vignetting and developed computational solutions [33], nor have we explicitly worked with (ix). However, we have noted that signals may vary slightly in shape and size. Perhaps this variation in shape can be used as a prior during the signal detection in later stages of the pipeline. We will now discuss standard methods for compensating many of these artifacts.

3.1 Removing the baseline

Removing the baseline from the images can be done in numerous different ways. The arguably most straightforward method that we used in Paper I and Paper VII is through a high-pass filter. This is useful for removing larger autofluorescence structures observed in the images and removing noise and additive offset in the data. At its core, a high-passed image can be computed as

\[ Y_{frci} = \max(X_{frci} - \sum_{j=1}^{n} X_{frcj}G_{ji}, 0) \]  \hspace{1cm} (3.1)

where \( i \) is a joint index for the spatial dimensions, and \( G \) represents the convolution with a low-frequency filter. Often, the filter is chosen as a Gaussian with a standard deviation larger than the width of the signal spots observed in the images. The Gaussian filter is also separable, meaning it can be efficiently applied as a series of one-dimension filters.

There are several alternatives to the very primitive high-pass filter. For example, the top-hat filter [23], the rolling ball filter [34], smoothing via regularized least squares [35], fitting of B-spline [36–38] or Gaussian mixtures [39]. Figure 3.1 (top row) shows a concrete example of the baseline and the effectiveness of the high-pass filter.
Figure 3.1. Example data from the *in situ* sequencing experiment performed in Paper VII. The first figure (A) shows a small tile of raw data. The red line corresponds to a 1D slice of observed intensities. The baseline is fitted using a simple Gaussian filter (B). The baseline-adjusted data is obtained by subtracting the baseline (and clipping negative values to zero), as shown in the right-most figure (C). The bottom row (D-E) shows intensities sampled at local maxima from two different fluorescent dyes in two baseline-adjusted channels. The histogram (F) shows the distribution of the intensities of the samples. It is clearly bimodal, with the upper modal corresponding to the intensities of the signals and the lower modal corresponding to the background. Notice that the upper modals do not align between the two dyes. Given the logarithmic scale of the horizontal axis, this suggests that a multiplicative scaling is necessary to avoid biases.
Figure 3.2. Example of crosstalk observed in an IST image. Red circles in (A) and (B) show two fluorescent signals that are visible in two channels. Crosstalk can clearly be visualized by plotting the strength of the signals in one channel versus the other channel (C). It is possible to compensate for the crosstalk by fitting a linear model to the observed intensities (C) and inverting it (D).

3.2 Scaling the images

The signal strength may vary between imaging rounds and channels by a multiplicative factor due to different spectral properties of the fluorophores and camera settings. This can be adjusted by identifying the average signal intensity per round and channel and scaling the images to be the same for each round and channel. This scaling is exemplified in Figure 3.1 (bottom row), where the points are sampled from the background and from local maxima from two different fluorescent dyes. The two violin figures show the distribution of the point intensities for the respective dyes. The distributions are shifted (in logarithmic scale), indicating a multiplicative scaling is necessary to avoid per-channel biases.

3.3 Color demixing

Typically, optical filters in microscopes are selected to isolate emissions from a single fluorescent dye per channel. However, the emissions from these dyes span a spectrum, leading to scenarios where a dye’s spectral range overlaps with the detection range of another channel. This overlap can cause signals intended for one channel to appear faintly in another, effectively mixing the signals between channels, see Figure 3.2A and 3.2B. The trend of crosstalk is visualized by plotting the intensity of peaks detected in one channel against the corresponding intensities in the other channels, see Figure 3.2C. The trend can be removed by simply fitting and inverting linear models [40–43].

To illustrate the color demixing, let us consider a simplified example (disregarding the different fields of view and imaging rounds) and consider a multi-channel image $X \in \mathbb{R}^{n_c \times n}$, where the signals are mixed in different channels. Assuming we can determine the mixing ratios between each pair of channels — for instance, by fitting a linear model to the intensities depicted in the figure.
ure — we could attempt to recover a demixed $Y \in \mathbb{Y}_{+}^{n_c \times n}$ image by solving the equation $Y = CX$, where $C \in \mathbb{Y}_{+}^{n_c \times n_c}$ contains the mixing ratios. However, this method can be improved by incorporating prior knowledge. Given that we are dealing with fluorescence microscopy data, where pixel values somewhat reflect the number of photons captured by the camera’s sensor, it is safe to reason that the image data should remain nonnegative and that there is some presence of noise. Hence, demixing is commonly framed as a nonnegative matrix factorization problem [44–47], aiming to deduce both the mixing matrix $C$ and the purified image $Y$ by minimizing a function that quantifies the disparity between the observed image and the reconstructed image, for example:

$$\arg\min_{Y \geq 0, C \geq 0} D(X, CY)$$  \hspace{1cm} (3.2)

Here, $D$ represents the error metric, often chosen as either the squared error

$$D(X, CY) = \|X - CY\|_{F}^2$$  \hspace{1cm} (3.3)

or the Kullback-Leibler divergence

$$D(X, CY) = \sum_{i=1}^{n} \sum_{c=1}^{n_c} \left( X \odot \log \left( \frac{X}{CY} \right) + CY - X \right)_{c, i}.$$  \hspace{1cm} (3.4)

The former assumes a setting with normally distributed noise, whereas the latter assumes a Poisson distribution. Solving the optimization problem often involves constraints and regularizations regarding $C$ and $Y$ [44–47]. In the following section, we will explore a specific algorithm tailored for another application that is somewhat analogous to the demixing problem.

### 3.4 Deconvolving the point-spread function

Notice from Figure 3.2 that the signals do not appear like small point sources but take the shape of small diffused blobs. The shape of the spots is determined by a point-spread function that depends both on the fluorophores’ spectral properties and the imaging system’s optics. This type of ’spreading out’ can become problematic if the signals are packed tightly or the signal-to-noise ratio is low. A way to improve the signal quality is to invert the effect of the point-spread function, resulting in images where the signals are represented as sharp point-sources.

