Towards comprehensive cellular atlases

High-throughput cell mapping by *in situ* sequencing

Xiaoyan Qian
Towards comprehensive cellular atlases
High-throughput cell mapping by in situ sequencing

Xiaoyan Qian

Academic dissertation for the Degree of Doctor of Philosophy in Biochemistry at Stockholm University to be publicly defended on Monday 25 November 2019 at 09.30 in Air & Fire, SciLifeLab, Tomtebodavägen 23 A.

Abstract
With recent technological advancements in single-cell biology, many aspects of individual cells are characterized with unprecedented resolution and details. Cell types in human and model organisms are redefined, and multiple organ-wide atlases are proposed to integrate different types of data to provide a comprehensive view of biological systems at cellular resolution. Incorporating location information of cells in such atlases is crucial to understanding the structure and functions. Several spatially resolved transcriptomics technologies may serve this purpose, and in situ sequencing (ISS) is among the most powerful ones.

ISS detects the expression of tens to hundreds of genes in situ, i.e. inside preserved cells and tissues. ISS is a targeted approach, using probes designed to identify specific transcripts. Its key advantages, as compared to other spatially resolved gene expression analysis methods, are high throughput, cellular resolution and tissue compatibility, making it a tool ideally suited for spatial cell mapping. The work included in this thesis aims to develop tools and methods for this application.

In paper I, a network analysis tool was developed to analyze ISS and other spatially resolved data. The tool enables smooth visualization of large datasets and generates networks based on colocalization. It also includes functions to test statistical significance and resolve tissue heterogeneity.

In paper II, we studied spatio-temporal patterns of immune response in tuberculosis granuloma by targeting immune markers with ISS. Using the tool developed in paper I together with other methods, we established an immune response time course at the granuloma sites and found histologically different granulomas based on transcriptional information. The paper demonstrated that ISS can robustly detect transcripts in formalin-fixed paraffin-embedded tissues across biological samples and reveal biologically relevant structures.

In paper III, we developed probabilistic cell typing by in situ sequencing (pciSeq), a method to spatially map cell types defined by single-cell RNA-sequencing. pciSeq is an integrated pipeline that includes gene selection, image analysis, barcode calling and cell type calling. We mapped closely related interneuron cell types of the mouse hippocampal CA1 region in 14 coronal sections and validated the results against ground truth.

In paper IV, we investigated the quantification bias of ISS resulting from the probe target selection. We developed a method to sequence in situ synthesized cDNA and found that the read coverage of in situ cDNA library reflected ISS counts more closely than conventional RNA sequencing, making it possible, to some extent, to predict a probe’s performance and guide the probe design.

Taken together, the developments described in this thesis comprise several tools that make ISS suitable for building cellular atlases via large-scale spatial mapping.

Keywords: Spatially resolved transcriptomics, in situ sequencing, cell type, spatial analysis.

Stockholm 2019
http://urn.kb.se/resolve?urn=urn:nbn:se:diva-174755

ISBN 978-91-7797-884-8

Department of Biochemistry and Biophysics
Stockholm University, 106 91 Stockholm
TOWARDS COMPREHENSIVE CELLULAR ATLASES
Xiaoyan Qian
Towards comprehensive cellular atlases
High-throughput cell mapping by *in situ* sequencing

Xiaoyan Qian
"If a thing is worth doing, it is worth doing badly."

-G. K. Chesterton

To my parents.
List of publications

This thesis is based on the following papers:


* These authors contributed equally.

Reprints were made with permission from the publishers.
Related work by the author


* These authors contributed equally.
# Contents

**Introduction** ................................................................................................. 11

**Single-cell technologies for molecular characterization of cells...** 15  
Cytometry ........................................................................................................ 15  
Single-cell sequencing ............................................................................... 16  
Single-cell proteomics and multimodal single-cell assays ..................... 17  
Seeing is believing ...................................................................................... 18  

**Spatially resolved technologies for locating molecules in tissue.** 19  
*In situ* multiplexed RNA detection .......................................................... 19  
  Hybridization-based detection ................................................................. 19  
  Enzyme-mediated detection .................................................................. 21  
  Spatially indexed transcriptomics ......................................................... 22  
  Complementary features of different spatial technologies ................. 24  
  Spatially resolved proteomics and epigenomics .................................. 28  
  Error detection and correction in combinatorial decoding ............... 28  
  Beyond spatial location of molecules .................................................. 29  

**Towards comprehensive cellular atlases** ............................................. 33  
  Multi-omics approach ............................................................................. 33  
  Combining spatial and non-spatial methods ......................................... 34  
  Requirements for high-throughput spatial mapping tools .................. 36  
  Challenges in data integration ............................................................... 38  

**Present investigations** ........................................................................... 41  
  Paper I - Network visualization and analysis of spatially aware gene  
      expression data with InsituNet .......................................................... 41  
  Paper II - Spatial and temporal localization of immune transcripts  
      defines hallmarks and diversity in the tuberculosis granuloma ....... 43  
  Paper III - Probabilistic cell typing enables fine mapping of closely  
      related cell types *in situ* ................................................................. 44  
  Paper IV - Target sequence design of padlock probes based on  
      experimentally determined *in situ* synthesized cDNA fragments .... 45  

**Conclusions and perspectives** ............................................................... 46  

**Populärvetenskaplig sammanfattning på Svenska** ............................. 48  

**Acknowledgements** ............................................................................. 49  

**References** ............................................................................................. 52
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATAC</td>
<td>Assay for transposase-accessible chromatin</td>
</tr>
<tr>
<td>bDNA</td>
<td>Branched-DNA</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence activated cell sorting</td>
</tr>
<tr>
<td>FFPE</td>
<td>Formalin-fixed paraffin-embedded</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence <em>in situ</em> hybridization</td>
</tr>
<tr>
<td>FISSEQ</td>
<td>Fluorescence <em>in situ</em> sequencing</td>
</tr>
<tr>
<td>ISH</td>
<td><em>In situ</em> hybridization</td>
</tr>
<tr>
<td>ISS</td>
<td><em>In situ</em> sequencing</td>
</tr>
<tr>
<td>LCM</td>
<td>Laser capture microdissection</td>
</tr>
<tr>
<td>MERFISH</td>
<td>Multiplexed error-robust FISH</td>
</tr>
<tr>
<td>NGS</td>
<td>Next-generation sequencing</td>
</tr>
<tr>
<td>RCA</td>
<td>Rolling circle amplification</td>
</tr>
<tr>
<td>SBL</td>
<td>Sequencing by ligation</td>
</tr>
<tr>
<td>scRNA-seq</td>
<td>Single-cell RNA-sequencing</td>
</tr>
<tr>
<td>seqFISH</td>
<td>Sequential fluorescence <em>in situ</em> hybridization</td>
</tr>
<tr>
<td>smFISH</td>
<td>Single-molecule FISH</td>
</tr>
<tr>
<td>ST</td>
<td>Spatial Transcriptomics</td>
</tr>
</tbody>
</table>
Introduction

Multicellular organisms, with few exceptions, are composed of diverse cell types that organize into complex tissue structures and give rise to biological functions. Studying where cells are located and how a tissue is organized, and subsequently how an organ is organized is crucial to understanding complex biological systems.

In the 18th century, French anatomist Xavier Bichat described dozens of tissue types based on their reactivity to various treatments and anatomical observations without the use of microscope. He classified the tissues independently of the organs they formed. Although many works were later revised, the idea that an organ was composed of different tissues and that there were universal tissue types across organs had a fundamental impact on modern histology (Simmons, 2002).

In the 19th century, German pathologist Rudolf Virchow advanced the knowledge from tissue level to cellular level with the help of improved microscopes and proposed several groundbreaking theories. Among many was his significant contribution to the cell theory, expressed by himself as “all cells come from cells” (Omnis cellula e cellula) which described cell division. Virchow proposed to reduce all tissues to a single element – cell. It was revolutionary at the time, and based on this view Virchow founded the field of cellular pathology (Simmons, 2002).

Numerous important works carried out after Virchow, for example by neuroscientist Ramón y Cajal, showed there are many different cell types in a tissue, and that they do not function independently. It is therefore important in modern science to incorporate both aspects, individual cells and higher-level structures formed by cells, when studying healthy as well as diseased tissues.

In 2013, our group developed the first tissue-compatible, high-throughput, highly multiplexed in situ gene expression profiling technology that also has cellular resolution: in situ sequencing (ISS) (Ke et al., 2013). ISS was demonstrated in three fresh-frozen breast tumor tissue sections by targeting the RNA from 39 genes. The method enabled simultaneous visualization of all 39 targeted genes, displaying high level of tissue heterogeneity within a tumor, and allowed close examination of gene co-expression at cellular level (Figure 1).
Around the same time, single-cell RNA-sequencing (scRNA-seq) entered a rapid developmental phase with improved cell isolation and sequencing library preparation (Hebenstreit, 2012), followed by immediate generation of new knowledge about cellular heterogeneity. Since then, new cell types have been continuously discovered among cells previously assumed to be homogeneous, and complex biological processes can be unraveled by interrogating individual cells (Kolodziejczyk et al., 2015).

As striking levels of cellular heterogeneity is revealed by scRNA-seq, the question arises where these cells are located within a tissue and how they contribute to the complex functions. This is impossible to achieve using scRNA-seq alone due to loss of spatial information when isolating single cells. This instead led to an explosion of new methods in the field of spatially resolved transcriptomics and proteomics (Lein et al., 2017; Strell et al., 2019). Although only a few technologies achieve an unbiased coverage of the

**Figure 1. In situ sequencing (ISS).** (a) Schematic of ISS. cDNA is synthesized in situ, and padlock probes hybridize and ligate onto cDNA targets, followed by rolling circle amplification. Modified from Mignardi et al., 2017. (b) Barcode in a padlock probe is sequenced by rounds of sequencing by ligation (SBL) and fluorescence imaging. Each spot is decoded, and a matching gene target is assigned based on the sequencing read. In the box are several transcripts detected from the same gene. (c) A plot showing all genes targeted in Ke et al., 2013. Nuclear DAPI stain is shown in the background as grayscale image. (d) tSNE visualization of single cells in (c) and the location of cluster 9 and 14 in the tissue.
transcriptome or proteome as the majority apply a targeted approach, the term “transcriptomics” or “proteomics” are often used to describe these methods.

