Applications of multisource data-based dynamic modeling to cell-cell signaling and infectious disease spreading

Chen, D.

Chapter 1

Introduction

Complex diseases and infectious diseases have consistently been significant issues threatening human survival and development. Understanding the mechanisms behind disease occurrence and predicting developmental trends from observational and experimental data are crucial research objectives in disease prevention and treatment. Statistical data analysis methods can quantitatively describe key information within datasets, thereby enhancing our comprehension of disease occurrence and development. Mathematical models, especially dynamic models derived from quantifying the disease occurrence process, can delve into the mechanisms of complex disease progression and infectious disease spread, predicting future disease states. Consequently, combining data with dynamic models can make full use of the information in the data to assist modeling, thereby improving the general applicability of dynamic modeling to diseases.

1.1 Background

Complex diseases significantly diminish the living quality of a patient and impose a substantial burden on the family of patients. Cardiovascular and cerebrovascular diseases alone account for over 10 million deaths annually, ranking as the leading global causes of death [1, 2, 3]. The incidence of deaths caused by malignant tumors is steadily increasing worldwide [4], with cancer-related deaths in the United States surpassing 600,000 in 2023 [5]. Cells, the fundamental biological units constituting all life, orchestrate crucial biological processes such as growth, development, metabolism, immunity, and the proliferation of living organisms [6]. In broad terms, complex dis-

eases arise because some cells in the patients are unable to perform normal biological functions [7]. Placing the cell at the center and exploring the cell-cell signaling mechanism, from gene dynamics to the collective behavior of cells, could yield new insights into the treatment of complex diseases.

Infectious disease is another major issue that threatens human health and social development. As of 2023, infectious disease epidemics, claiming more than 1 million lives, have occurred at least 17 times in human history [8]. The ongoing AIDS pandemic has resulted in more than 36.5 million deaths [9]. In many developed countries, the cost of preventing and treating infectious diseases exceeds 100 billion dollars annually [10]. The global COVID-19 epidemic, in recent years, has not only claimed over 10 million lives [11] but has also led to staggering economic expenditures and increased unemployment [12, 13]. Confronting the threats posed by infectious diseases, it is particularly crucial to explore their causes, reveal patterns of spread, and evaluate the effectiveness of intervention measures.

There has been a long history of using mathematical models to investigate key problems in infectious disease transmission [14] and cell-cell signaling [15]. As early as the 18th century, Bernoulli proposed a mathematical model to describe the spread of smallpox [16]. In 1927, Kermack and McKendrick proposed the famous SIR infectious disease model [17]. Since then, various types of infectious disease models have emerged [18, 19, 20]. In recent years, these basic infectious disease models have been developed into more realistic dynamic models [21, 22, 23]. The first type of extension considers pathogen dynamics within infected individuals, developing models that include pathogen evolution [24], age structure [25] and multi-scale dynamics [26]. The second type of expansion mainly considers macro-level influencing factors, such as contact patterns [27], population movement [28], environmental conditions [29] and various intervention measures against infectious diseases [30, 31]. Mathematical analysis of infectious disease models can provide sufficient conditions for disease extinction in a population [32, 33], which provide reference opinions and scientific guidance for disease prevention and control.

Exploring the cell-cell signaling mechanism from genes to the collective behavior of cells involves multiple biological scales, and dynamic models play a pivotal role at each scale. At the smallest length scales, dynamic models are employed to characterize gene transcription oscillation [34], switching dynamics [35], and nonlinearity [36]. On the cellular scale, dynamic models [37, 38] have been instrumental in exploring the determinants of cell shape [37], cell motility [39], and environment-dependent cell specialization [40]. Moving to tissue scales, dynamic models [41, 42] are also employed to

address inquiries in angiogenesis [43] and embryonic development [44]. On population scales, dynamic models are applied to study the collective behavior of animals [45] and the superspreading of pathogens [46]. Equally important, prospective prediction of biological consequences can expand our understanding to biological systems and guide our decisions in application areas [47, 28]. Equally important, the prospective prediction of biological consequences can enhance our understanding of biological systems and guide decision-making in application areas [47, 28]. For instance, Bentley et al. [41], based on model simulations, predicted that the level of protein Dll4 in endothelial cells would fluctuate under the stimulation of high concentrations of vascular endothelial growth factor (VEGF) in the extracellular matrix, limiting vascular branching. Later, they confirmed their predictions experimentally [48].