Let us again simplify the notation and assume that $X \in \mathbb{R}_{+}^{n}$ is a simple image with $n$ number of pixels. Let us also assume a linear point-spread function that disperses signal intensities across adjacent pixels. We can model the point-spread function as a matrix $C \in \mathbb{R}^{n \times n}$, and set up an equation for the clean, in-focus, image $Y \in \mathbb{R}_{+}^{n}$ as

$$X = CY.$$  \hspace{1cm} (3.5)
The point-spread function is often chosen as a spatially invariant function and can be modeled as a convolution. Fourier transform can efficiently solve the above equation in such cases. Unfortunately, the above equation only makes sense in a world where the acquired image data is free from noise and the PSF is exact. In reality, we are constantly subjected to some degree of measurement noise. To obtain a desirable solution, we must take this noise into account. As discussed in previous sections, the noise is often assumed to be Poisson distributed, Gaussian, or a mixture of the two.

To keep things simple, let us assume a setting with only Poisson distributed noise, which is typically considered to be the dominant type of noise [26]. The observed image data can be considered as a realization from Poisson distributed parameterized by the rate $\lambda = CY$, that is

$$X \sim \text{Poisson}\left(\lambda\right). \quad (3.6)$$

The goal is thus to find $Y$, and thereby parameters of the Poisson distribution, $\lambda$, so that the observed data is as probable as possible. The quantity that measures this probability is the likelihood. Our objective is thus to find the clean image $Y$ that maximizes the likelihood. Since the likelihood function of a Poisson distribution is well-known, we end up with the optimization problem:

$$\arg\max_Y \prod_{i=1}^{n} \left( (CY)^X \circ e^{-CY} \div X! \right)_i \quad (3.7)$$

where the exponent, divisions, factorials, and multiplication ($\circ$) are element-wise. It is more convenient to work with the logarithm of the above equation. Taking the logarithm, dropping and adding some constant terms, and multiplying with $-1$, we end up with the equivalent minimization problem:

$$\arg\min_Y D(X, CY) \quad (3.8)$$

where

$$D(X, CY) = \sum_{i=1}^{n} \left( X \circ \log \left( \frac{X}{CY} \right) + CY - X \right)_i \quad (3.9)$$

is the Kullback-Leibler divergence [48]. The attentive reader will notice that this problem is rather similar to the one in the previous section (Eq. 3.4), where we discussed spectral demixing. The only difference is the interpretation of the matrices. However, before, we merely stated the optimization problem. Now, let us derive an incredibly simple algorithm for minimizing the above function. There are several ways to derive this algorithm, but the easiest is via the well-known gradient descent formulation. The gradient descent algorithm, given by its name, relies on descending the objective function by walking along its negative gradient. It can be formulated as

$$Y^{(t+1)} \leftarrow Y^{(t)} - \tau \nabla_Y D(X, CY^{(t)}) \quad (3.10)$$
where $\nabla$ refers to the gradient operator, $t$ is the optimization iteration, and $\tau$ is a parameter controlling the step size. The choice of $\tau$ is crucial. A value that is too small will lead to slow convergence. A value that is too large will lead to over-shooting and divergence. Plugging in the gradient gives us

$$Y^{(t+1)} \leftarrow Y^{(t)} + \tau \left( C^T \frac{X}{CY} - C^T 1 \right). \quad (3.11)$$

It turns out that convergence can be guaranteed if we choose $\tau = Y C^T 1$ \cite{48}. The update becomes

$$Y^{(t+1)} \leftarrow Y^{(t)} \odot \frac{C^T \frac{X}{CY^{(t)}}}{C^T 1} \quad (3.12)$$

where $1$ is a vector of ones, the divisions are elementwise, and $\odot$ is the elementwise product. While this algorithm was popularized by the famous non-negative matrix factorization papers by Lee et al. \cite{48,49}, it was originally introduced by Lucy \cite{50} and Richardson \cite{51} during the 1970s. It is often referred to as the Richardson-Lucy deconvolution algorithm. It is still used today for the deconvolution of microscopy data \cite{52–54}. Interestingly, recent studies have shown that replacing the transposed convolution, $C^T$, with a different filter can dramatically reduce the number of iterations needed for convergence \cite{55}.

This iterative process possesses certain implicit biases beneficial for processing IST data. Provided the initial guess and the point-spread function are nonnegative, the algorithm’s multiplicative updates preserve nonnegativity. Additionally, the algorithm is biased towards sparse solutions, which is desirable given that IST images contain sparse point sources rather than large coherent structures. Lastly, the total amount of fluorescent signal remains constant over the updates, that is:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} Y_{j}^{(t+1)} = \sum_{i=1}^{n} X_{i} \quad (3.13)$$

This can be verified by multiplying the left- and right-hand-side of the update (Eq. 3.12) with $C$ and summing over all pixels, that is

$$\sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} Y_{j}^{(t+1)} = \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} Y_{j}^{(t)} \frac{\sum_{k=1}^{n} C_{kj} X_{k}}{\sum_{m=1}^{n} C_{mj}} = \sum_{k=1}^{n} X_{k}. \quad (3.14)$$

Since all variables are nonnegative, we can replace the sums with the $\| \cdot \|_1$ norm, and we end up with $\| CY \|_1 = \| X \|_1 = \text{const}$. This type of constraint is both a curse and a blessing. On one hand, it naturally leads to sparsity. On the other hand, noise reduction does not occur automatically because the total fluorescent signal remains constant across updates. This algorithm has been of great value in this thesis. It has been used in Paper I and Paper VII.
4. Registration

Previously, we discussed artifacts related to related to signal intensities. Remember that our goal is to localize and identify the different mRNA molecules combinatorially labeled by fluorescent dyes in different imaging rounds. To identify this combinatorial code, the fluorescent signal from a particular rolling circle product must appear at the same location in the different imaging rounds. This is generally not the case due to several factors:

(i) **Field of view positional offset**: Mechanical limitations in the imaging system make it difficult to determine each field of view’s position precisely. Due to the rigidity of the sample and the construction of the stages where the sample is mounted, this offset in position is purely translational [44, 56].

(ii) **Between rounds positional offset**: Between the rounds, the sample is removed from the microscope for stripping and re-staining. Placing the sample back in the same position with submicrometer precision is extremely difficult (impossible). There will, therefore, be a positional offset between rounds. Moreover, repeated rounds of hybridization can slightly alter the shape of the sample, which cannot be modeled using rigid translations.

(iii) **Optical aberrations**: Imperfections in the microscope’s optics, such as the lenses, can result in distorted images. Chromatic aberrations were, for example, noted in the study by Qian et al. [44].