ISS shares many common features with other spatially resolved transcriptomic technologies, but also has its unique features, namely high throughput, high specificity, targeted approach and low detection efficiency. These features make ISS very attractive as a spatial cell mapping tool that brings insights into tissue organization and microenvironment.

Apart from scRNA-seq, other single-cell technologies were also developed to study the genome, epigenome and proteome of single cells. The ability to study complex tissues at cellular resolution not only helps to build cell type taxonomy, but also makes it possible to closely examine communication between different molecular layers of DNA, RNA and protein without being confounded by cellular heterogeneity, making multi-omics studies all the more attractive.

With the wide array of technologies available to date, integrating different molecular and computational tools, and creating synergy, will bring the most benefit to the broad scientific community. It also becomes a reachable goal to build comprehensive atlases from molecular and spatial information of cells. The work included in this thesis aims to develop tools that enable ISS to be applicable and contribute to large-scale spatial mapping efforts to build cell atlases.
Single-cell technologies for molecular characterization of cells

Unlike traditional morphological assessments, molecular profiling of cells offers an unbiased way to characterize cells, classify cells into cell types, and has the potential to reveal cellular functions. This was, however, difficult to achieve in individual cells because most assays require a certain amount of input material equivalent to thousands of cells. Hence, bulk measurement and comparison have been the most common strategy in marker discovery. Using this approach, many cell type- or disease-related markers have been discovered, often followed by visualization under the microscope or detailed characterization of marker-positive cells, with the final goal of deriving “pure” profiles.

The last decade witnessed rapid technological advancements that not only increased the number of parameters measured per biological sample, but also brought single-cell resolution to the assays with genome-wide coverage. This led to the description of high-resolution cellular profiles and gradual establishment of a cell type catalog (Regev et al., 2017). Molecular definition of cell types is key to locating various cell types because it can provide refined target lists that maximize the use of targeted in situ assays.

Cytometry

Flow cytometry and fluorescence-activated cell sorting (FACS) are widely used to characterize individual cells on properties like size, DNA/RNA content and the amount of protein. Modern instruments can be equipped with as many as 17 fluorescence channels (Perfetto et al., 2004), greatly increasing the resolution in separating cell populations. Imaging flow cytometry (Barteneva et al., 2012) incorporates fluorescence imaging into FACS and provides extra image-based information like cell morphology and intracellular localization of protein. Mass cytometry (instrument known as CyTOF) (Bandura et al., 2009) detects more proteins simultaneously by using antibodies coupled to isotopes instead of fluorophores and detected by mass spectrometry. Currently it can detect around 40 antibodies, limited by the availability of high-purity stable isotopes. Using this technology, as an example, researchers found previously unobserved continuum in hematopoiesis (Bendall et al., 2011) and could build a developmental
trajectory (Bendall et al., 2014), highlighting the power of highly multiplexed assays.

Although flow cytometry and mass cytometry can analyze single cells, their approach of targeting only tens of proteins fundamentally limits their use in the discovery of new cellular profiles and markers, which should ideally be achieved through non-targeted, unbiased approaches.

**Single-cell sequencing**

scRNA-seq detects mRNA molecules in a cell and provides quantitative measurement of gene expression. Currently, many cell isolation platforms and sequencing library preparation reagents are commercially available, enabling the wide application of scRNA-seq. The introduction and validation of single-nucleus RNA-sequencing (Bakken et al., 2018; Grindberg et al., 2013) further expanded the applications to fixed cells (Habib et al., 2016) and frozen human samples (Lake et al., 2016). The sensitivity of mRNA detection has improved over the years, with ~10% commonly achieved in the 3' end sequencing protocol (Zheng et al., 2017) and 30% for Smart-seq2 full-length library preparation (Picelli et al., 2013). Lately, as high as 50% sensitivity was reported from MATQ-seq (Sheng et al., 2017). The throughput capabilities also dramatically increased thanks to novel cell indexing systems, with data from two million cells recently reported in a single study (Cao et al., 2019). However, the cost of sequencing increases significantly with the number of cells profiled, forcing researchers to carefully balance between sequencing depth and sampling breadth. Several studies support that low-coverage sequencing is sufficient for finding cell types (Cao et al., 2019; Jaitin et al., 2014; Pollen et al., 2014) because of the intrinsic low dimensionality in gene expression data (Heimberg et al., 2016) (Figure 2), but is insufficient to identify all biologically relevant genes (Li et al., 2016) or isoforms (Arzalluz-Luque and Conesa, 2018). Apart from cell type discovery, scRNA-seq contributed significantly to the understanding of dynamic biological processes and gene regulation by providing untangled data (Kolodziejczyk et al., 2015).

Single-cell DNA sequencing provides an opportunity to dissect cell populations based on genetic alterations and construct evolutionary trees (Navin et al., 2011). However, it is more challenging than scRNA-seq in terms of assay sensitivity, sequencing depth and analysis, and thus has been less widely adopted. But, as more evidence supporting genetic diversity and mosaicism in healthy individuals emerges (Lodato et al., 2015), it is likely to gain popularity in the near future for studying developmental processes and aging (Zhang et al., 2019b). Besides, adaptation of many epigenetic assays
(bisulfite sequencing, ATAC-seq, etc.) to the single-cell format has enabled investigation of diverse epigenetic regulations with unprecedented resolution (Kelsey et al., 2017) and may provide more stable cell type profiles. More importantly, single-cell epigenomics has a great potential to bring new mechanistic insights into gene regulation, cell fate decision and epigenetic memory.

Just a few years ago, single-cell sequencing was considered too noisy and expensive to carry out routinely, but with continuous development it has become a common tool in biological studies and also frequently included in clinical studies (Keener, 2019).

**Single-cell proteomics and multimodal single-cell assays**

There is a long observed discrepancy between protein and RNA abundance, which can be attributed to technical as well as biological factors (Liu et al., 2016). This raises the question whether using the transcriptome can sufficiently approximate the proteome and phenotypes, and has led to many groups attempting to resolve the proteome with single-cell resolution. The most common approach so far is to use oligonucleotide to barcode antibodies.
The number of proteins measured simultaneously is not limited by readout labels as in flow and mass cytometry, but by antibody specificity. Proof-of-concept studies have also been shown for single-cell western-blot (Hughes et al., 2014) and mass spectrometry (Budnik et al., 2018), but these technologies are still in their infancy.

Multimodal single-cell assays simultaneously investigate more than one aspect of a single cell. Many of them provide one or two modalities in addition to scRNA-seq, partially thanks to the huge success of scRNA-seq technology, but mainly because of the bridging role of RNA between DNA and protein that makes it ideal to include. Single-cell genome-transcriptome assays can reveal aneuploidy and gene dosage effect (Dey et al., 2015; Macaulay et al., 2015). Epigenome-transcriptome assays are used to study gene regulation, but the sparsity of epigenomic data with even more decreased sequencing depth makes it extremely difficult to infer epigenome-transcriptome relationship in single cells, which is instead achieved by aggregating similar cells (Angermueller et al., 2016; Cao et al., 2018; Clark et al., 2018). Transcriptome-multiplexed protein assays are useful for studying dynamic processes but are still far from whole proteome coverage (Frei et al., 2016; Genshaft et al., 2016; Mimitou et al., 2019; Stoeckius et al., 2017). On the other hand, many assays can be easily combined with upstream FACS to add phenotypic information (Buenrostro et al., 2018; Chen et al., 2018a).

**Seeing is believing**

Microscopy imaging has always been the complementary tool to visualize and validate newly discovered markers. Single-cell resolution can be easily achieved with modern microscopes, and the throughput that imaging provides is several magnitudes higher than in vitro single-cell profiling. Moreover, tissue dissociation during single cell isolation often induces changes in cells and leads to skewness in profiles, while analyzing cells in their native context can efficiently avoid this (van den Brink et al., 2017).

In recent years, validating new cell types from scRNA-seq is typically achieved by detecting 3-4 proteins together using immunofluorescence or 3-4 RNA species using RNAscope assay (see next section). However, it is becoming clear that cell types are defined by a combination of markers and therefore methods are needed with the capability to detect more genes from the same samples.
Spatially resolved technologies for locating molecules in tissue

Understanding spatial distribution of cells is as important as understanding molecular profiles of cells since all complex functions in multicellular organisms arise from coordinated efforts from multiple cells, and tissue structures are often crucial to the functions. Many spatially resolved technologies, including ISS, seek to provide spatial information in addition to molecular information to help answer questions about tissue organization.

**In situ multiplexed RNA detection**

Technologies in this category determine the spatial location as well as RNA species using microscopy imaging. Most modern technologies are developed on the concept of single-molecule detection and color combinations and/or permutations.

**Hybridization-based detection**

Since the first demonstration of RNA fluorescent *in situ* hybridization (FISH) (Singer and Ward, 1982), FISH-based RNA detection methods evolved rapidly. Single-molecule RNA detection was shown by hybridizing multi-labeled detection oligos (Femino *et al.*, 1998), or several individually labeled oligos (Raj *et al.*, 2008) to RNA. This technique is now widely known as single-molecule FISH (smFISH).

