Universiteit Leiden
The Netherlands

Integration and disentanglement of single-cell and spatial transcriptomics in health and disease

Novella Rausell, C.

Citation

Novella Rausell, C. (2025, May 28). *Integration and disentanglement of single-cell and spatial transcriptomics in health and disease*. Retrieved from https://hdl.handle.net/1887/4247894

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4247894

Note: To cite this publication please use the final published version (if applicable).

1

INTRODUCTION

UNDERSTANDING health and disease is the main focus of modern medicine. Early practitioners studied disease by comparing sick patients to healthy individuals and as early as 1753 controlled clinical trials were performed to treat scurvy[1]. However, treatments were based on observation and intervention rather than on a mechanistic understanding of disease. It was not until 1949 that the first molecular basis of disease was described by Linus Pauling and his colleagues after discovering that the hemoglobin from individuals with sickle-cell anemia was different at the molecular level from those of healthy individuals[2]. Shortly after, Francis Crick and James Watson gave an accurate description of the DNA molecule and its double-helical structure[3], laying the foundations of molecular biology and offering a blueprint that explained how information is stored and transmitted across generations.

Today we know that every cell in a living organism contains DNA. This molecule stores heritable information (i.e., genes) on function and survival scattered around 3.1 billion base pairs in the genome (i.e., the DNA makeup of an organism). Using the simplest view of the central dogma of molecular biology, a gene is transcribed into messenger RNAs (mRNAs) and these in turn are translated into proteins, which carry specific functions in every cell[4]. This is a tightly regulated process, and errors in any of the steps or components can lead to a cascade of events that ultimately leads to the loss of homeostasis in a cell. Eventually, such cells can no longer perform their physiological function and adapt to environmental cues, which causes disease.

In a healthy state, gene transcription (i.e., gene expression) is intricately regulated to meet the physiological demands of the cell environment. A neuron and an epithelial cell in the kidney contain the same set of genes, but due to transcriptional regulation only a fraction of these will be transcribed in each respectively[5] (**Figure 1**). Some genes, such as ATP1B, which encodes an ion channel, are constitutively expressed across various organs. While this gene is essential for basic cellular functions, it shows increased expression in specific organs like the kidney where its role in ion transport and fluid balance is crucial. These gene regulatory processes are influenced not only by intrinsic molecular mechanisms but also by extrinsic factors such as nutrients, hormones, and environmental conditions, ensuring that gene expression is finely tuned[6–8]. This results in a variability in gene expression that extends beyond tissues and organs, as cells in close proximity to another can uniquely express certain genes or show different levels of expression, specialising in distinct functions[9].

Similarly, in disease states, gene expression can be disrupted by intrinsic factors like genetic mutations (i.e., changes in the sequence of the DNA) or by extrinsic factors such as viral infections or chemical exposures. These disruptions can alter the normal regulatory processes, leading to the abnormal transcription of genes critical for maintaining cellular homeostasis. In cancer for example, changes in the DNA can lead to overexpression of certain genes which eventually causes cells to proliferate and overgrow, producing tumours[10,11]. Most of the time, however, changes in gene expression are not the primary cause of disease but serve as biomarkers that reflect various stages, predispositions, or the progression of a disease[12]. Therefore, establishing a connection between gene expression patterns and specific disease phenotypes remains essential for understanding the underlying mechanisms and potential therapeutic targets.