(iv) **Signal jittering**: The sample is taken out of the microscope between the imaging rounds for signal removal and restaining. This process may lead to a slight stochastic shift in the position of individual signals between rounds [57].

Methods for correcting these misalignment errors are often referred to as image registration methods. Regardless of the type of transformation causing the misalignment, starting the registration process with a basic alignment using only translation and then moving on to more intricate transformations is common practice. The phase correlation algorithm is a particularly effective and straightforward algorithm for image alignment.

4.1 Phase correlation

Phase correlation is a useful method for determining the relative shift between two images [58, 59]. It utilizes Fourier transform properties to compute
the displacement between two images using phase information, providing resilience against noise and variations in intensity. The magic behind phase correlation is that a phase shift in the Fourier transform indicates a translation relationship between two images.

Mathematically, for two images with respective Fourier transform $F_1$ and $F_2$, the phase correlation function can be expressed as

$$R = \mathcal{F}^{-1}\left(\frac{F_1 \odot F_2^*}{|F_1 \odot F_2^*|}\right)$$ (4.1)

where $*$ represents the complex conjugate and $\mathcal{F}^{-1}$ is the inverse Fourier transform. The translational shift is easily identified as the peak in $R$. Since it operates in the frequency domain and relies on phase information, it is less affected by noise than spatial domain techniques. This, together with the speed of the fast Fourier transformation, makes it an incredibly useful tool.

4.2 Stitching

The phase correlation allows us to stitch together fields of view into a large mosaic image. This is done by simply applying the phase-correlation algorithm between pairs of adjacent fields of views and estimating their relative translation. By defining one field of view to be positioned at the origin of a global coordinate system, the remaining fields of view can be aligned relative to it. However, there are more pairs of fields of view and, therefore, more estimated translations than there are fields of view. The question then becomes, which of the estimations do we choose when stitching the field of views together? This leads to an overdetermined problem. Two popular algorithms, ASHLAR [56] and MIST [60], solve this in a combinatorial way with minimum spanning trees. Qian et al. [44] solved the overdetermined system using least squares. We found ASHLAR to be particularly useful, especially for the stitching of images acquired in Paper VII.

4.3 More degrees of freedom

While stitching and registering images using phase correlation is efficient, both computationally and with regard to noise, it is limited to finding only translational shifts. It has been argued in a few studies that a simple translational alignment is not sufficient, as the samples might be slightly deformed between staining rounds or subjected to chromatic aberrations [44]. In such cases, higher degrees of freedom are necessary to align the images. Qian et al. [44] assumed affine transformations between the fields of view. The transformations were fitted by first identifying the location of signals, effectively
creating a point cloud, which was aligned using the iterative-closest point algorithm [61].

The software package accompanying the study by Lee et al. [62] included a registration model that assumed local translations, which, on a global scale, could also model non-linear transformations.

It has also been noted that individual signal spots can jitter between sequencing rounds. In general, this jittering is relatively small (few pixels) and is hard to compensate for using conventional registration algorithms. Instead, this is dealt with during later stages of the pipeline, during decoding, where we locate the signals and identify their combinatorial label [57, 63].
5. Decoding

We have now discussed the noise related to the raw image data and how to register the signals. At this point in the pipeline, we can assume that the images are aligned properly and adjusted for different types of noises. Let us, therefore, drop the field of view index from our image data (as the fields of view are all stitched into a single, coherent image) and instead work with a stack of images with dimensions for rounds, channels, and pixels. The goal now is to extract a sparser and "easier-to-interpret" representation of the data. Recall that the molecular types, that is, the mRNA species, are encoded in a combinatorial manner, as is shown in Figure 2.2. Our objective is, therefore, to localize signals and identify combinatorial signal patterns.

5.1 Signal localization

The challenge of localizing fluorescent signals is a classical image analysis problem that continues to attract research interest today, largely thanks to advancements and applications of smFISH methods. Classical approaches, which are still employed today, primarily focus on identifying signal peaks through local maxima identification via Laplacian or difference of Gaussian filters [64]. However, we have seen a shift toward more complex techniques in the past years. Deep learning-based techniques, in particular, have emerged as powerful tools, offering enhanced accuracy and robustness in signal identification [65–67]. These approaches are still rivaled by advanced classical methods [68, 69]. If each fluorescent dye could uniquely label a distinct molecule, the decoding process would boil down to only the detection of signals. However, combinatorial labeling techniques are employed to increase the number of detectable molecules. This complexity requires an additional step: identifying signal patterns to determine the mRNA species.

5.2 Signal identification

Once the signals are detected, the next step is to compare the signal patterns with a known set of reference codes, where each reference code is linked to an mRNA species. We call this set of reference codes the codebook. It can be expressed as a matrix of size $\mathbf{0, 1}^{n_g \times n_r \times n_c}$ where $n_g$ is the number of unique codes in the codebook. Let us assume that the signal has been detected at
Figure 5.1. The decoding process exemplified. The left panel (A) shows example data with one spatial dimension (columns) and rounds and channels concatenated into a single dimension (rows). Two different mRNA molecules are visible in this image, as depicted by the two patterns of present and absent signals along the rounds and channels. To decode the data, the first step is to detect locations where the signal is present, as shown in the middle panel (B). Once the locations are detected, the next step is to compare the signal pattern with a set of reference patterns that indicate the mRNA species (C).

To determine the code, i.e., the mRNA species, we can identify which code from the codebook has the largest inner product with the observed signal intensities in the rounds and channels at location $i$, that is

$$
\text{arg min } j \sum_{r=1}^{n_r} \sum_{c=1}^{n_c} D_{jrc}X_{rci}.
$$

(5.1)

Suppose the inner product passes some threshold or some other hand-crafted quality measure. In that case, we say that we have identified a particular mRNA species, and we store the molecule’s location and species in a list. Here, we chose the inner product as an example of quantifying the similarities between the observed fluorescent patterns and the references in the codebook. In literature, we have seen many different types of similarity functions [15, 19, 22, 23, 44, 70, 71]. Fortunately, a few years back, Starfish [64], a Python framework, was developed to unify all pipelines into a single framework. We have exemplified this decoding process in Figure 5.1.