DNA/DNA or DNA/RNA hybridization is reversible and sequence-specific, making it an extremely flexible system to control how RNA molecules are labeled, allowing color combinations (Levsky *et al.*, 2002; Lubeck and Cai, 2012) and multiple rounds of hybridization-stripping (Lubeck *et al.*, 2014). Via several rounds of multi-color imaging, coding capacity scales up rapidly. However, as more transcripts are visualized, optical crowding also becomes an issue. This is especially prominent in smFISH where the estimated detection efficiency is 95% (Femino *et al.*, 1998).

In recently developed smFISH methods, crowding is often tackled by the super-resolution philosophy: only a small subset of mRNA is “lit up” in each imaging round and extensive rounds of imaging are performed. Both MERFISH (multiplexed error-robust FISH) (Xia *et al.*, 2019a) ([Figure 3a](#)) and SeqFISH+ (Eng *et al.*, 2019) used this approach recently to increase the number of target genes to 10000 ([Table 1](#)). Another way to avoid crowding is
to simply use a non-combinatorial coding scheme, as in osmFISH (cyclic-ouroboros smFISH) (Codeluppi et al., 2018) (Figure 3b), which does not provide genome-wide expression but can quantify lowly abundant transcripts more accurately.

The abovementioned smFISH methods produce very weak fluorescence signals and require imaging with high magnification and long exposures. Two main hybridization-based methods have been developed to amplify signals in situ. One is the modification of branched-DNA (bDNA) signal amplification technology (Player et al., 2001) to have higher sequence specificity (commercial product known as RNAscope, ACD) (Wang et al., 2012). The other is based on hybridization chain reaction (HCR) (Choi et al., 2014). Previously, both suffered from poor signal removal and were not used in highly multiplexed FISH. But recently, SeqFISH (Figure 3c) demonstrated the use of HCR in multiple hybridization rounds by enzymatically removing DNA probes (Shah et al., 2016), and bDNA was also tested in MERFISH setting via fluorophore cleavage (Xia et al., 2019b). Nonetheless, the commercially available RNAscope assay is still limited to the number of fluorescence channels equipped in a microscope.

Figure 3. Hybridization-based in situ multiplexed RNA detection. Each panel illustrates the detection and coding scheme. (a) MERFISH uses combinatorial sparse labelling. (b) osmFISH uses non-combinatorial sequential hybridization. (c) seqFISH amplifies signal and uses combinatorial coding. Modified from Lein et al., 2017.
**Enzyme-mediated detection**

Although hybridization is much more efficient than enzymatic reactions, use of enzyme has its own advantages, namely high specificity, ability to query unknown sequences and lower demands for synthesizing complex probe pools.

Signal amplification using enzymes is easy to achieve and require minimal number of input oligonucleotides although it also results in higher variation of intensities (Xia *et al.*, 2019b). In smFISH, signal detection relies on multiple successful recognition events along a transcript that jointly create intensities above detection threshold and background noise. On the other hand, the majority of enzyme-mediated methods utilize circularization of linear probes as a pre-requrement for successful signal amplification. The circularization process is mediated by high-fidelity enzymes and makes it possible to generate specific signals based on short stretches of nucleotide.

ISS was the first technology to use enzymatic reactions to approach multiplexed spatial expression profiling. Built on an earlier technology based on padlock probes and rolling circle amplification (RCA) (Larsson *et al.*, 2010), ISS integrated sequencing-by-ligation (SBL) chemistry from next-generation sequencing (NGS) to readout short barcode sequences designed into probes. For each RNA species, as short as 30 to 40 nucleotides are enough to target and generate signals. In a padlock-based method, linear probes with target-binding sites at both ends are hybridized to in situ reverse transcribed cDNA and circularized by high-fidelity ligase, followed by RCA using the circularized probe as the template. Each RCA product is analogous to an amplicon in NGS technology, which can be sequenced in a similar way using a fluorescence microscope (Figure 1). The method was demonstrated in human breast cancer tissue sections targeting 39 genes (Ke *et al.*, 2013). Additionally, ISS can detect single-nucleotide variants and splice variants in situ (Kiflemariam *et al.*, 2014). The major drawback of the method is its low detection efficiency. A successful detection of signal requires all following events: cDNA synthesis, probe hybridization, probe ligation and RCA. Slight efficiency loss in any step decreases the observed transcript counts drastically. Therefore, in Paper III we developed a new probe design software to allow multiple probes to hybridize along a transcript to minimize the loss caused by failure of any step at a single target site.

A few new methods sought to work around reverse transcription, which is often considered the efficiency-limiting step. The newly characterized RNA-templated DNA ligase SplintR (Lohman *et al.*, 2014) opened up the opportunity to circularize probes directly on RNA target in situ (Schneider and Meier, 2017), but high false-positive rate (Krzywkowski and Nilsson, 2017)
resulting from unspecific ligation makes it difficult to use in a quantitative highly multiplexed gene expression profiling assay. Although the approach was attempted in a modified version of ISS – BOLORAMIS, no decoded result was shown in the preprint (Iyer et al., 2018). Notably, SplintR was proven very efficient in a low-plex non-digital setting, e.g. chromogenic readout (Jiang et al., 2019). A similar idea was pursued using “click chemistry” to circularize the probes followed by exponential amplification by more rounds of “click chemistry” (Rouhanifard et al., 2018). STARmap (spatially-resolved transcript amplicon readout mapping) (Wang et al., 2018) circumvents reverse transcription by identifying RNA targets via two independent hybridization events, and enzymatic ligation serves as a locking mechanism to preserve successful double-positive detection for subsequent amplification, followed by an improved version of SBL, SEDAL (sequencing with error-reduction by dynamic annealing and ligation) to readout the barcode.

Aiming for unbiased transcriptome-wide coverage, FISSEQ (fluorescent in situ sequencing) technology (Lee et al., 2014) was developed to amplify cDNA itself (circularized by intramolecular ligation) and use SOLiD sequencing chemistry (a chain-growing form of SBL to read longer sequence) on tissue sections. In FISSEQ, around 30 nucleotides are sequenced for each cDNA, and the reads are aligned to the genome or transcriptome as in conventional NGS. The method detected up to 8102 genes in human primary fibroblast cells, while no sequencing result was shown for tissue sections. Although cDNA synthesis was initiated by random priming, an uneven representation of RNA species was observed, which led to the authors’ conclusion that transcripts involved in RNA or protein processing are less accessible to FISSEQ and thus underrepresented.

**Spatially indexed transcriptomics**

Technologies based on in situ readout are superior in terms of spatial resolution but limited when it comes to discovering new sequence and structural variants. And they almost always require high-resolution imaging setup and are quite time/labor consuming. To overcome these, a range of technologies are developed in which spatial information is encoded into DNA sequence and sequencing reads are re-located to the tissue space using these “spatial barcodes”.

The mainstream technologies use pre-prepared slides to achieve this goal. Spatial transcriptomics (ST) (Ståhl et al., 2016) utilizes array-printed glass slides on which tissue sections are mounted and imaged with histological
staining. Each spot in the array has pre-determined spatial barcodes and can capture mRNA once tissue is lysed, followed by cDNA synthesis, release and sequencing by NGS. Reads are later assigned to spatial locations using the spatial barcodes. The main drawback is the difficulty to produce high-quality dense arrays. Currently, each array spot (diameter 100 µm) captures mRNA from around 30 cells and many cells are lost due to the uncovered area between spots (center-to-center distance 200 µm). Slide-seq (Rodrigues et al., 2019) overcomes the problem by using densely packed 10 µm beads. The random process of beads packing requires identifying spatial barcodes in situ after slide preparation prior to sample mounting. This is currently achieved by low-resolution imaging and SOLiD sequencing of 14-nt spatial barcodes, while the subsequent steps resemble ST. It has been shown to be able to resolve fine structures like a single-cell layer in a complex tissue. Newly developed high-definition ST (HDST) (Vickovic et al., 2019) pushes the resolution limit to 2 µm by depositing barcoded beads into dense hexagonal array wells. The spatial barcodes are readout in situ by multiple rounds of hybridization. However, the resulting data was sparse, and the authors later binned the data from 5 wells to “enhanced bins” (on average around 44 molecules detected) or to “cells” based on overlap between bins and nuclei.

A new approach, INSTA-seq (in situ transcriptome accessibility sequencing) (Fürth et al., 2019) uses transcript itself as a spatial barcode. The method is built on FISSEQ with several modifications followed by tissue lysis and NGS. INSTA-seq greatly shortens in situ imaging cycles by sequencing only 12 bases with a highly efficient SBL chemistry, termed PRICKLi (paired-end ribonucleotide-Inosine cleaved k-mer ligation). The short in situ sequencing length can be used for sequence alignment but mainly serves to barcode each transcript which is later identified in NGS. The method set out to answer the RNA accessibility question posed by FISSEQ by taking advantage of NGS but is still restrained by what is resolvable in situ.

Another strategy is to use precisely controlled light or ion beam to sample systematically over a tissue area. Commercially available Nanostring Digital Spatial Profiling technology (Merritt et al., 2019) uses photo-cleavable readout oligos and precise delivery of light to cleave, collect and quantify transcripts. Imaging CyTOF (Schulz et al., 2018) works in a similar way but uses isotope-labeled readout oligos and is again limited by the availability of isotopes. Notably, these methods can also detect protein if used jointly with antibodies carrying compatible labels. In some applications, the regions are preselected manually under the microscope to allow more targeted analysis (Decalf et al., 2019), similar to laser capture microdissection (LCM). Traditionally, researchers will manually outline cells based on their
morphism for LCM in counter-stained brightfield images, but it is becoming more common to do multi-channel fluorescence imaging, automatically classify and pick cells (Brasko et al., 2018).