Compared with biostatistics [49] and bioinformatics [50], the contribution of dynamic models to life science mainly shows advantages in two aspects: (1) deciphering the mechanisms and rules that govern some biological phenomena; (2) prospectively predicting various biological consequences by mimicking the experimental conditions that are currently not easy to meet in reality. Traditionally, limited by the availability of data, dynamic modeling for biological systems has heavily relied on mechanistic assumption [51], that is, a conjectured approximation of the systems in the real world. The small amount of data available is mainly used to estimate the parameters in dynamic models [52] rather than to infer the mechanisms of biological processes involved in a biological system. Although such modeling provides us with a new view to understand the specific biological phenomena, it is common that multiple models based on different assumptions have the ability to explain the same phenomenon [53, 54, 55]. Consequently, facing a zoo of mathematical models without support from data, it is difficult to derive general rules governing the biological processes. Furthermore, overfitting of mathematical models to data also reduces the dependability of prospective predictions of some biological consequences based on the mathematical model [56]. To address these concerns about dynamic modeling, measuring and collecting data from different aspects of related systems is necessary, which could falsify or confirm the proposed dynamic models [57].

With the development of data measurement and collection techniques in recent years, multi-source data related to one biological process have become available [58]. During angiogenesis, for example, the strength of intercellular Notch signaling [59] and intracellular VEGF receptors (VEGFRs) signaling [60], the dynamics of actin polarization [61], as well as the quantification of Notch-related cell motility [62] and Notch-controlled cell rearrangement [63] can all be obtained in experiments. Similar

examples also appear in epidemiology. During an infectious disease outbreak, in addition to the cases of infected individuals, viral dynamics within the host [26], contact pattern between host [27], traffic flux [28] and meteorological data [29, 64] are also available. These data, originating from different scales and sources, on the one hand provide us with opportunities to improve modeling by reducing the unreality and uncertainty of mechanistic assumptions, but on the other hand pose challenges for us in terms of information integration.

The transmission of infectious pathogens in a population could also induce complex diseases in infected individuals, for example, human papillomavirus (HPV)-associated cancer [65]. Furthermore, the development of human immunodeficiency virus (HIV) within host [66] and the transmission of HIV between host [9] has characteristics of both a complex disease and an infectious disease. Thus, there is a potential link between cell-cell signaling, closely related to complex disease, and the spread of infectious disease. In this thesis we study two application areas separately, where the interplay between multi-source data analysis and dynamic modeling yields novel perspectives for understanding cell-cell signaling and infectious disease spreading, which could provide insights for designing novel interventions against complex disease development and infectious disease transmission. In the next section we introduce the biological state of the art and modeling studies for the first application area which is about Notch signaling that is a central cell-cell signaling in Metazoa development. Next we present previous modeling studies for the second application area which is about the spread and control of infectious disease.

1.2 Modeling cell-cell signaling

All tissue development and replenishment relies upon the differentiation of progenitor cells into more specialized cells [7]. Cell-cell signaling, especially the Notch signaling system is one of the main engines driving this process [67]. Notch signaling was the first cell receptor signal transduction pathway to be discovered, more than a century ago [68], and decades of research since then have established that it is a central regulator of cell fate that underpins normal embryo maturation and tissue homeostasis, from controlling the fine-grain patterning of insect wings to orchestrating human organ development [69]. An example related to this is that Notch signaling pathway together with vascular growth factor receptors (VEGFRs) signaling pathway controls the specialization of endothelial cells during blood vessels growth [70]. Endothelial cells specialize into either endothelial tip cells or endothelial stalk cells [71]. Tip cells

with high motility play a leading role guiding cell movement to form new vascular branches and stalk cells with high proliferation rates develop the new branches. In **Part I** of this thesis, by integrating experimental data at both the cellular level and tissue level, we develop dynamic models for Notch signaling and VEGFRs signaling with a particular focus on cell specialization and tissue development.