5.3 Joint localization and identification

Decoupling the tasks of signal detection and decoding enhances computational efficiency significantly. By identifying signals across various rounds and channels individually, there is no need to load the entire dataset into memory simultaneously. Nonetheless, one could argue that decoding efficiency could benefit from analyzing patterns across all rounds and channels simultaneously. This idea has been explored by several researchers [72–75].
Figure 5.2. The decoding problem can be framed as a nonnegative matrix factorization problem. The observed image data can be modeled as a product of the codebook, the demixed image $Y$, and the point-spread function. The goal is to find the demixed image in which individual peaks directly indicate different mRNA species. The red box highlights an observed combinatorial signal pattern corresponding to an mRNA species. The blue boxes show the corresponding code in the codebook and the one-hot representation in the demixed image.

We proposed in Paper I a simple methodology that introduces a straightforward decoding strategy inspired by the Richardson-Lucy deconvolution technique, echoing the concepts discussed in previous sections but incorporating both the codebook and the point-spread function into the observation model. We let our image data, $X \in \mathbb{R}^{n_r \times n_c \times n}$, be represented as an image stack with images for all rounds and channels and stipulate the observation model

$$X_{rci} \sim \text{Poisson}(\lambda_{rci}), \quad \lambda_{rci} = \sum_{g=1}^{n_g} \sum_{j=1}^{n} D_{rg} Y_{gj} P_{ji}. \tag{5.2}$$

Our objective is to deduce $Y$, another image stack, indicating the signal intensity of each codebook entry at every pixel. This demixing process translates the combinatorial encodings in $X$ into straightforward, one-hot representation in $Y$, making it easy to identify the labeled molecule types based on the peaks in $Y$. This matrix factorization approach is illustrated in Figure 5.2.

We opted for the previously described Richardson-Lucy algorithm due to its ease of implementation and desirable implicit biases for both sparsity and non-negativity. By simultaneously taking into account signals from several rounds and channels, these matrix factorization-based decoding algorithms can dramatically improve decoding in crowded signals or low signal-to-noise ratios. However, the iterative nature of the optimization makes them computationally expensive, requiring dedicated hardware for speeding up the process. They also inherently require the images to be properly aligned.
5.4 Quality control

Visual Inspection

How do we evaluate the quality of the decoding? One of the simplest quality control methods is visual inspection. Here, the data’s large spatial scale and high dimensionality introduce challenges. We are tasked with examining signal spots that are submicrometer in size within multi-channel images that can span millimeters captured in different staining rounds. To address this needle-in-a-haystack problem, we enhanced our image viewer, TissUUmaps [76, 77] (Paper III), which is tailored for visualizing exceptionally large images. A notable addition to this tool is the Spot Inspector plugin, which facilitates the convenient examination of images across multiple rounds and channels. Figure 5.3 shows the graphical user interface of TissUUmaps.

False Discovery Rate

As a precautionary measure, integrating unexpected codes into the codebook — codes that are not anticipated to be present in the dataset — has, in our research, proven to be extremely useful for finding errors in the decoding. The relative frequency of these unexpected codes versus the expected ones serves as a useful metric for detecting discrepancies or errors in the decoding process.

After decoding, we finally end up with a set of spatial points with categorical labels. In the following chapters, we will discuss how to analyze these point-patterns.
6. Spatial analysis

We have now reached the fourth segment of our pipeline. In this segment, we are working with data that consists of spatial points, each representing the presence of an mRNA species. It is important to note that mRNA molecules act as messengers, carrying genetic instructions from DNA to the machinery within cells that synthesize proteins. This connection is why we will now change terminology and refer to these mRNA species more simply as genes, as each mRNA molecule represents a gene being expressed. Thus, each spatial point effectively marks the location of an expressed gene. We let $e_i \in \{0, 1\}^{n_g}$ denote one-hot-encoded label for the $i$-th point’s corresponding gene, where $n_g$ is the number of unique genes targeted in the experiment. The first analysis step is often to identify regions, such as cells or tissue-level structures, containing similar compositions of genes.

6.1 Expression-based region identification

It is possible to identify different cell types and tissue-level structures based on the composition of genes within small spatial regions. A common yet powerful way of doing this analysis is by counting the frequency of the different genes within spatial bins. The frequency of the different genes within the bins serves as a genetic signature of the local neighborhood. We can then identify repeated gene expression signatures by clustering the bins. Since each bin is associated with a spatial location, we can map these clusters back onto tissue, effectively creating a map indicating regions with similar gene expression (and thus similar composition of cell types). The frequency of the different genes within different bins can be computed as

$$x_i = \sum_{j=1}^{m} w_{ij} e_j$$

(6.1)

where $x_i$ is the $i$-th bin, $m$ is the number of points in the data, and $w \in \mathbb{R}^{n_b \times m}$ is a matrix that assigns each point to a bin. The bin assignment matrix, $w$, can be defined in several ways. It depends on where and how we place the bins in the tissue and assign the points to each bin. We have attempted to summarize the different methods used in literature in Figure 6.1. One particularly common method is to assign genes to cells defined by cell segmentation.
Figure 6.1. A common procedure for identifying regions with similar gene expression patterns is computing the frequency of different genes within spatial bins, followed by clustering. Bins can be placed differently within the tissue. Common examples are: On a regular grid [70, 78–80], on a diamond [81], per point [57, 82], based on cell segmentation [18, 19, 44, 83, 84], or based on peaks in density [85], as is shown in the top row (A). The assignment of genes to bins can also be done differently (B). Common methods are based on nearest neighbor [82], based on distances [70, 78–81], based on the borders defined by cell segmentation [18, 19, 44, 83, 84], or based on probabilities derived from co-localization patterns in the data [44, 82, 86]. The bottom row (C) shows an example of the clustering procedure using a \( k \) nearest neighbor assignment method. Bins are placed on top of each point. Here, only two bins are shown (red and purple), where the dotted lines indicate which points are assigned to them. The resulting gene-count-per-bin vectors are used in clustering to reveal regions with similar gene expression compositions.
Bins defined by cell segmentation

Cell segmentation is arguably the most common problem in biomedical image analysis. As a colleague of mine — who works with providing image analysis support to bioimaging researchers at Uppsala University — says: "If I get ten requests for support, ten include cell segmentation".