**Complementary features of different spatial technologies**

So far, there is no one-size-fits-all solution when it comes to highly multiplexed spatial gene expression assays. Traditionally, smFISH-based methods are considered low-throughput and cannot detect many genes simultaneously although they have excellent detection efficiency, while enzyme-based methods are suited for high throughput but cannot achieve high sensitivity (Crosetto et al., 2015). But new studies are challenging this view, especially with many smFISH-based methods achieving throughput of thousands of cells and targeting tens of thousands of genes (see Table 1 for summary of recent literature reporting experimental data of highly multiplexed in situ assays). This of course comes at a price of longer imaging and turn-around time. For studying a small area of tissue (a few mm²), this might not pose any issue, but for some applications (e.g. human tissues) this can be a limiting factor. Another consideration is tissue preparation. ST, Slide-seq and HDST all require samples to be mounted to specially prepared slides, which limits their use in already prepared and archival material. MERFISH and SeqFISH+ requires tissue clearing or expansion, which may introduce distortions or RNA/protein loss. So far, only ISS and RNAscope demonstrated the compatibility with formalin-fixed paraffin-embedded (FFPE) materials (Carow et al., 2019; Kiflemariam et al., 2014; Wang et al., 2012), which makes it extremely attractive for cancer research. General technical comparisons between different technologies are summarized in (Strell et al., 2019).

Combining different spatial technologies, on the other hand, presents an opportunity to study many genes simultaneously at different scales and resolutions. A recent study combining ST and ISS (Chen et al., 2019) makes use of the transcriptome-wide coverage by ST to identify genes of interest near amyloid-beta plaque, followed by high-resolution detection of these genes using ISS, to reveal spatially restricted and coordinated responses of glia cells near plaques in the Alzheimer’s disease. ST/Slide-seq and ISS have similar throughput and therefore easy to apply to consecutive sections and take advantage of both methods, while HDST will not benefit as much thanks to its fine resolution and similar detection efficiency as ISS. In the newly announced 10x Visium system (commercial solution based on ST) cDNA is
kept on the slide instead of being cleaved off, which offers an interesting possibility to detect in situ anchored cDNA with any in situ readout methods after ST on the same slide. This will be ideal in case some features need to be “zoomed in” and also offers a way to “demultiplex” cell mixture in one ST array spot. In ISS, random cDNA priming and targeted padlock probe binding means the non-targeted cDNA molecules will be largely intact in situ, providing possibility to simultaneously detect lowly expressed genes with more sensitive methods like RNAscope or osmFISH. Combining with high-resolution methods like MERFISH and SeqFISH+ will also be useful as these methods are focusing more on inferring cellular dynamics from subcellular localization of RNA molecules. Generally, smFISH-based methods should be able to integrate more easily because RNA molecules are never degraded during the procedure and stripping of already bound oligos is not challenging. The idea of INSTA-seq can also be applied to ISS by introducing a random spatial barcode in padlock probes. By relatively sparse in situ labeling and spatial proximity, e.g. diffusion and ligation as in DNA microscopy (Weinstein et al., 2019), it is theoretically possible to resolve more molecules using NGS, which will address the limitations set by optical crowding.

What requires a lot more development is true 3D profiling methods. Until now, only RNAscope, STARmap and expansion FISH (ExFISH) (Chen et al., 2016a; Gross-Thebing et al., 2014; Wang et al., 2018) achieved true volumetric imaging in tissue. However, both RNAscope and ExFISH are limited to detecting 2-3 genes, while STARmap detected 28 genes in 150 µm-thick section.
Table 1. Data from recent publications in which highly multiplexed in situ RNA detection technologies were used to spatially locate cell types.

<table>
<thead>
<tr>
<th>Publication</th>
<th>Technology</th>
<th>Material</th>
<th># of genes</th>
<th># of cells</th>
<th>Detection efficiency</th>
<th>Readout error rate</th>
<th>Reads per cell</th>
<th>Microscope</th>
<th>Imaging†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xia et al., 2019a</td>
<td>MERFISH</td>
<td>U2-OS cells, expanded</td>
<td>10050</td>
<td>1368</td>
<td>80%</td>
<td>4%</td>
<td>92000 ± 32000</td>
<td>Widefield 60x</td>
<td>3c * 6z * 23r</td>
</tr>
<tr>
<td>Moffitt et al., 2018</td>
<td>MERFISH</td>
<td>Mouse hypothalamic preoptic region, cleared</td>
<td>155</td>
<td>~1 million</td>
<td>6-8x scRNA-seq</td>
<td>NA</td>
<td>Median ~200</td>
<td>Widefield 60x</td>
<td>2c * 7z * 8r + 2c * 1z * 11r</td>
</tr>
<tr>
<td>Eng et al., 2019</td>
<td>seqFISH+</td>
<td>Mouse brain, cleared</td>
<td>10000</td>
<td>2963</td>
<td>49%</td>
<td>0.22 ± 0.07/cell/bar code</td>
<td>5615 ± 3307</td>
<td>63x</td>
<td>3c * 2z * 80r</td>
</tr>
<tr>
<td>Shah et al., 2016</td>
<td>seqFISH</td>
<td>Mouse hippocampus &amp; cortex</td>
<td>125</td>
<td>14908</td>
<td>78.9%</td>
<td>4.6 ± 4.7 /cell</td>
<td>914.8 ± 570.5</td>
<td>Widefield ?x</td>
<td>5c * ?z * 4τ + 5c * ?z * 5τ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>249</td>
<td>2050</td>
<td>71%</td>
<td>NA</td>
<td>2807 ± 1660</td>
<td>Spinning disk confocal ?x</td>
<td>5c * ?z * 5τ + 5c * ?z * 7τ</td>
</tr>
<tr>
<td>Codeluppi et al., 2018</td>
<td>osmFISH</td>
<td>Mouse cortex, cleared</td>
<td>33</td>
<td>4839</td>
<td>~4x scRNA-seq</td>
<td>NA</td>
<td>435.5</td>
<td>Widefield, 100x</td>
<td>3c * 43z * 12r</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mouse cortex, cleared</td>
<td>160 (1020*)</td>
<td>3142</td>
<td>≥ scRNA-seq</td>
<td>1-4% after error rejection</td>
<td>Median 0.67</td>
<td>Scanning confocal 40x</td>
<td>4c * ?z * 6r</td>
</tr>
<tr>
<td>----------------</td>
<td>----------</td>
<td>-----------------------</td>
<td>-------------</td>
<td>------</td>
<td>-------------</td>
<td>---------------------------</td>
<td>--------------</td>
<td>----------------------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>Wang et al., 2018</strong></td>
<td>STARmap</td>
<td>Mouse cortex, cleared 150 um-thick</td>
<td>28</td>
<td>32845</td>
<td>NA</td>
<td>NA</td>
<td>No counting</td>
<td>Scanning confocal 25x</td>
<td>4c * ?z * 7r</td>
</tr>
<tr>
<td><strong>Paper III</strong></td>
<td>ISS</td>
<td>Mouse hippocampal CA1</td>
<td>99</td>
<td>27338</td>
<td>~0.2x scRNA-seq</td>
<td>NA</td>
<td>~20</td>
<td>Widefield 20x</td>
<td>5c * 7z * 5r + 2c * 7z * 1r</td>
</tr>
</tbody>
</table>

NA: not available. Note: if multiple datasets are presented in a study, the data collected from tissues are prioritized.

† The minimum amount of imaging required for barcode calling and error correction. c: channels, z: z-stacks, r: imaging rounds. Information that cannot be easily found in the corresponding publication is shown as question marks.
Spatially resolved proteomics and epigenomics

Highly multiplexed protein imaging to detect nearly 100 proteins has been demonstrated using iterative antibody incubation, imaging and bleaching/elution cycles (Gerdes et al., 2013; Gut et al., 2018; Schubert et al., 2006). Recently demonstrated CODEX (co-detection by indexing) (Goltsev et al., 2018) system uses oligo-tagged antibodies and the tags were identified by sequential hybridization and non-combinatorial labeling. In CODEX, all antibodies are incubated at the same time prior to oligo hybridization, greatly shortening the assay time. Imaging CyTOF (Angelo et al., 2014; Giesen et al., 2014) is another alternative to immunofluorescence, but requires specific instruments.

Few methods exist that provide spatially resolved epigenome information. ATAC-see (Chen et al., 2016b) visualizes open chromatin by preloading fluorescently labeled oligos to Tn5. The resolution is very low, but it reveals higher-order conformation of chromosomes that can reflect cell states. Intron SeqFISH (Shah et al., 2018) visualizes transcription active sites in the chromosomes by targeting only 5’ UTR of transcripts. Although the data is extremely noisy in inferring higher-order genome structures, it remains as an interesting approach to study relative localization of gene loci. A slightly modified version of MERFISH (Wang et al., 2016) is more powerful in perusing 3D chromosome structure, but questions arise whether hybridizing to DNA will alter the 3D structure.

Error detection and correction in combinatorial coding

In combinatorial coding, minor errors in each imaging round quickly accumulate and compound after just a few rounds of imaging. Methods having both in situ reads and “true” sequencing reads, e.g. FISSEQ and Slide-seq, generally suffer less from this problem since the match can still be found by having long enough reads and loosening criteria in read alignment. However, this poses a big problem for in situ-only technologies, because it can result in misidentification or detection loss.

Several error detection and correction strategies were presented in the last few years. Original ISS publication simply leaves enough unused barcodes to allow the assessment of error rate and only the genes with counts more than top unassigned barcode are accepted (Ke et al., 2013). MERFISH takes into account the different “on” and “off” rate in hybridization and leaves hamming
distance 4 between any barcode pair to allow error correction. The authors also targeted 10 times more genes with only error detection using a hamming distance of 2, while the percentage of barcodes detected with higher confidence ratio than unassigned barcodes dropped sharply from 91% to 73% (Chen et al., 2015). SeqFISH essentially takes the same approach as ISS but the fact that hybridization is more prone to “off” error makes it easier to perform error correction (Shah et al., 2016). STARmap inherits the advantage of SOLiD sequencing by interrogating the same base twice, which can efficiently reject sequencing errors, but in crowded and highly multiplexed detection, 40% of reads were still rejected (Wang et al., 2018).