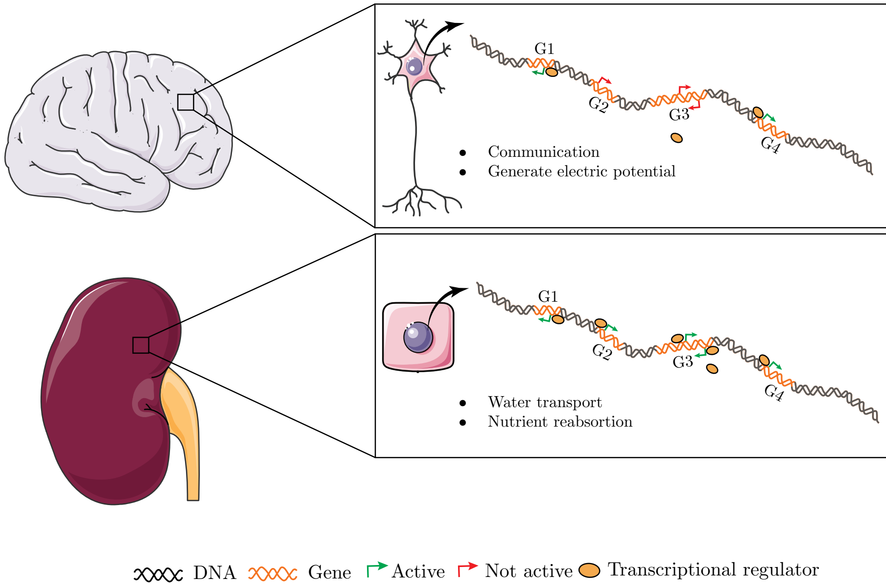

Figure 1.1: *Transcriptional regulation during homeostasis*. Given the same set of genes (G1, G2, G3 and G4), a cell in the kidney and a neuron carry different functions due to transcriptional regulation.

1.1 AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is a genetic disorder characterised by the formation of fluid-filled cysts in the kidneys, with a prevalence of approximately 1 in 1000 individuals worldwide [13]. These cysts progressively expand, impairing normal renal structure and function, ultimately leading to end-stage renal disease (ESRD). The primary genetic causes of ADPKD are mutations in the PKD1 or PKD2 genes, which encode Polycystin 1 (PC1) and Polycystin 2 (PC2) respectively [14]. These proteins play crucial roles in various cellular processes, including cell-cell and cell-matrix interactions, cell proliferation, and fluid secretion [15,16].

Cyst development in ADPKD is a complex process involving multiple cellular and molecular mechanisms. Key pathological features include aberrant cell proliferation, increased fluid secretion, extracellular matrix (ECM) remodelling, chronic inflammation, and dysregulated cellular metabolism [17]. Changes in the composition and organisation of the surrounding tissue microenvironment, including extracellular matrix remodelling, inflammatory cell infiltration, and secreted factors, further contribute to cyst expansion and fibrosis in the adjacent tissue [18]. The cyst microenvironment is a critical driver of ADPKD progression. As cysts enlarge, they impose mechanical stress on the surrounding tissue, which can

initiate injury responses that, in turn, stimulate further cyst development[19]. This feedback loop establishes a self-sustaining cycle of cyst expansion and tissue damage. Additionally, the microenvironment is enriched with inflammatory cells and cytokines, which further contribute to disease progression[20].

Understanding the complex interplay of molecular and cellular processes in ADPKD is crucial for developing effective therapeutic strategies. Current research efforts are focused on elucidating the mechanisms of cyst formation and growth, identifying key signalling pathways involved in disease progression.

1.2 Challenges to be addressed by single-cell and spatial techniques in ADPKD

Recent advances in next-generation sequencing (NGS) technologies have equipped researchers with powerful tools to analyse the genetic and transcriptional landscapes of both healthy and diseased tissues, including ADPKD. Initially, bulk RNA sequencing (bulk RNA-seq) of kidney tissue samples provided valuable insights into the overall transcriptional changes associated with ADPKD[21]. However, bulk RNA-seq has limitations in its ability to capture the cellular heterogeneity within ADPKD kidneys, as it provides an average gene expression profile across all cells in a sample, potentially masking important cell-type specific changes.

To overcome these limitations, single-cell mRNA sequencing (scRNA-seq) has emerged as a powerful tool, greatly deepening our understanding of functional heterogeneity in ADPKD tissues. scRNA-seq enables the grouping of cells into physiologically relevant types based on their gene expression profiles, revealing the diverse cellular populations present in ADPKD kidneys[22]. Recent single-cell studies in kidney injury models have identified Failed Repair Proximal Tubules (FR-PT)[23], cells that escape repair after injury and may contribute to chronic kidney disease progression. Moreover, scRNA-seq revealed the same FR-PT cell type in ADPKD patients, potentially contributing to cyst growth[24]. Understanding the emergence and behaviour of FR-PT-like cells in the context of ADPKD could provide crucial insights into the mechanisms of cyst formation and disease progression. However, while scRNA-seq offers high cellular resolution, it lacks spatial context, limiting our understanding of how these cells interact with their microenvironment—a critical aspect in ADPKD where spatial organization of cells in cystic and non-cystic regions profoundly influences disease outcomes.