In IST, cell segmentation is done based on nuclear-stained images obtained in parallel with the experiment. This is often done using deep-learning powered techniques [87] such as StarDist [88, 89] or Cellpose [90]. Genes are then assigned to segmented nuclei based on spatial proximity. This segmentation method assumes that all cells are convex and have nuclei that can be easily detected.

In reality, cells are not necessarily convex [86], and simply assigning genes based on distances might not be optimal. Moreover, detecting the nucleus might be difficult due to the quality of the sample, the density of cells, and the fact that cells can have part of their cytoplasm within the image while the nucleus is simply outside the imaged region. As such, researchers have experimented with using the points for gene expression as input for segmenting the data [82, 86, 91–93]. In Paper IV we explored the idea of using a signed graph-partitioning algorithm for segmenting cells without relying on nuclei segmentation.

Clustering the bins

Whether the bins are derived from cell segmentation or other methods, the subsequent step involves clustering these bins based on the gene composition within them. The clusters can indicate different biological objects depending on the size of the bins, as well as the minimum distance for which genes are assigned to the bins. For example, this type of clustering can be used to label subcellular structures [81], cell types [94], or larger tissue domains [95].

There are many different techniques for clustering the bins. For example using mixture of Gaussian models [70, 96], clustering done using community detection on graphs [85, 97–103] or clustering done using factor analysis [81, 104–107]. We also find applications where graph-neural networks are used before the clustering to integrate spatial information with gene expression information [70, 94, 95, 108–110].

Clustering the bins can help the researcher dissect their high-dimensional data into a few comprehensible components. In Paper II, we wanted to address the issue that many methods are often limited to a particular spatial scale, whether it be cell typing or tissue-domain identification. This, combined with lengthy optimization and the lack of interactive visualization, limits these methods’ exploratory capabilities. As such, we developed Points2Regions, a simple and versatile tool that is fast, works across scales, and can be used directly within popular image viewers.
The next step after clustering is to explore the clusters within a spatial context. This leads us to the next section, visualization, which is crucial for exploring and identifying spatial patterns.

6.2 Visualization

Visualizing IST data is fundamentally challenging. A single experiment can generate millions of points for hundreds of different genes. Such quantities require computationally efficient solutions for rendering. Moreover, patterns can be explored at different spatial scales. Ranging from subcellular [111] patterns up to patterns on a millimeter scale. Interactive solutions are, therefore, essential in the exploration of these datasets. This led us to further develop TissUUmaps: a user-friendly software [77, 112] tailored for datasets with large numbers of points.

Napari [113] stands out as another valuable image viewer, excelling particularly in displaying multi-dimensional image data, although it is not optimized for datasets with large quantities of markers. Additionally, credit is due to Vitessce [114], an impressive framework for creating customizable and interactive visualizations.

The exploration of IST data through clustering and interactive visualization techniques lays the foundation for identifying biologically interesting patterns. Upon uncovering these patterns, a new form of confirmatory or quantitative analysis comes into play.

6.3 Quantifying spatial patterns

There are several methods for quantifying spatial patterns [115]. Methods such as Ripley’s function [116], Newman’s assortativity [117], and centrality scores [118] can be used to assess if points of a particular label exhibit a spatially random pattern or not. Cluster co-occurrence probability [84] and object-object correlation analysis [119] are other valuable methods for assessing the non-random distribution of points of different types in comparison to the distribution of other points within the tissue.

One specific method we have explored a bit more extensively in Paper V is the neighborhood enrichment test.

Neighborhood enrichment test

Neighborhood enrichment tests [84, 102, 120–122] are valuable for quantifying enrichment between spatial points with categorical labels. In our context, these points usually correspond to cells or genes of different types. At its core, the neighborhood enrichment test checks how often points with a particular
Figure 6.2. The computation of the neighborhood enrichment score. (A) Points corresponding to different genes are located in tissue. Points are said to be neighbors if they are connected by an edge in a graph (B). Here, the edges are generated arbitrarily, but in practice, they are defined based on some criteria related to the spatial distance between the points. For example, if two points are within a set distance, they are considered neighbors. The number of neighbors between reference points (Gene A) and query points (Gene B) are computed (C). The same quantity is computed in several scenarios where the label of the points has been scrambled (D). The expectation and standard deviation in the number of neighbors during randomization are estimated to compute the z-score (Eq. 6.2).

reference label are neighboring points with a particular query label. The observed number of neighbors is compared against several baselines where the labels of the points have been shuffled. The test output is a z-score that indicates if points with the two labels are spatially enriched (co-localized). To use this test, one must define what constitutes a neighbor. This can be done in different ways. For example, if a point with a query label is within a threshold distance of a point with a reference label, the two are defined as neighbors. Mathematically, the neighborhood enrichment score given by the test is

$$z_{AB} = \frac{c_{AB} - \mu_{AB}}{\sigma_{AB}}, \quad (6.2)$$

where $c$ is the observed number of neighbors between points with label A and points with label B. The variables $\mu_{AB}$ and $\sigma_{AB}$ are the expectation and standard deviation in number of reference-query neighbors observed when the points’ labels had been scrambled. These variables are usually estimated using several scramblings of the labels. Figure 6.2 shows a graphical example of the computation of the enrichment scores.

In Paper V, we introduced an alternative resampling strategy. Rather than permuting all labels (keeping the label abundances fixed), we propose to resample based on the empirical probability of finding a label in the dataset. This allowed us to compute an analytical expression for the enrichment score, dramatically improving computational time and scalability, while preserving information.
### Table 6.1. List of toolboxes for IST analysis.

<table>
<thead>
<tr>
<th>Toolbox</th>
<th>Language</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scanpy [99]</td>
<td>Python</td>
<td>Statistical analysis of omics data.</td>
</tr>
<tr>
<td>Squidpy [84]</td>
<td>Python</td>
<td>Spatial analysis of omics data.</td>
</tr>
<tr>
<td>Matisse [79]</td>
<td>MATLAB</td>
<td>Spatial and statistical analysis of single-cell data.</td>
</tr>
<tr>
<td>Giotto [102]</td>
<td>R</td>
<td>Spatial and statistical analysis of omics data.</td>
</tr>
<tr>
<td>Seurat [123]</td>
<td>R</td>
<td>Statistical analysis of omics data.</td>
</tr>
<tr>
<td>Starfish [64]</td>
<td>Python</td>
<td>Image analysis for decoding toolkits of IST data</td>
</tr>
<tr>
<td>pySpacell [124]</td>
<td>Python</td>
<td>Spatial analysis of cell images.</td>
</tr>
</tbody>
</table>

### Toolboxes

While the field of IST is relatively young, there are still plenty of available toolboxes that unify many of the spatial analysis methods into a single framework, see Table 6.1.
7. The end of the pipeline

We have reached the end of a comprehensive pipeline, encompassing stages from data image restoration, registration, decoding and followed by downstream analysis and visualization. In this concluding chapter, I will summarize my contributions across various research papers in the thesis.