Some decoding errors result from misalignment between imaging rounds and chromatic aberrations. Multi-channel imaging efficiently scales up the throughput and coding capacity but also brings chromatic aberrations and channel crosstalk. To address this, in Paper III, we developed a new integrated image analysis and decoding pipeline that utilizes point cloud to alleviate minor misalignment and chromatic aberrations. Since ISS leaves more than one base difference between genes in a barcoding scheme, which gives room for error correction, we used a barcode searching approach that assigns the best matching barcode to any read followed by quality filtering. With these improvements, ISS can detect more signals that enables cell type mapping without losing throughput.

**Beyond spatial location of molecules**

Detecting DNA, RNA or protein in situ is a step towards understanding their functional roles in the native environment. It is therefore a logic step to group them into the functional unit – a cell. However, identifying all molecules belonging to a cell based on an image remains a demanding task in tissue. Most technologies rely on DAPI staining to segment nuclei and a secondary staining to identify cell bodies if applicable (Figure 4), e.g. Nissl stain used in the brain tissue. A membrane stain is well suited to identifying cell boundaries (Goltsev et al., 2018) but, without z-resolved imaging or deconvolution, it provides ambiguous information regarding cell volume (Figure 4a). A cytoplasmic stain is less than ideal because it cannot resolve boundaries among tightly packed cells in a tissue (Figure 4c), except for tissues composed of similar, near-spherical cells where it’s considered less relevant (e.g. cancer). But in many applications (e.g. brain), the distance between cells is large enough this doesn’t pose any issue. Notably, many stains selectively label cells which can assist accurate segmentation of cells of interest but may also lead to undesired bias in quantification. Therefore, many
studies look for alternatives than conventional staining, including using autofluorescence and endogenous cellular features combined with machine learning (Sommer et al., 2011) or deep learning (Moen et al., 2019). Recent approach of using fluorescently labeled poly(T) oligos to detect cytoplasm on the premise that mRNA is spread throughout cell body has been proven very useful (Codeluppi et al., 2018; Moffitt et al., 2018). This is however difficult to achieve with low magnification imaging due to low signal-to-noise ratio (Figure 4b), and instead an extended concept based on reads density can be used (Figure 4e) (Shah et al., 2016). Another strategy, less explored, is based on the detected molecules to segment cells. It is conceivable that RNA or protein molecules should reflect the compartments they reside in, since a cell’s identity is defined by the coordinated activities of all these molecules and there should be little “spill-over” between different cell types. In an unsupervised approach, this could be theoretically achieved by identifying neighbor information of molecules and drawing lines between unmixable “neighborhood”. The concept is not new, but in practice the error is often too big to yield meaningful results. On the other hand, in a supervised approach, prior knowledge of co-expression can be used to improve the process, which was pursued with great success in Paper III.

Following cell segmentation, it is common to classify cells prior to further investigating the spatial arrangement of cells. Now that many technologies allow highly multiplexed detection and/or have excellent detection efficiency, unsupervised clustering is often used to identify in situ cell types. The results are, to some extent, in line with previous knowledge or single-cell defined cell types, but discrepancies are often observed (Cembrowski and Spruston, 2017; Shah et al., 2016), even in many cases where the targets are selected based on previous studies (Wang et al., 2018). It is frequently attributed to the differences in dynamic range and detection sensitivity, which complicates data interpretation. Nonetheless, the results on broad cell type level are generally in agreement between orthogonal methods (Chen et al., 2015; Codeluppi et al., 2018).

At the cellular level, the spatial pattern has been most frequently analyzed by looking at their immediate neighbors, e.g. finding “spatial domains” that consist of cells of the same type or of different types that repeatedly appear in the vicinity of each other (Zhu et al., 2018) and finding changes in cell-cell contact upon disease (Goltsev et al., 2018). Investigating cellular contact is important to understand tissue microenvironment, and this will provide valuable information about disease onset and progress, especially around lesions in damaged tissues.
On the other hand, in many applications, data of cellular resolution is unavailable and high-multiplex data is only available by aggregating multiple experiments, or simply the task of cell segmentation is overwhelmingly cumbersome. In such cases, coarse tissue structures can still be revealed by binning data and applying dimensionality reduction or clustering to find bins with similar profiles. By visualizing bins in their original dimension using colors corresponding to the reduced dimension, all major anatomical regions in the brain could be visualized (Mahfouz et al., 2015; Ortiz et al., 2019; Partel et al., 2019) and disease-relevant structures could be visualized in cancer (Svedlund et al., 2019). Another strategy is to identify non-randomly distributed genes and find regions corresponding to these genes. Early spatial statistics like Ripley’s K function can achieve this, but applying it to high-multiplex data becomes a tedious process and prone to false discoveries. In

**Figure 4.** Staining and cell segmentation. (a) Some antibody staining (shown Kv2.1 potassium channel) delineates cell soma boundaries clearly but fails to label all cells. (b) Labeling mature mRNA using fluorescently labeled poly(T) oligo is an alternative method to stain cytoplasm. The signals are often very weak and again only soma can be identified using this method. (c) Cytoplasmic staining (shown NeuN) in a dense area provides little help for segmentation. (d) DAPI segmentation and expansion from boundaries to mimic cells. Identified boundaries are shown on the right side, with decoded RNA molecules shown in different symbols. (e) Cell boundaries inferred from RNA molecule density. (f) Assignment of reads using prior knowledge. Gray lines connecting reads and cells indicate which cell a read belongs to. Shown in the figure is an O/LM cell to which many Sst reads are assigned correctly. The initial segmentation based on DAPI is shown in dotted gray lines.
addition, tissue structures make the theoretical random distribution meaningless and therefore randomized simulations are often required to devise accurate statistic (Mignardi et al., 2017). New methods have been developed to address some of the issues (Edsgärd et al., 2018; Rodrigues et al., 2019; Svensson et al., 2018) but the field largely remains unexplored. Coarse structures are rarely defined by a single gene and variations in expression level should not obscure the structure. With this in mind, in **Paper I**, we developed a tool for network analysis based on co-localization with a sliding window function, reasoning that stable gene-gene interaction pattern will consistently show up in network structures and thus can be identified. And in **Paper II**, we utilized this tool to find similar and dissimilar immune responses at tuberculosis granuloma sites at different time points in different mouse models.

Moving from spatial molecular detection to structural or functional inference is key to condensing information from massive amount of data points for multi-omics integration and comparison. It is heavily dependent on computational tools. Although novel analysis methods have been proposed, more need to be developed, with accompanying statistical tests to provide robust measuring of spatial patterns.
Towards comprehensive cellular atlases

With the development of new, robust, high-throughput technologies, large-scale studies have been conducted to construct atlases and databases, for example the Allen Brain Atlas with RNA ISH data from >20,000 genes in the mouse brain (Lein et al., 2007), Tabula Muris with scRNA-seq data from 20 mouse organs (Tabula Muris Consortium et al., 2018), and the Human Protein Atlas with ~17,000 antibody staining data across 32 human tissues (Uhlén et al., 2015). These resources are extremely valuable to the scientific community and often serve as a reference. However, most of them are limited to one type of molecular information and in order to cross-reference, a user must visit multiple databases and atlases and often struggles with different data formats before getting an overview. Now that many technologies are available at affordable cost and several consortia and initiatives are launched to coordinate the global scientific efforts, building comprehensive atlases is no longer a difficult goal to reach. Such atlases should contain molecular, functional as well as spatial data and integrated in a meaningful and user-friendly way to maximize the functionalities.

Multi-omics approach

Multi-modal single-cell sequencing is a powerful way to investigate the direct connections between several molecular layers of a single cell, but as previously discussed, the resulting data is often too sparse to draw conclusions at single-cell level. Therefore, combining data from individual modalities is a more viable approach. When integrating data, using cell types as anchor points to connect different modalities is the most straightforward way. Although there is no full consensus on cell types, some degree of agreement has been reached between independent studies and between different molecular modalities (Buenrostro et al., 2018; Welch et al., 2019). In this way, data types that traditionally cannot be directly linked, e.g. spatial distribution of cells that have specific DNA methylation pattern, can potentially be all “anchored” to scRNA-seq cell types, allowing the investigation of the relationship of the two indirectly linked data types.
Combining spatial and non-spatial methods

When combining spatial and non-spatial methods, researchers often treat them as orthogonal methods to validate each other’s results. But the distinct advantages from each of them make it more attractive to combine both.

In recent years, spatially mapping newly discovered cell types is of paramount interest. Computational reconstruction based on “landmark genes” determined from spatial profiling (e.g. RNA ISH, LCM) is feasible, but mostly limited to healthy tissue or simple structures and have relatively low spatial resolution (Achim et al., 2015; Moor et al., 2018; Satija et al., 2015). Some studies took the approach of clustering cells from scRNA-seq and cells from spatial profiling in parallel, followed by cell type comparison between the two and finding the best match (Codeluppi et al., 2018; Moffitt et al., 2018). The initial independent clustering allows discovery of potential new cell types, but finding the unique match between fine clusters is generally difficult. Other studies do not rely on parallel clustering but aim to assign individual cells profiled in situ to their best matching single-cell cell types. This has been demonstrated by binary combinations of a few carefully selected markers (e.g. gene A-positive & gene B-negative = cell type C) (Hodge et al., 2019; Tiklová et al., 2019), by finding the best correlating cell types (Soldatov et al., 2019), or using more advanced machine learning methods (Rodriques et al., 2019; Zhu et al., 2018). But none of the above has demonstrated the ability to spatially map closely related, fine cell types, largely due to semi-optimal gene panel design and skews in counts across methods. Although some spatial technologies have nearly transcriptome-wide coverage, they are limited in throughput and not suitable for large-scale cell mapping in tissue (Table 1), and therefore target selection is not a trivial step. In Paper III, we address this by a new computational approach in which gene selection is adjusted for differences in quantification arising from technical factors and optimized for a medium-sized panel (~100 genes) for ISS, and the assignment of a cell to scRNA-defined cell type is based on Bayesian inference. This enabled spatial mapping of closely related hippocampal CA1 interneuron subtypes in 14 coronal sections.