To address this gap, researchers have started integrating scRNA-seq data with spatial transcriptomics (ST) techniques, which allow for the study of gene expression while preserving the spatial relationships between cells[25,26]. This integrated approach is particularly beneficial for investigating FR-PT-like cells in ADPKD, as it enables a deeper understanding of how these cells are distributed relative to cystic structures and their interactions with neighbouring cell types (**Figure 2**). For instance, exploring the spatial distribution of FR-PT

cells in relation to cysts and examining their communication patterns could reveal critical aspects of ADPKD pathogenesis.

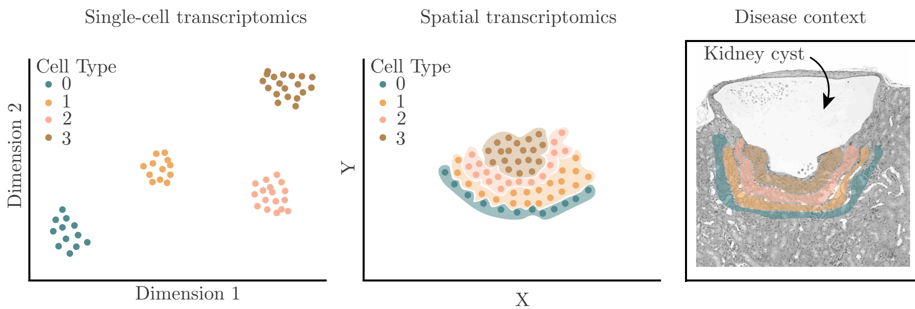

Figure 1.2: *Cell localisation in disease*. Cells are coloured by cell type. Dimension 1 and Dimension 2 correspond to an arbitrary 2-dimensional representation in which similar profiles are close by in space. Shadows illustrate the mapping between identified disease-associated niches and the tissue slide. Tissue slide courtesy of the Division of Translational Toxicology, National Institute of Environmental Health Sciences

Cell-cell communication is regulated through interactions between cell surface receptors and ligands—molecules secreted by neighbouring or distant cells—which activate signalling pathways that shape cellular behaviour and fate[27]. Analysing these spatially resolved interactions is crucial for understanding how local tissue environments modulate cellular responses, particularly in the context of ADPKD, where the proximity of cells to expanding cysts may influence their transition into a failed repair state. Applying ST techniques, researchers can map these communication events in the cyst microenvironment, which may elucidate how signalling networks contribute to the maintenance of cystic growth and the emergence of FR-PT-like cells.

Single-cell and spatial approaches can address the challenge of understanding the temporal dynamics of ADPKD progression. By analysing tissues at different stages of the disease, researchers could track how cellular states evolve over time, revealing key transition points in disease progression and opportunities for early intervention. This includes monitoring the emergence and behaviour of FR-PT-like populations, their spatial distribution relative to cysts, and their interactions with surrounding cells. Such analyses promise to uncover critical insights into the initiation and progression of ADPKD, potentially identifying new targets for therapeutic intervention.

1.3 Representation learning

The advent of single-cell and spatial transcriptomics technologies has generated vast amounts of high-dimensional data. A typical study might quantify around 25,000 genes across thousands of cells or spatial locations, resulting in hundreds of millions of measurements[28]. This high dimensionality, coupled with technical variability presents significant challenges

for data analysis and interpretation.

To address these challenges, dimensionality reduction techniques have become essential in the computational analysis pipeline. These methods aim to summarise the observed variability into a lower-dimensional representation while preserving the most relevant biological information. One common approach is to focus on highly variable genes, which are likely to contain the most informative biological signals[29] (**Figure 3**). This step reduces the data to a more manageable set of 2,000-4,000 genes. Representation learning[30] extends this concept by using machine learning algorithms to find optimal low-dimensional representations of the data. These learned representations can capture complex, non-linear relationships and facilitate visualization and clustering[31–33]. Additionally, these can also enable advanced analyses such as learning cell type identities or inferring developmental trajectories[34,35] (**Figure 3**).