7.1 Summary of publications and contributions

Paper I — ISTDECO: In situ transcriptomics decoding by deconvolution

In this project, we set out to frame the decoding of IST images as a type of matrix factorization problem, as described in Chapter 5. By combining a priori known factors, such as the codebook and the point-spread function, into a unified optimization problem, we developed a decoding method resilient to overlapping signals and noise compared to approaches where detection and identification are treated as distinct steps, as shown in Figure 7.1.

Initially, we employed an optimization approach with strict constraints for sparsity, namely, using a matching pursuit algorithm. While the matching pursuit algorithm was good, it was not much better than heuristics, such as finding peaks, making it hard to motivate why one would opt for an optimization approach. Later, we sought inspiration from deconvolution literature (as opposed to dictionary learning). We found that the Richardson-Lucy algorithm works well despite its simplicity. The matrix factorization approach works incredibly well in situations with tightly packed signals; see, for example, Figure 7.1. It, however, requires that certain assumptions are met, such as the registration of images. Solving the optimization problem is also expensive, given the large number of pixels, as pointed out in [125].

For this manuscript, I am the primary author and wrote the first draft with supervision from co-authors. The algorithm was developed in collaboration with co-authors and implemented by me. I implemented the methods and carried out the evaluation.

Paper II — Points2Regions: Fast, interactive clustering of imaging-based spatial transcriptomics data

While engaged in collaborative projects like Paper VII, we frequently employed clustering techniques to pinpoint areas with distinct gene expression
patterns. It became evident that these patterns manifest at various scales, such as micrometer-scale patterns defining different cell types and millimeter-scale patterns delineating tissue-level structures. However, existing tools often leaned heavily on advanced machine learning and additional data types like single-cell RNA sequencing or cell segmentation, limiting their effectiveness in identifying patterns across multiple spatial scales.

In this project, we set out to develop a tool that would favor exploratory analysis. We suggested that efficient exploratory analysis would greatly benefit from a tool that can be applied at different scales, is fast to compute, and provides clear data visualizations. As such, we develop Points2Regions (Paper II), a clustering tool for identifying regions with similar gene expression patterns.

In simulated data tests, we observed that Points2Regions, despite relying on straightforward machine learning, demonstrated performance comparable to many complex algorithms limited to a particular spatial resolution. Additionally, real-world dataset tests validated that the clusters produced were consistent with previously published analyses.

To enhance accessibility, Points2Regions has been released in various forms: as an interactive plugin within two widely used visualization software, Napari, and TissUUmaps, allowing for interactive exploration; an online, install-free service; and a user-friendly Python program.

**Figure 7.1.** Experimental results from Paper I. The top row shows a channel-wise max projection from one of the rounds in simulated IST datasets. The datasets were simulated in four scenarios, each scenario with a different total number of signals. The red dots mark the center of the signals. Method performance was assessed using $F_1$-score on 25 simulated datasets for each scenario. Both methods, ISTDECO (ours) and BarDensr [72] are based on similar concepts (non-negative matrix factorization) and perform remarkably well despite a large number of overlapping signals.
I am the primary author of this manuscript. I wrote the manuscript (with input from co-authors) and developed and implemented the algorithm and experiments. The plugins were implemented in collaboration with co-authors.

Paper III — TissUUmaps 3: Improvements in interactive visualization, exploration, and quality assessment of large-scale spatial omics data

While engaged in the development of Paper I, we encountered challenges in visual quality control. The process of simultaneously visualizing multiple signals from various rounds and channels proved to be cumbersome. By also recognizing the need for software capable of efficiently displaying millions of different points, we set out to enhance our existing tool, TissUUmaps. The software had a robust foundation but faced limitations concerning user-friendliness and the capacity to visualize numerous points. I am the primary author, along with Nicolas Pielawski. My contributions were conducting the experiments, writing the manuscripts, and developing the software. However, I want to especially acknowledge former colleagues Dr. Fredrik Nysjö and Dr. Christophe Avenel for their excellent work in improving the software. The software continues to grow, becoming more widely adopted within spatial omics. The software today is available at https://github.com/TissUUmaps/TissUUmaps.

Paper IV — Cell segmentation of in situ transcriptomics data using signed graph partitioning

The idea behind this manuscript came from Paper II, where we noticed that simple machine learning was sufficient to cluster the data. While Paper II can be thought of as a method for semantic segmentation, here, we wanted to address the problem of instance segmentation. Specifically, the goal was to segment individual cells based on the points for the different genes. We chose to work with the Mutex Watershed algorithm [126] since it does not rely on any seeds, is generally fast, and relies on signed graph partitioning.

The idea was thus to create a signed graph where strongly positive (attractive) edges should connect points for different genes belonging to the same cell. Oppositely, points belonging to different cells should, ideally, be connected with strongly negative edges. We define the strength of these edges by comparing local gene expression distributions. Here, the advantage of signed-graph partitioning comes into play. It allows us to leverage both positive and negative information. For example, if two points are separated by a short distance and surrounded by similar types of genes, we know they likely should belong to the same cell. On the contrary, if a long distance separates two points, we know that they should belong to different cells.
We benchmarked our method against Baysor [82], an established method, and found that the median intersection over the union of segmented labels was around 80% for two datasets.

Our method is still crude and can be improved in terms speed and the prediction of edge strengths. I am the primary author, along with Andrea Behanova. My contributions here were developing the methods, conducting the experiments, and writing the manuscript.

Paper V — An Analytical Neighborhood Enrichment Score for Spatial Omics

The idea behind the fast neighborhood enrichment test came after working with Ref [120], where an interactive tool for visualizing the enrichment scores was proposed. However, this tool required pre-computed scores, making it less accessible for non-programmers. I noticed that the enrichment score could be computed analytically by changing the resampling from permutation to sampling based on empirical probability. This dramatically sped up the computation while preserving the information of the original score. I am the sole contributor to this project (which is ongoing).