Many single-cell sequencing results have observed continua of cells in their transcriptomic profiles, contrary to the discrete profiles expected for individual cell types. This makes categorizing cells into cell types very challenging and often creates confusion between studies depending where the boundaries are drawn. One logical way to understand the continua is to see if a transcriptomic continuum corresponds to spatial continuum. Studies have verified the presence of spatial gene-expression gradient (no clear subdivision) within a cell type which may contribute to the functional...
heterogeneity (Cembrowski et al., 2016). A new model suggests that in a tissue where cells multitask, spatial position creates performance gradient in cell tasks due to differential access to nutrient and signals, which in turn creates gene-expression continuum (Adler et al., 2019). It is, however, unclear if the model holds true in a tissue composed of highly specialized cells. Hence, spatially mapping the cells observed in transcriptomic continua and examining the gradient genes provides an attractive solution to systematically test the hypothesis whether there is connection between the two and investigate the potential role of microenvironment in generating the subtle diversities. Continua are also observed during biological dynamic processes. Although both single-cell technologies and spatial technologies capture only a snapshot (unless genetic manipulations or perturbations are introduced early on), during dynamic processes, like cell differentiation, cells of continuous states and transition between them are captured altogether. This allows subsequent computational reconstruction of cell trajectory by aligning cells along the process based on transcriptomic or epigenomic profiles. Trajectory inference is becoming a popular tool to analyze biological processes, but few have mapped the process in situ and compared trajectory and cell migration or compartmentalization (Soldatov et al., 2019), which may provide insights into molecular/cellular events that lead to the commitment of a cell to a specific fate.

Non-spatial methods can provide detailed information about molecular events, but many processes are also chemically and physically restricted in vivo. For example, bulk and single-cell genome sequencing have brought great insights into clonal expansion and evolution of tumor, but the growth, dispersion and invasion of tumor cells are restricted by various barriers, and failure to take it into account makes it rather hard to reach a plausible explanation for tumor progression (Waclaw et al., 2015). Placing the cells in tissue context with genetic and transcriptional information and combining it with surrounding tissue structures will provide a holistic picture of disease progression. Furthermore, by combining it in the future with high-resolution imaging mass spectrometry, spatial distribution of metabolites and signal molecules can also be integrated to infer relationship between a cell’s environment and its phenotype.

In the long run, a consensus on healthy cell types across all organs may be reached (Regev et al., 2017) and it is foreseeable that corresponding cell markers (epigenetic, transcriptional or protein) will be developed. This will make targeted and image-based approach attractive again due to low cost, compatibility with multiple molecular types and ability to accurately find rare cells. It is yet unclear if most diseases are caused by the right cells in the wrong
place at the wrong time or by dramatic phenotypic changes of cells. The answer will get more clear as more diseased tissues are being analyzed in depth, but it is likely that some properly functioning cells are always involved (e.g. immune cells). An image-based screening using targeted approach therefore offers an alternative strategy to scale up single-cell sequencing and expand its use to less well-known systems.

**Requirements for high-throughput spatial mapping tools**

The goal of spatial mapping is to locate molecules, cells or structures of interest in large scale to assist tissue- or organ-wide systematic investigation. Scalability and robustness are key requirements for such tools. RNA ISH has been extremely successful in creating atlases, but it’s not fully quantitative and lacks single-cell resolution when comparing gene co-expression.

As more evidence supports that cell type identification is feasible with shallow sequencing (Figure 2), the same argument can be used for spatial cell mapping. High detection efficiency is desirable, but not necessary for cell type mapping. ST was recently used to build an adult mouse brain atlas that resolves finer anatomical regions than the Allen Brain Atlas (Ortiz et al., 2019). Remarkably, it was achieved using only 75 coronal sections of a single brain hemisphere with extremely low detection rate (on average 2-3 reads per gene was detected for ~4500 genes in each ST array spot), and the unsupervised clustering used only the top 50% variable genes. Hence, all clusters were presumably defined by the most strongly expressed region-specific genes, which also suggests genome-wide expression data is not necessary.

For cell mapping, a technology with cellular resolution is strongly preferred with no doubt. The resolution of ST is not enough to resolve single cells and therefore the annotation is often limited to anatomical regions. Slide-seq, on the other hand, has the potential since the size of beads is close to the size of a cell and there is no space between beads. By leveraging prior cell type information, it has been shown to be able to “demultiplex” a bead (Rodriques et al., 2019). In situ technologies generally outperform spatially barcoded technologies in terms of resolution but with lower throughput.

For mapping purposes, *de novo* identification of markers and cell types from *in situ* data is unnecessary. Based on the existing single-cell data, targeted approach with up to a few hundred genes should be enough, and as multiple studies demonstrated, gene panel size can be substantially decreased by selecting optimal targets (Figure 5). Among currently available
technologies, ISS, MERFISH, SeqFISH and STARmap all demonstrated spatial cell mapping using a few hundred genes, but the results from some of them are not in line with prior knowledge (Cembrowski and Spruston, 2017; Shah et al., 2016; Wang et al., 2018) highlighting the need to test the robustness of the methods. In Paper III, we addressed the issue by validating the ISS-based cell type mapping approach in previously well studied hippocampal CA1 area.

Another consideration is the ability of a technology to distinguish closely related cell types. This is generally more dependent on the detection specificity since similar cell types often differ by a small set of genes, but sensitivity may also play a role. Depending on how finely a mapping approach should distinguish cell types and how many diverse cell types it should cover, it can become unrealistic to aim to map all fine cell types in one experiment because of potential marker crosstalk. In this case, integrating multiple experiments, from e.g. RNAscope, and aligning them to a common coordinate framework is an alternative, although this will yield limited information about neighboring cells.

Figure 5. Gene panel size required for cell type identification. (a) MERFISH achieves nearly maximum performance with half of the genes it originally used. Modified from Moffitt et al., 2018. (b) SeqFISH takes random set of genes and the performance increases gradually with no obvious saturation. Modified from Shah et al., 2016. (c) Similar to MERFISH, pciSeq only requires roughly half of the genes to reach maximum performance. Modified from Paper III.
Challenges in data integration

Integrating data from multiple experimental batches, from multiple technological platforms, from multiple labs and even across molecular layers is essential to build comprehensive atlases, but it has proven not to be an easy task.

For data integration, several computational methods have been proposed to perform data normalization between experimental batches and technological platforms for single-cell data (Butler et al., 2018; Haghverdi et al., 2018; Lin et al., 2019), but there are still challenges remaining (Laehnemann et al., 2019). Recently, this has been extended to different data modalities (Stuart et al., 2019; Welch et al., 2019), which also integrates spatial data with non-spatial data, paving the way for multi-omics integration. Both of the methods (Stuart et al., 2019; Welch et al., 2019) efficiently aligned scRNA-seq data with STARmap data but also revealed the difference in scale and variation across technologies. This is not completely unexpected, and in Paper III, we also modeled the efficiency weight given to each gene to account for such difference between datasets. And in Paper IV, we asked more specifically whether we can improve the alignment between datasets by better designing probe target sites.

Another challenge is to ensure the robustness of cell types. Identifying cell types by unsupervised clustering is sensitive to data quality and to some extent clustering methods. It has therefore been proposed to take consensus cell types from multiple clustering results (Kiselev et al., 2017; Tasic et al., 2016). Yet between studies, there are often differences in the number of proposed cell types and profiles (Tasic et al., 2016; Zeisel et al., 2015). Often the discrepancy arises from different annotation of cell clusters. Some methods have been developed to identify highly similar clusters between data sets (Crow et al., 2018; Tasic et al., 2018), and supervised approaches were proposed to “BLAST” a cell type (Abdelaal et al., 2019). New methods were also developed to automate the cell type annotation procedure, hoping to increase the quality (Abdelaal et al., 2019; Zhang et al., 2019a).

Between spatial data, there has been no unified file format that allows easy data transfer and downstream analysis. Starfish aims to provide such standard by creating a modularized image analysis pipeline compatible with most image-based transcriptomics and proteomics methods (Perkel, 2019), but to date it requires a certain input format which requires pre-processing from data generators, which results in reluctance to adopt such a pipeline. However, with more researchers interested in importing the spatial technologies to their labs, changes need to happen in the near future for broad and quick adoption. Another outstanding issue in integrating spatial data is the lack of common
coordinate framework or the tool to align data to the framework. The neuroscience field has advanced quite far in setting up the framework and providing tools (Fürth et al., 2018; Kuan et al., 2015; Lein et al., 2007), but is much more difficult to achieve in tissues with less rigid and patterned structure. In order to efficiently align spatial data to a meaningful coordinate system, such framework needs to be developed soon. There have been 3D anatomical models built within individual projects (de Bakker et al., 2012), but a standard system that is widely accepted is necessary to set.
Present investigations

**Paper I - Network visualization and analysis of spatially aware gene expression data with InsituNet**

Many spatially resolved transcriptomic assays are developed but there are few intuitive and easy-to-use analysis tools available that visualize and find patterns in the data generated from such assays. In this study, we aimed to develop such a tool based on network analysis.

The tool was developed as an application of the existing network analysis software Cytoscape to enable full access to other analysis tools already developed and deposited in Cytoscape. It includes many functions for intuitive visualization and is optimized for speed and performance. The network construction is based on colocalization of transcripts. First, for each spot, all neighbors within a fixed distance are identified using given X and Y coordinates. Then using the neighbor information, a colocalization matrix is derived, and a network is built based on the colocalization information. In the network, each node represents a gene and each edge represents the colocalization between two genes. In visualization, the size of a node represents the frequency of the gene in the given dataset and the thickness of an edge represents how often the two genes colocalize.