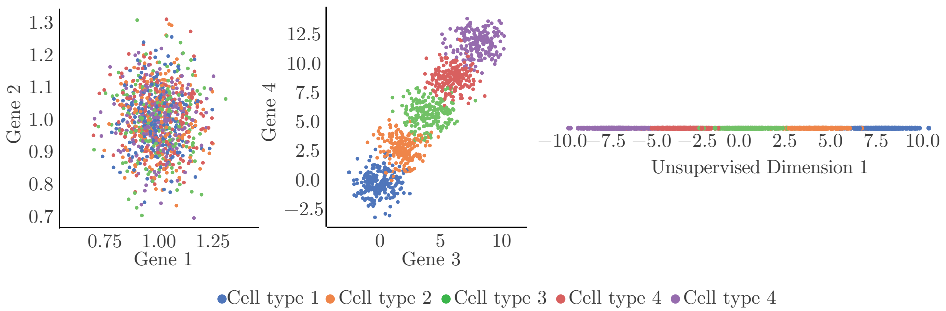

Figure 1.3: *Highly variable genes and dimensionality reduction.* Dots represent individual cells coloured by cell type. Genes 1 and 2 are housekeeping genes. Genes 3 and 4 are highly variable. Axes represent normalised gene expression. After keeping Genes 3 and 4, the dimensionality is further reduced to a 1-dimensional space with representation learning techniques.

In addition, representation learning techniques can be designed to address specific challenges in single-cell and spatial transcriptomics, such as batch effects. These are systematic variations between datasets that are not due to the biological differences we are interested in. In downstream tasks, batch effects are difficult to separate from true biological signal and can arise from various sources such as differences in sample preparation, sequencing protocols, or the specific equipment used.

Taking batch effects into consideration is particularly important when integrating data from multiple sources to create cell atlases. These aim to provide a reference map of cell types and states. Leveraging advanced representation learning techniques that can handle batch effects, enables researchers to create robust and comprehensive atlases. These resources serve as valuable references against which disease-specific changes, such as those observed in ADPKD, can be compared. Moreover, these integration approaches are critical for handling complex experimental designs, where samples may be processed using different protocols or processed in different pools.

1.3.1 DEEP REPRESENTATION LEARNING

To introduce non-linearities in the learned mapping function, several approaches have been explored in the literature. One such approach is the use of kernel methods, which map the original data into a higher-dimensional space (hence introducing non-linearities) where linear separation is possible[36]. More recently, deep neural networks (DNNs) – the focus of this thesis – have gained popularity for their ability to learn complex non-linear mappings directly from data through multiple processing layers. For a comprehensive review of the inner workings of neural networks we refer the reader to other resources[37,38]. Nevertheless, a simple abstraction is to understand neurons as computational units that perform a weighted sum of their inputs and then apply a non-linear function to this sum to produce an output. Connecting several neurons into layers allows the network to learn more complex patterns than linear models (**Figure 4A**).

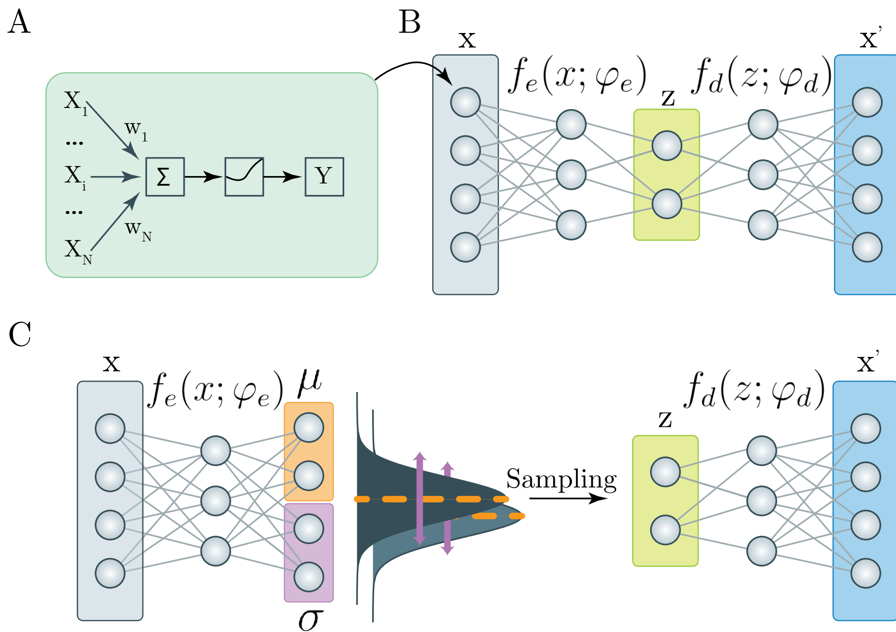

Figure 1.4: *Unsupervised learning using neural networks.* **A**. Illustration of a neuron. X represents the input features and w are the learned weights. After the weighted sum, outputs are fed through a non-linear function to obtain output values Y. **B**. Architecture of an Autoencoder. Spheres represent neurons, organised into layers. **C**. Architecture of a Variational Autoencoder. Circles represent neurons, organised into layers.