Paper VI — Transcriptome-Supervised Classification of Tissue Morphology Using Deep Learning

In this publication, we wanted to explore alternative ways of training deep neural networks to classify tissue morphology. Rather than relying on manual annotations, we wanted to propose the idea of models trained with markers from an ISS experiment.

We noticed that we could, with relative ease, classify two anatomical regions in a mouse brain using a small deep-learning model trained on patches extracted based on gene expression. Today, *in situ* sequencing is expensive (a single experiment costs around 40kSEK), making this a rather expensive and complicated replacement for manual annotations. However, more accessible methods are being developed. With deep learning, this study hints that one might be able to predict gene expression from tissue morphology.

I am the primary author. My contributions were writing the manuscript (with input from co-authors), implementing the methods, and running the experiments.
As of today, colorectal cancer is the third leading cause of cancer-related mortality. Roughly 30% of patients will develop colorectal liver metastases (CLM). In the study of CLM, researchers focus on two specific ways tumors grow, known as histopathological growth patterns (HGP): the desmoplastic and replacement patterns. The desmoplastic pattern (DHGP) is characterized by a protective rim that separates tumor cells from healthy liver cells and is linked to positive outcomes after liver surgery. In contrast, the replacement pattern (RHGP) lacks this rim. In RHGP, tumor cells infiltrate liver plates, replacing healthy cells and relying on existing blood vessels for survival.

In this project, we explored the molecular mechanisms behind these two growth patterns and, perhaps, answer the question: Can the DHGP be induced? And could there be novel oncological therapy targets, thereby improving CLM patients’ survival?

This interdisciplinary collaboration started after helping Dr. Olga Surova, at Stockholm University with the computational part of her in situ sequencing samples. This project involved several samples, generating 75TB of data using three in situ sequencing panels. Our initial study, Paper VII, focused on the samples showing DHGP and the molecular diversity within the protective rim.

Key findings from the study were different areas within the protective rim.

The outer part, close to the liver, showed a general response to liver injury, where specific genes like NGFR were identified. These genes are known to play roles in fibrosis. In contrast, closer to the cancer cells, the inner part of the rim seems to consist of a stroma that is educated by adjacent tumor cells.

I am the primary author, along with Maria Conde. My contributions to this manuscript were doing all computational analysis interactive visualizations, writing many technical parts, and preparing figures for the manuscript. This project will continue with additional in situ sequencing panels to further elucidate the molecular diversity of the two growth patterns.

Related Paper I — End-to-end Multiple Instance Learning with Gradient Accumulation

In this project, we proposed a gradient accumulation strategy that enables end-to-end training of a particular kind of deep neural network architecture used in attention-based multiple instance learning [127]. These architectures are frequently used to classify histopathological slides since they only require slide-level annotations but can provide explainability in the form of attention-heatmaps on a patch level. In brief, the architecture consists of a feature extractor that transforms patches extracted from a slide into lower-dimensional representations. The representations are then pooled together using a weighted
average, where the weights are learned and interpreted as attention (how much the network attends to each patch). The pooled attention is fed through a multilayer perceptron (MLP) to give the classification score. We show that the entire model, starting from the feature extractor and ending with the MLP, can be trained jointly and is only limited in memory by the size of an individual patch.

I am the primary author, along with Nadezhda Koriakina. My contributions to this manuscript were developing the method, implementing the method and experiments, and writing the manuscript.

Related Paper II — Visualization and quality control tools for large-scale multiplex tissue analysis in TissUUmaps 3

In this manuscript, we develop several interactive visualizations in TissUUmaps for quality control for large-scale tissue analysis in TissUUmaps 3. My contributions were helping in the development of some methods and assisting in writing the manuscript.

Related Paper III — Optimizing Xenium In Situ Data Utility by Quality Assessment and Best Practice Analysis Workflows

This manuscript contains benchmarking of several methods in Xenium data [128] (the latest in situ transcriptomics machine developed by 10x Genomics) to determine best practices in terms of analysis workflows. My contributions were implementing and evaluating one of the segmentation-free methods (Points2Regions) and performing parts of the sub-cellular cluster analysis.

Looking ahead

This thesis has been centered around analyzing and developing tools for data generated from IST techniques, focusing on in situ sequencing. The evolution of IST techniques has been significant throughout my PhD journey, with several new derivative methods [21, 22] having been developed. As of very recently, stand-alone machines have been launched. These machines streamline the entire process, from inputting tissue to conducting image analysis and staining, ultimately outputting gene points with their respective positions [128]. Despite these cutting-edge advancements, these machines remain unaffordable to many laboratories. Thus, decoding and quality control tools are therefore still valuable for laboratories doing manual in situ sequencing. Here, I am incredibly enthusiastic about the ongoing efforts by the molecular diagnostic group at Stockholm University, actively working on improving the
chemistry of the *in situ* sequencing techniques, making it cheaper and more accessible.

However, a notable limitation with techniques such as *in situ* sequencing is the extensive image analysis required for data processing. While there are toolboxes like Starfish [64], a user-friendly end-to-end framework tailored for non-programmers is still lacking, making it hard for labs without dedicated computer scientists to conduct the experiments. Such a software could, perhaps, be achieved by stitching several pieces of software into a combined one. For example, ASHLAR [56] for image registration and stitching, Starfish [64] for decoding, Cellpose [90] cell segmentation, and SquidPy [84] for downstream analysis. It would, of course, require an extensive amount of dedication to develop and maintain such software.

During my PhD studies, we briefly discussed if the pipeline could be simplified by introducing a new data format. As the data size often exceeds the computers’ working memory, one must employ different tiling procedures, resulting in many copies of the already large dataset saved to disk. However, the data is sparse in the spatial domain. Most pixels belong to the background. This made us wonder if the data could leverage this sparsity to create a compressed, easier-to-work-with format. However, this is something we only discussed, but it could be an exciting direction for future research.

Throughout the course of my PhD studies, I have noticed two reoccurring problems related to multiplexed fluorescence imaging — whether it be IST or proteomics.