In order to assess the statistical significance of edges, we implemented two ways of testing. One is based on label permutation repeated 1000 times, where the location of all spots is kept but the gene labels are shuffled. This preserves gene frequency and restrictions on spatial distribution of spots, but sensitive to the location of very lowly expressed genes. The other test implemented is hypergeometric test, which also takes into account the frequency of genes. After multi-test correction, edges considered statistically not significant are removed from the network. And the resulting network structure can be investigated using graph-based methods to look for structures like motifs.

One major drawback of such co-localization network is that it cannot reflect spatial heterogeneity. We implemented two methods to take into consideration tissue heterogeneity. One is manually drawing different regions of interest based on morphology and the other is a simple sliding window. For a given tissue, a sliding window with no overlap is created and network for each sliding window is generated. Then all networks can be used for comparison to find universal and unique connections.
In conclusion, we developed an analysis tool tailored towards spatially resolved transcriptomics. In addition to network analysis, the tool enables smooth visualization. We demonstrated the tool with already published breast cancer data (Ke et al., 2013) and could find and visualize different colocalization patterns.
Paper II - Spatial and temporal localization of immune transcripts defines hallmarks and diversity in the tuberculosis granuloma

In this study, we sought to use ISS and several developed analysis methods to study tuberculosis (TB) granuloma microenvironment. TB granuloma is a compact aggregate of immune cells. It is traditionally considered a mechanism that a host develops to “wall off” bacteria from spreading, but growing evidence shows bacteria can also trigger T cells to drive necrosis in granuloma to assist the spread of bacteria. Patient autopsies show huge diversity in granuloma histology, but the complete molecular mechanism is unclear.

We designed a panel of 34 immune markers and applied ISS on mouse lung tissues from three different time points after infection to study microenvironment changes in granuloma maturation. Using density-based screening, we visualized different immune cells recruited to TB granuloma as it matures, establishing an immune time course. By using unsupervised clustering on binned data, we found different spatial clusters and their corresponding molecular profiles and could match the profiles with previously described epithelioid and lymphoid granulomas. We then used the tool developed in Paper I to search for different colocalization patterns in each type of granuloma and found central regulators at different time points and different complexity of immune responses. We validated major findings with immunofluorescence.

We combined bacterial staining with ISS to specifically study immune markers present close to bacteria, and could conclude that activated macrophages colocalize with bacteria, probably encapsulating them. We also included another mouse strain in the study since it is known that single mouse model cannot faithfully represent immune response against TB in human.

In conclusion, we explored the potential of applying ISS to study heterogeneous structures like TB granuloma. Our spatio-temporal approach built the immune time course during granuloma maturation, revealed morphologically different epithelioid and lymphoid granulomas using transcriptional information only and characterized different immune components in them.
Paper III - Probabilistic cell typing enables fine mapping of closely related cell types

In this study, we aimed to specifically locate scRNA-seq defined cell types in tissue. We developed a probabilistic approach termed probabilistic cell typing by in situ sequencing (pciSeq) with the purpose of mapping closely related interneuron cell types in mouse hippocampal CA1 area. pciSeq includes gene selection, image analysis and cell typing in addition to ISS.

The gene selection algorithm finds a set of genes that can identify cells from scRNA-seq to their original cluster, taking into consideration the lower efficiency of ISS by down-sampling the scRNA-seq counts. Using this approach, we designed a panel of 99 genes including some manually selected ones to spatially map CA1 interneurons.

We developed a new probe design software and modified parts of ISS experimental procedure to achieve higher sensitivity with bigger gene panel. A new image analysis pipeline was also developed to better automate the procedure and include novel processing steps to efficiently deal with channel crosstalk, chromatic aberrations and minor misalignment across imaging rounds. Decoding was also changed to allow minor errors in reads. All these changes enabled higher reads density with more genes, without the need to use higher magnification, making full advantage of the high-throughput nature of ISS.

The cell typing is based on a probabilistic approach, which includes assignment of reads to cells, assignment of cells to cell types and estimation of gene-specific detection efficiency in each iteration until convergence. Assigning reads to cells based on gene co-expression in scRNA-seq data efficiently addresses the difficulty of accurate cell segmentation. And by adding an “unassigned” class, cell types with no marker designed to target them can be efficiently filtered out. We visualized the final results in a spatial map in which each cell is represented by a pie chart showing probability of belonging to individual cell type.

We then validated the method by first aggregating the 70 fine cell types into canonical “superclasses” and comparing the spatial location of each class to known ground truth.

In conclusion, we developed a new cell typing approach based on ISS, which allows high throughput mapping of closely related cell types defined by scRNA-seq. To our knowledge, this is the first demonstration of well validated spatial cell mapping method that can also resolve fine cell types.
Paper IV - Target sequence design of padlock probes based on experimentally determined in situ synthesized cDNA fragments

As a targeted approach, ISS target sites may be limited to the accessible parts of RNA. We observed extreme cases during study III when some probes targeting the extreme 5’-end failed to yield any signal due to the use of alternative transcription initiation sites in the brain. That observation prompted this study in which we aimed to understand the quantification bias and find transcript targets that are more readily accessible for padlock probes to bind.

We therefore developed a new method to investigate the available cDNA fragments after in situ reverse transcription, by releasing cDNA molecules from the tissue and sequencing. The results show generally very short cDNA molecules are synthesized, indicating the low efficiency of in situ reverse transcription. And when the reads are mapped to the genome, there are clear hot spots, indicating uneven representation of transcriptome by in situ synthesized cDNA. We then compared the in situ read counts observed in Paper III with both Smart-seq2 and cDNA-release library preparation and observed slightly higher correlation between in situ counts and the latter. This led to the reasoning that what is observed in ISS is limited to cDNA fragments and by targeting cDNA hot spots detected in the cDNA-release sequencing we could increase the detection success rate of padlock probes.

We then tested the idea with probes targeting different Actb exons in the mouse brain sections and observed generally good correlation but also noticed the insufficient predictive power of cDNA-release sequencing for in situ counts. But clearly, by avoiding target sites with no sequencing reads coverage, the chance of detecting a gene in situ is significantly higher. Although the question about quantification bias remains largely unanswered, the approach at least provides a way of designing probes for ISS which should result in lower dropout rate and thus ensure generally better correlation with in vitro methods.
Conclusions and perspectives

The overall aim of this thesis is to develop tools surrounding ISS to enable large-scale application of ISS for spatial mapping. The whole spatial field was still in its infancy when my PhD education started five years ago, with relatively simple way of processing images and very few analysis tools available for spatial data. In that context, I started developing various analysis tools and improving the molecular assay in parallel. Within the last five years, efforts from our group and other groups have pushed the field forward. From my own perspective, this can be summarized as a “zooming in” process. In the beginning, limited by detection efficiency and difficulty of segmenting cells in tumor samples, we were mostly analyzing the data at regional level, as demonstrated in Paper I and Paper II. As the ISS method improved, we were slowly able to zoom in to single cells. However, the targeted approach and moderate-sized gene panel used in ISS made it difficult to discover previously unknown co-expression patterns. We then saw a new opportunity as single-cell field was growing rapidly. We actively sought to use ISS as mapping tool to complement single-cell sequencing to provide the missing spatial information. We achieved the goal with success, as described in Paper III. At the same time, many questions appeared about the quantification of ISS. And therefore, in Paper IV we closely zoomed in to cDNA molecules to try to answer what determines ISS counts and how we can better design the assay to provide biologically relevant answers. With these improvements, ISS is gaining its popularity as a mapping tool and used in several collaborative projects. We are hoping to see fruitful results from these collaborations in the next few years.

There is still room for improvement in ISS. First is detection sensitivity. Many aspects of the protocol can still be improved. With improved sensitivity, ISS will have a strong advantage over hybridization-based methods thanks to its ability to distinguish single-nucleotide variants and isoforms. Second is 3D imaging. Combining ISS with gel embedding and thick section imaging has been proven tricky, but it is still an important aspect to achieve. Considering some enzyme-mediated methods have already been demonstrated in gel-embedded tissue, it should not pose an improbable challenge. Third is combination with antibody staining. Immunofluorescence has been tested previously after ISS, but the results largely depend on the stability of epitope. If antibody binding can precede reverse transcription, CODEX or similar system can be an alternative solution. Successful detection of lowly expressed proteins can also pose lower requirements for sensitivity of ISS.
The field of spatially resolved transcriptomics has developed rapidly during the last few years, but it is far from reaching the limit. Combined transcript and protein detection will probably be actively pursued, especially with subcellular localization, to answer the question of different functional roles and regulation of RNA and protein. BaristaSeq (Barcode in situ targeted sequencing) (Chen et al., 2018b) type of approach combined with lineage tracing may attract a lot of attention as many single-cell studies are moving forward to study developmental progress and disease. Spatially mapping genetic clones is still of interest in cancer research, and it may also be used to identify somatic mutations when studying aging or disease-prone mutation accumulation processes, but almost no existing technology can achieve this goal with high sensitivity and confidence in mutation calling. Profiling in thick sections combined with 3D imaging has shown some preliminary proofs-of-concept, but a lot more need to be done to increase the number of genes and eventually achieving whole-mount throughput.