Taking a neuron as the main building block, several schemes (or architectures) have been proposed to tackle different problems and types of input data[39–42]. Among these, autoencoders (AEs) have emerged as a powerful tool for unsupervised representation learning when large amounts of unlabelled data are available. The goal of autoencoders is to learn a function that maps input data to a lower-dimensional representation by means of an encoder network $f_e(x;\mathbf{e})$. Here x represents the input data and $\mathbf{e}$ are the parameters of the encoder network

(e.g., weights). This function has many potential configurations since different parameter setups can lead to similar outputs. To constrain the training process, a decoder network $f_d(z;\mathbf{d})$ is also learned. In this case, z corresponds to the lower-dimensional representation of the data x and $\mathbf{d}$ are the parameters of the decoder network. The decoder projects the lower-dimensional representation back to the original feature space, obtaining reconstructed data x' (**Figure 4B**). In the simplest case, training aims to optimise the latent representation z and network parameters $(\mathbf{e},\mathbf{d})$ to minimise the loss function of the reconstruction (e.g., the Mean Squared Error):

$$\mathscr{L} = \frac{1}{n}\sum_{i=1}^{n} \|x_i - x_i'\|^2, \tag{1.1}$$

where x_i is the original input, x_i' is the reconstructed output, and n is the number of samples.

Several alternatives of the autoencoder architecture are available in the literature[43,44]. However, this thesis extensively uses Variational Autoencoders (VAEs), an application of autoencoders using Variational Inference[45]. The key difference between AEs and VAEs is that while AEs learn a latent representation z directly through the encoder function $f_e(x;\mathbf{e})$, VAEs first assume that z follows a normal distribution, specifically $z \sim \mathscr{N}(\mu,\sigma^2)$. Variational Inference is then used to estimate the true posterior distribution $p(z)$ with an approximate posterior $q(z|x)$, which is assumed to closely resemble a standard normal distribution, a widely-used prior. The encoder network is then used to learn the parameters of $q(z|x)$, namely μ and σ. The training process is constrained by minimising the Kullback-Leibler (KL) divergence—a measure of similarity between two distributions—between $q(z|x)$ and $p(z)$: $D_{\mathrm{KL}}(q(z|x)\|p(z))$.

Instead of using a deterministic decoder to reconstruct the original data, VAEs use a probabilistic decoder that maps the latent representation z to the parameters θ of a distribution from which data can be generated, namely $p(x|z)$ (**Figure 4C**). Similarly to AEs, an additional reconstruction loss term is added to the overall loss function. In this case, the reconstruction loss is the log-likelihood of observing the data x under the parameters of the learned generative distribution. We include an expectation (i.e., $\mathbb{E}_{z\sim q}[\cdots]$) to account for the probabilistic nature of z, ensuring that the loss function captures the average effectiveness of the decoder across various possible latent representations.

Since we want to maximise this likelihood, we include the term as a negative log-likelihood. We define the following loss function:

$$\mathscr{L} = \mathscr{L}_{\mathrm{KL}} - \mathbb{E}_{z\sim q(z|x)}\left[\log p(x|z)\right]. \tag{1.2}$$

Here, $\mathscr{L}_{\mathrm{KL}}$ is the KL divergence term, and the second term represents the reconstruction loss. Together, these ensure that the model learns a latent representation z that is both informative

and conforms to the desired prior distribution.

Several frameworks and methods have been developed that use the architecture of VAEs to learn deep representations in scRNAseq data[46–49]. Generally, these methods incorporate modelling choices that align with the technical limitations associated with scRNAseq data (i.e., sparsity and zero-inflation, Section 1.4). For example, the generative model used in most methods maps the latent representation z to the parameters of a zero-inflated distribution suited to the count-based nature of scRNAseq data[50]. However, scVI[51] stands out as the first model based on VAEs that can perform integration of cells displaying significant batch effects (i.e., non-biological variation). The model achieves batch-correction by leveraging Conditional VAEs (CVAE)[52]. Briefly, the model includes the batch covariates as extra variables into the encoder and decoder inputs. Importantly, since the decoder can use the batch information directly to reconstruct the data, the latent representation is freed from having to capture these technical differences and can instead focus on the underlying biological signal.