The first problem is related to stitching and registration. While ASHLAR [56] stands out as a valuable tool for stitching and registration, it is limited to two degrees of freedom, specifically translations. However, assuming purely translational offsets might prove insufficient when optical aberrations are present or when the sample undergoes slight deformation across consecutive staining cycles. This limitation could be mitigated by employing alternative registration techniques or combining multiple methods. A unified framework for registering multiplex fluorescence microscopy data would be of great value.

The second problem involves signal heterogeneity between fluorophores, tissue samples, and samples collected in different tissue microarray cores. An interesting direction for future research projects could be developing a model grounded in assumptions about the underlying physics of the data. Leveraging these assumptions, one could formulate an inverse problem to standardize the data. This approach is akin to our work in Paper I but with a broader application, not limited to IST decoding.

I am glad to announce that work related to Paper VII will continue together with Hanna Nyström and Maria Conde, with additional *in situ* sequencing panels designed to decipher the molecular diversity behind the histopathological growth patterns found in colorectal liver metastasis. Throughout our interdisciplinary project, we have benefitted greatly from our image-viewer TissU-Umaps, allowing us to share interactive viewports of the image data online.
Allowing the less "tech-savvy" collaborators to explore the data conveniently. Given the activity on GitHub and the interests of external people, I am positive that this software will continue to grow.

As for now, I am mostly excited to spend time with my daughter and watch her grow.
8. Summary in Swedish


Här behövs det digital bildanalys för att automatiskt lokalisera alla ljusprickar, samt bestämma färgmönstret. Denna process kallas för avkodning, och här har vi ett av den här avhandlingens olika bidrag.

I Artikel I presenterar vi en metod där vi utnyttjar information om ljusprickarnas form samt förväntade färgmönster för att effektivt avkoda bilderna. Genom att formulera problemet som ett slags optimeringsproblem får vi en effektivare avkodning med avseende på mätningsbrus och överlappande prickar.

Avkodningen resulterar i punktmoln, där varje punkt visar var en viss typ av gen är uttryckt. Dessa punktmoln är väldigt stora. Här krävs det nya verktyg för såvåg visualisering som automatiskt analys.

I Artikel II presenterar vi ett beräkningsverktyg som automatiskt delar in vävnaden i regioner med liknande genuttryck. Metoden är beräkningsmässigt väldigt effektiv, vilket gör att man enkelt kan identifiera olika regioner på olika skalar.

I Artikel III har vi dessutom utvecklat en visualiseringsmjukvara för att enkelt kunna visualisera dessa stora punktmoln.

I Artikel IV presenterar vi ytterligare ett beräkningsverktyg för att identifiera individuella celler baserat på genuttrycket.

I Artikel V introducerar vi en analytisk variant av ett mått som används för att mäta, exempelvis, huruvida två gener ofta uttrycks tillsammans. Vanligtvis approximeras detta mått med stokastiska metoder. Fördelen med ett analytiskt mått är att det är deterministiskt och går mycket fortare att beräkna.


resultera i en behandling som får tumören att övergå till det mer fördelaktiga växtmönstret, och på så vis förbättrar överlevandsprognosen.

I den här studien har vi studerat tumörer som växer enligt det föredelaktiga växtmönstret. Med hjälp av in situ-sekvensering har vi undersökt genuttrycket i mellanskiktet mellan frisk lever och tumör. Nära tumören har vi hittat ett intressant genuttryck som vi tror är en respons på cancerceller som utbildar friska celler till cancerceller. I framtida projekt vill vi fortsätta att utforska processen bakom denna ”utbildning”.

Summerat innehåller den här avhandlingen flera verktyg, mjukvara och tillämpningar kopplade till in situ-sekvensering.
9. Acknowledgement

This PhD journey has been an incredible experience, and I feel truly blessed to have crossed paths with so many wonderful people.

First and foremost, I extend my heartfelt gratitude to Carolina Wählby for her exceptional guidance as my supervisor. In academia, individuals are often recognized for their intelligence, wisdom, and competence. While Carolina possesses all these qualities, her true strength as a supervisor lies in her kindness.

I also want to express my thanks to my co-supervisors, Anna Klemm and Filip Malmberg, who have consistently been there to offer their support.

A special acknowledgment goes to the fellow PhD students and colleagues who have been part of this journey: Kimmo Kartasalo, Hanqing Zhang, Nicolas Pielawski, Gabriele Partel, Anindya Gupta, Raphaela Heil, Håkan Wieslander, Erik Hallström, Elisabeth Wetzer, Nadezhda Koriakina (and Vangjush Komini), Karl Bengtsson Bernander, Jonas Windhager, Bozhi (Big Boz) Liu, Johan Öfverstedt, Leslie Solorzano, and Love Nordling.

I want to extend my special gratitude to Christophe Avenel and Fredrik Nysjö for our enjoyable discussions and your assistance with TissUUmaps.

Special thanks to Ingela Nyström for the exciting work with CBA. Also, thanks to Nataša Sladoje & Joakim Lindblad for the engaging discussions during modern and classical papers (and the nice collaboration with CBA).

I am also thankful to inspiring senior researchers and teachers: Robin Strand, Gunilla Borgefors, Ewert Bengtsson, Anders Hast, Orcun Göksel, and Ida-Maria Sindtorn.

I would also like to dedicate a heartfelt thanks to all my collaborators: Hanna Nyström, Olga Surova, Maria Escriva Conde, Peter Vermulen, Fred Hamprecht, Diego Ferran, Mats Nilsson, and Sergio Marco Salas.

A special shoutout to our PhD football team, Gupta FC. Winning our first game after 21 consecutive losses in the Uppsala Korpen league was a true moment of triumph. Siiuuu! Special thanks to my colleague Ankit Gupta for being an amazing friend (and goalkeeper). The 21 consecutive losses were not your fault.

Also, a big thank you to Can Deniz Bezek for all the interesting discussions regarding convex optimization and football and to Eduard "Eduroam" Chelebian and Andrea "Andreus" Behanova for the joy and laughter you brought into my life during these past years.

Thanks to the Larsson sisters for all the fun summer vacations. Special thanks to Konrad Steinvall. How long have we been friends? 20 years? It is weird how we always end up at the same place.
To my family — mom, dad, and my sister — thank you for always being there for me.

Lastly, my deepest and dearest thanks to Saga and our daughter Rut for the unconditional support and boundless love you both provide.
References


53


Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2371*

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)