1.2.1 Notch signaling

Notch and Notch ligands are transmembrane protein families found in metazoan species [72]. The basic paradigm of Notch signaling is that Notch ligand expressed in one cell binds to and trans-activates Notch receptor in neighboring cells. The activated Notch receptor releases Notch intracellular domain (NICD) which would trans-locate to the cell nucleus to up-regulate or down-regulate the expression of many genes including Notch and its ligands (Fig 1.1). This process is called trans-actication of the Notch receptor [72]. In addition to the trans-activation mediated by Notch ligand from neighboring cells, Notch activity can be inhibited by the ligands expressed in the same cell, which is called cis-inhibition of the Notch receptor [73]. Based on this simple paradigm of Notch signaling (Fig 1.1), Collier et al. [74] proposed the first dynamic model, a set of ordinary differential equations, to investigate Notch signaling in the

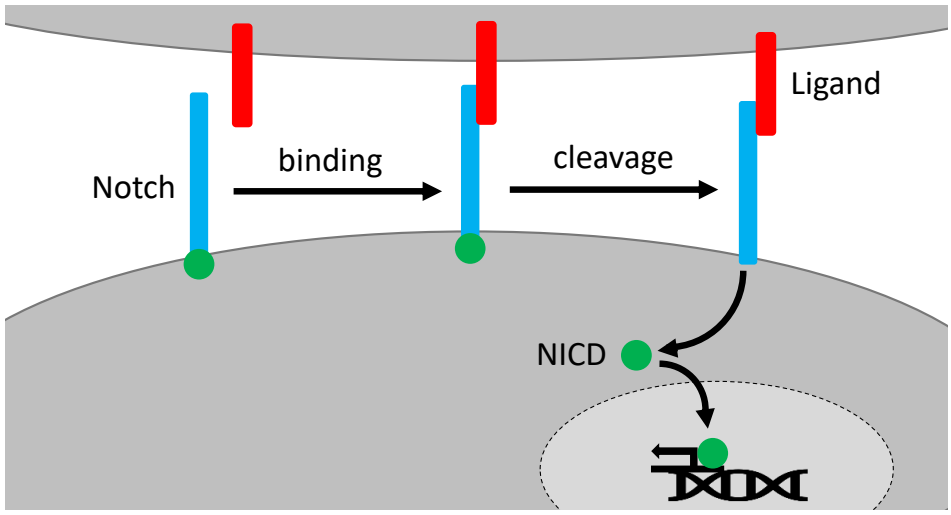

Figure 1.1: The basic paradigm of Notch signaling. Binding of Notch ligand from one cell to Notch in a neighboring cell results in the cleavage of Notch intracellular domain (NICD) which regulates genes expression in the nucleus.

context of multicellular pattern formation. In any cell, they focused on the amounts of Notch ligand and Notch signal (e.g., free NICD). Mathematically,

$$\begin{aligned} \frac{\mathrm{d}L}{\mathrm{d}t} &= F(S_n) - \beta L, \\ \frac{\mathrm{d}S_n}{\mathrm{d}t} &= G(L_{ext}) - \beta_S S_n. \end{aligned} \tag{1.1}$$

Here the production rate of Notch ligand (L) is negatively ($F'(S_n) < 0$) regulated by the Notch signal (S_n) in the same cell. The intracellular Notch signal (S_n) is triggered ($G'(L_{ext}) > 0$) by trans ligands (L_{ext}) from neighboring cells. Parameters β and β_S quantify the first-order degradation rate of Notch ligand and Notch signal (e.g., free NICD) in the cell. This model recapitulates key aspects of multi-cellular patterning in the vertebrate nervous system [34] and other tissues [75, 76, 77]. Because the cis-inhibition of Notch activity is not considered in this model, the model can not explain experimental observations related to cis-inhibition [78]. For example, it can not recapitulate Notch ligands Delta/Serrate-mediated cis-inhibition of Notch signaling in *Drosophila* wing vein patterning [79, 80].