1.3.2 INTEGRATION ACROSS STUDIES

Given the ever-increasing number of scRNA-seq datasets available in public repositories, interpreting shared cellular phenotypes can help us better understand both health and disease by identifying common molecular signatures across populations and conditions. Recent efforts have been directed towards performing integrative analysis of within-organ datasets[53] or whole-organism datasets[54] to obtain reference atlases. These efforts typically use unsupervised learning to obtain a latent representation that captures shared biology while accounting for technical differences between datasets. Such atlases can then facilitate the interpretation of new datasets by automatically classifying cell types or uncovering previously overlooked cell populations[55].

However, integrating cell type information from different sources is challenging due to labelling inconsistencies driven by the unsupervised nature of cell type annotation. To address this, researchers have designed models that can handle cell type discrepancies by constructing hierarchies during the integration process[35,56]. These hierarchies help identify inconsistencies and similarities across datasets and harmonise cell types. Ideally, one would use such references to characterise previously overlooked populations or obtain shared markers across studies, furthering our understanding of molecular cellular phenotypes. Additionally, label transfer models[57–59] can annotate new datasets using these harmonised labels.

1.3.3 INTEGRATION ACROSS TECHNOLOGIES

Despite significant progress in Spatial Transcriptomics, current whole-transcriptome approaches still do not provide single-cell resolution. To address this limitation, researchers

have leveraged scRNA-seq data by learning joint representations that align both technologies[60–62]. These representations enable tasks such as estimating cell type proportions or imputing missing genes in targeted Spatial Transcriptomics[60,61], as well as enriching scRNA-seq with inferred spatial context[63,64]. However, most existing approaches assume that the aligned technologies capture similar cellular populations or that both technologies have been applied to the same set of samples. As integrated reference atlases increasingly support tasks like automatic cell type annotation, it becomes crucial to develop cross-technology integration methods that account for the distinct cellular composition observed by each technology.

As mentioned in Section 1.3, the additional spatial context in spatial transcriptomics datasets facilitates the estimation of communication events. By incorporating known ligand-receptor pairs as well as cellular niches or spatial proximity between pairs of cells, various methods have effectively leveraged this information to learn interaction-informed representations[65–67]. Given the increased number of spatial transcriptomic studies that aim to characterise tissues in health and disease, some methods have been introduced that allow for the contextualization of communication events among samples or conditions[68]. However, there remains a need to apply these methods specifically within tissues, particularly in complex environments such as tumours and cysts, to better understand the local dynamics of cellular communication.

1.4 Rerpesentation disentanglement

Disentanglement aims to identify and separate the fundamental factors that generate variation in data into distinct, interpretable components. In single-cell datasets from different species, conditions, or modalities, this concept extends beyond finding a single representation - it would be highly beneficial to separate the data into shared and unique components. Ideally, a shared representation would learn common cellular identities, while private representations would capture species-, condition-, or modality-specific factors of variation (**Figure 5**). Such disentangled representations could provide deeper insights into conserved cellular functions across species, disease-specific alterations, and technology-specific signals.

One of the first models to achieve such properties was β-VAE[69], which introduces a hyperparameter β that emphasizes the regularization of latent space to enhance factor separation. A classical application of such models is learning a deep representation from large portrait image datasets. In this case, disentangling the learned latent variables to learn characteristics such as mood, face shape, or hair color can help interpret and generalize these models to tasks outside their training datasets.

More recently, disentanglement techniques have expanded to multi-view datasets, where each sample is represented in different modalities. In such cases, training a single multi-modal or multi-view VAE may lead to a unified representation that obscures unique variations observed only in specific views. For instance, in a dataset with image-text description pairs, it is desirable to obtain a shared representation that captures the item described in each pair,