ISS and padlock probe-based assays are generally very flexible. With continuous innovations the method has great potential to gain higher sensitivity and can be used in a wide range of applications in conjunction with other technologies.
Populärvetenskaplig sammanfattning på Svenska

Flercelliga organismer som människan består av många olika celltyper. Med nya tekniker är det nu möjligt att studera enskilda cellers molekylära innehåll. Detta har lett till att man har hittat nya celltyper och att traditionellt klassificerade celltyper kan omdetermineras till molekylärt definierade celltyper. För att förstå komplexa biologiska processer på organnivå så är det viktigt att utöver celltypsklassificering se hur cellerna är organiserade i organen genom att lokalisera dessa celler i vävnader. Processen att lokalisera och kartlägga celler i vävnader kallas för cellkartläggning. ”In situ sequencing” (ISS) är en teknik med stor potential att möjliggör detta på ett effektivt sätt. Detta eftersom ISS detekterar RNA molekyler från hundratals gener samtidigt direkt i vävnaden (”in situ”) med så kallade ”padlock probes”.

I denna avhandling beskriver jag utvecklingen av olika analysmetoder och tillämpningar av ISS vid cellkartläggning. I samarbete med andra forskare har vi utvecklat ett nätverksanalysverktyg som hittar spatiale mönster baserat på samlokalisering. Detta verktyg användes sedan för att studera immunsvar på tuberkulosgranulom som bekräftade tidigare studie samt lokalisering av histologiskt och molekylärt distinkta granulom. Vidare, för att identifiera exakta celltyper i vävnader, förbättrade vi den molekylära metoden och utvecklade en ny analysmetod för att bestämma sannolikheten för att en cell är av en viss celltyp. Metoden användes och verifierades i hippocampus CA1-regionen i mushjärnan, och tillämpades sedan för kartering av hjärnbarken. I den sista studien, utvecklad vi en ny metod för att förbättra designen av ”padlock probes” vilket riktar in sig på RNA-molekyler, i hopp om att förbättra analyskänsligheten av ISS.

Med dessa nya framsteg, har ISS visat förmågan att kartlägga celltyper med hög kapacitet. Detta kan komma till stor användning inom ett stort antal biologiska och medicinska fält inom en snar framtid.

48
Acknowledgements

This thesis would not have been possible without the scientific contribution and support from many other people. I would like to take the opportunity and express my gratitude.

First of all, Mats, thank you for offering me the opportunity to do my PhD in your lab. It was a truly amazing journey. I would have never thought I would achieve this much, knowing how easily I get diverted. Your encouragement, optimism, and rare push were quite crucial, I believe. Thank you very much for also putting so much trust in me and giving me the freedom to pursue my scientific interests.

Carolina and several members of Wählby group: Petter, Gabriele, Leslie and Maxime, thank you all for the in-depth discussion about image analysis and pattern recognition. It’s a pity that we didn’t collaborate more, but all the knowledge that I gained from you during my first two years of PhD laid a solid foundation for many of my works later.

Kenneth, I would like to thank you for your enthusiasm and patience apart from your scientific insights. The cell typing paper will not exist without you, and I learned so much from you, in person and from your code. It still amazes me that you as a PI, in London, were debugging my code at 10 pm via skype. It’s good to see finally our paper got accepted!

To all my collaborators, thank you so much - John and David for the nice work put into making InSituNet, Berit and Thomas for believing in ISS and troubleshooting every step to make ISS work for TB-infected lung samples, Dimitris for developing visualization tools for the cell typing application and for being always available when I had questions, Nathan and Jens for the valuable feedback on the cell typing project during early days, Dan for the great work put into detecting influenza genome segments, Elin for the inspiring discussions we had while finishing the sexual dimorphism project, Jessica and Carina for all the interesting and innovative ideas that came up during various oncology-related projects, Metka for years of hard work to detect Ephrin receptor mutations, Chenglin for the great effort in generating data for the heart development project, Stefania and Michaela for explorative work combining several spatial technologies, and last but not least Di for inventing the cDNA sequencing method.

Special thanks to Daniel and Markus. Thank you for your precious feedback on the thesis and sorry for imposing the horrible first draft on you. Also, big thanks to Malte and Pia for feedback, to Amitha, Nicolas and Daniel for thorough proofreading and to Christoffer for correcting (aka rewriting) the Swedish summary. Thank you very much!
To the members of Nilsson group, thank you so much for creating such an enjoyable work atmosphere. Chika, thank you for being the “mom” of the group and taking care of many things for years, ranging from buffers to snacks (equally important!). Amita, your positivity and composure are so catching that sometimes when I’m stressed I just look at you to calm down. Di, I learned from you that there is no boundary in scientific imagination. I wish we could have pursued some ideas a bit further and developed new cool tools. Felix, you and your fabulous aura literally lights up the office and thanks for providing delicious ca-bu-qiu-nuo. Kamila, thank you for the nice afterwork badminton, and appropriate and inappropriate jokes. Nara, you are such an energetic person and a living encyclopedia. Anastasia, big thanks for streaming the famous TV show. It was a lot of fun watching with you guys. Jessica, you are always so positive and full of energy! Thank you for all the patience towards my forgetfulness and for nice discussions about ISS applications. Olga, thanks for lots of good recommendations and your meticulous attitude towards research has been very inspiring to me. Erik, thank you for great discussions about literature, culture and history; it has been truly educational. Daniel, big thanks for always providing what I ask for despite last-minute notification. Elin, thank you so much for sharing the excitement, satisfaction as well as disappointment and frustration in purusing PhD degree. Your engagement in the broader scientific community and in the society has always amazed me and reminds of our duty beyond day-to-day lab work. Markus, you are simply an amazing person, although not always easy to reach when I need you. Thanks for numerous discussions about technology and biology and your insights about neuroscience. Flor, thanks for bringing the vibes into the group that are as colorful and ever-changing as your hair color. And the new talents, Christoffer, Sergio, Hower and Milda, I feel so much younger and cooler with you guys around. Thank you for the good beer time that we had together and I’m looking forward to the scientific breakthroughs that you will bring forth!

To the past members of Nilsson group, Rongqin, Camilla, Marco, Anja, Anna, David, Annika, Thomas, Pavan, JP, Eva, Anne, Malte, Tom, Iván, Sibel, Carina, Navya, Chenglin, and our honorary member Tim, you all had some influence on me, one way or another. Some of you gave me the entry course to the in situ field, some shared your PhD survival strategies and work-life philosophies, some introduced me to new sports, and so on and so on. It was a pleasure to work with you and I feel extremely lucky to have friends like you. I’m also very excited that I can continue to work with some of you in the future. My two students, Hanyu and Dave, thank you for your hard work and believing in my supervision although I was often busy or just impatient.

And my colleagues at Cartana, Jana, Johanna, Toon, Anton, Iván, Johan, Malte, Francesca, Mats B, Mårten, Alexia, thanks a lot for your support and understanding. It has been very exciting and eye-opening to work with all of you. Thank you for believing in our technology, especially Malte and
Francesca, for devoting yourselves to the cause despite many doubts. I have so much more to share with you and I’m looking forward to working fulltime with you all soon.

My dear friends Haixia, Fan, Xin, Jenny, Annika, Magda, Dan, Hao, Xiang, Yue, Yilin, David, Pontus, Kostas, Xia, Wenjiang, Zheng and Linus, thank you for all the good times we spent together and occasional scientific inputs over the years. I wouldn’t have enjoyed my PhD education in Sweden as much without you.

Nicolas, words cannot express my gratitude for your support and company over the last few years. You believed in me when I doubted myself. Thank you so much for your encouragements during difficult times and always bringing out the positive side of me. Danièle and Éric, thank you for your hospitality whenever I visited you and for the remote support that you constantly sent me. Merci beaucoup!

My family, 엄마 아빠, 지금까지 절 믿어 주셔서 감사합니다. 점에 단하나인 밀로서 멀리 떠나 외국에서 생활하는데 미안한 것도 있지만 그동안 계속 지원해 주셔서 고맙습니다. 비록 멀리 떨어져서는 있지만 엄마 아빠 사랑하는 마음은 계속 그대로입니다. 박사공부를 끝 맞힌 후 앞으로도 계속 엄마 아빠의 자랑스러운 딸이 되고 싶습니다.
References


Schneider, N., and Meier, M. (2017). Efficient in situ detection of mRNAs using the 
Chlorella virus DNA ligase for padlock probe ligation. RNA N. Y. N 23, 250–256. 
Schubert, W., Bonnkeoh, B., Pommer, A.J., Philipson, L., Böckelmann, R., Malykh, 
Analyzing proteome topology and function by automated multidimensional 
Schulz, D., Zanotelli, V.R.T., Fischer, J.R., Schapiro, D., Engler, S., Lun, X.-K., 
Jackson, H.W., and Bodenmiller, B. (2018). Simultaneous Multiplexed Imaging of 
mRNA and Proteins with Subcellular Resolution in Breast Cancer Tissue Samples 
of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. 
Neuron 92, 342–357. 
Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.-H.L., Koulena, N., 
the Nascent Transcriptome by Intron seqFISH. Cell 174, 363-376.e16. 
Sheng, K., Cao, W., Niu, Y., Deng, Q., and Zong, C. (2017). Effective detection of 
variation in single-cell transcriptomes using MATQ-seq. Nat. Methods 14, 267– 
270. 
Simmons, J. (2002). Doctors and Discoveries: Lives that Created Today’s Medicine 
(Houghton Mifflin Harcourt). 
muscle tissue culture by using in situ hybridization with a biotinated nucleotide 
Soldatov, R., Kaucka, M., Kastriti, M.E., Petersen, J., Chontorotzea, T., Englmaier, 
Spatiotemporal structure of cell fate decisions in murine neural crest. Science 364, 
eaa9536. 
Interactive learning and segmentation toolkit. In 2011 IEEE International 
Stähl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., 
and analysis of gene expression in tissue sections by spatial transcriptomics. 
Science 353, 78–82. 
Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, 
Strell, C., Hilscher, M.M., Laxman, N., Svedlund, J., Wu, C., Yokota, C., and 
Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 
Svedlund, J., Strell, C., Qian, X., Zilkens, K.J.C., Tobin, N.P., Bergh, J., Sieuwerts, 
to spatially resolve gene expression profiles of diagnostic and prognostic markers 