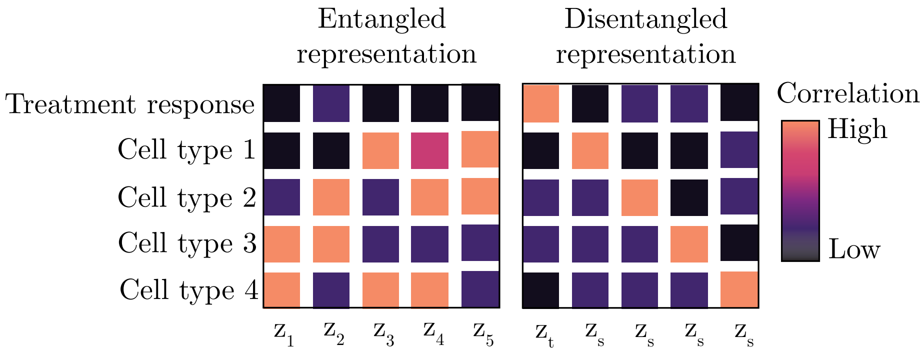

Figure 1.5: *Entangled and Disentangled representations.* **Left**: Entangled representation. Heatmap of the correlation between the learned latent factors ($z_1 - z_6$) and the unobserved true biological factors of variation. The treatment response is masked in the entangled representation. **Right**: Disentangled representation. Heatmap of the correlation between the treatment-specific (z_t) and shared latent factors (z_s) and the unobserved true biological factors of variation. The treatment response is learned by the treatment-specific latent factor.

and distinct representations for the visual and textual elements [70].

Some methods have been developed that try to disentangle representations through various architectural and loss choices [71,72]. These models, however, assume the availability of either paired multi-modal data or single-modality datasets in which the same set of features are measured across different conditions or groups. Disentangling these complex datasets remains a challenging and active area of research. This is largely due to the difficulties in meaningfully integrating and, crucially, separating diverse representations.

The advancement of disentanglement methods capable of handling varied features across different samples holds significant potential for enhancing our understanding of health and disease. In cancer research, for example, disentangled representations could separate tumor-specific alterations from common cellular responses to stress. In the context of ADPKD (Autosomal Dominant Polycystic Kidney Disease), disentanglement could potentially separate cyst-specific gene expression patterns from general kidney injury responses, potentially uncovering new biomarkers for disease progression.

1.5 Contributions of this thesis

Throughout this thesis, we address the challenges mentioned above using computational approaches that refine the integration of scRNA-seq and spatial transcriptomics data.

In **Chapter 2** we introduce the Mouse Kidney Atlas. Here we leveraged deep representation learning to integrate healthy single-cell RNA-seq datasets from different studies with a wide range of technologies. Additionally, we harmonise labels between datasets in a two-step

fashion: we first learn a cell type hierarchy between datasets using a previously published method[34] and perform extensive manual curation of cell types. We compare the classification potential of our learned representation against the only other (human) reference available[73] for kidney tissue and show improved performance with a variety of classifiers[74]. We leverage the increased resolutive power of our reference to better characterise overlooked populations and detect markers conserved across studies. In **Chapter 3** we investigate how established cysts reshape the molecular landscape of their surrounding tissue in ADPKD. We first fine-tuned our trained model with disease-associated populations from several studies to create a reference for deconvolution. Our integration approach specifically accounts for differences in cellular composition between kidneys. Additionally, we make use of a contextualised cellular communication approach to study discrete communication events in the cyst microenvironment. We then examine gene expression dynamics near cystic structures by fitting Generalised Additive Models, offering new insights into the molecular mechanisms driving cystogenesis. In **Chapter 4** we investigate the cellular mechanisms driving early stages of ADPKD, as the initial changes that lead to cyst formation remain poorly understood. To do so, we used single-nucleus RNA sequencing and evaluated different experimental approaches to minimise technical variation. This included comparing nuclei isolation protocols from fresh and frozen tissue, as well as testing modern sample multiplexing methods such as antibody hashing and probe-based demultiplexing to reduce batch effects.

In **Chapter 5**, we introduce spVIPES, a novel computational method to learn disentangled representations from scRNA-seq with unpaired features. After validating our model using simulated data and demonstrating that it outperforms state-of-the-art methods[51,59,71,75], we apply spVIPES to three biological contexts: (i) comparing kidney cell types across evolutionarily distant species without requiring ortholog mapping, (ii) studying shared and time-specific cellular responses during kidney regeneration following acute injury, and (iii) dissecting general and treatment-specific effects of IFN-*beta* stimulation in immune cells. Through these applications, we demonstrate how disentangled representations can provide new biological insights when analysing diverse experimental conditions.

Finally, in **Chapter 6** we provide a discussion of this thesis as well as potential future computational research directions in single-cell and spatial transcriptomics in the context of kidney physiology and ADPKD.