More recent models have helped solved some of these problems, Sprinzak et al. [81] found that the mechanism of Notch receptor cis-inhibition [82] is because of the mutual inactivation of Notch and its ligands in the same cell. By taking the cis-inhibition of Notch receptor into account, they proposed a new mathematical model including both trans-activation (k_t) and cis-inhibition (k_{ci}) of Notch receptor in the same cell. The model is described mathematically as,

$$\begin{aligned} \frac{\mathrm{d}L}{\mathrm{d}t} &= b_L - \beta L - k_t N_{ext} L - k_{ci} N L, \\ \frac{\mathrm{d}N}{\mathrm{d}t} &= b_N - \beta N - k_t L_{ext} N - k_{ci} L N, \\ \frac{\mathrm{d}S_n}{\mathrm{d}t} &= k_t L_{ext} N - \beta_S S_n, \end{aligned} \tag{1.2}$$

where b_L and b_N are the production rates of Notch ligand (L) and Notch (N) in the cell. Trans ligand (L_{ext}) from neighboring cells mediates tans-activation of Notch, which generates a Notch signal (S_n). Similarly, Notch ligand can also bind to the Notch (N_{ext}) from other cells in trans. Notch ligand and Notch receptor mutually inactivate each other in the same cell ($k_{ci}NL$). Parameters β and β_S quantify the degradation of proteins in the cell. This model recapitulates the properties of Notch receptor cis-interactions [81] and the sharp boundary of *Drosophila* wing vein patterning [79, 80].

If including a feedback loop from Notch signal to the production rate of Notch ligand and Notch receptor, this model would also recapitulate the patterning of other tissues [83].

Cis-activation, where Notch ligand activates Notch receptor in the same cell, has recently been reported [84, 85]. However, the strength of this autonomous Notch signaling in cells is very low compared to the signaling resulted from ligand-mediated trans-activation of Notch receptor [85]. Therefore, the cis-interactions between Notch and Notch ligand typically lead to inhibition of Notch receptor in the same cell [78, 86]. Consequently, the canonical Notch signaling in cells is largely governed by a balance between trans-activation and cis-inhibition of Notch receptor [73], and the Sprinzak model (1.2) has served as a foundation for subsequent modeling extensions in other Notch-related biological processes [87, 42, 88].

Because of the mutual inactivation of Notch and its ligand in the same cell, one prediction of the Sprinzak model (1.2) is that a cell is either in a ligand-dominated sending state or a receptor-dominated receiving state, depending on the relative production rates of Notch receptor and its ligand in the same cell [81]. However, this mutually exclusive behaviour of Notch ligand and Notch receptor in cells is not consistent with the observations in Notch-controlled dorsal/ventral boundary patterning in the wing disc of *Drosophila* where the levels of both Notch ligand and receptor are high in ventral cells [89, 90, 91, 92]. In Chapter 2 of this thesis, we experimentally show that Notch ligands self-associate through homo-dimerization or homo-oligomerization, which is a previously unreported process in Notch signaling. Through a combination of experimentation and dynamic modeling, we have broken new ground by refining the Sprinzak model (1.2) and propose a new model where Notch signaling in a cell is due to ligand monomer-mediated Notch trans-activation, dampened by ligand dimer-mediated cis-inhibition of Notch receptor. The new model recapitulates our data as well as published data related to Notch signaling, including the dorsal/ventral boundary patterning in the wing disc of *Drosophila* [47].

1.2.2 Notch-related cell fate decision

The large variety of cell types in a multi-cellular organism are formed through a series of cell fate decisions [93]. These are driven by chemical and mechanical signals within and between cells [94, 95]. Delta-Notch signaling-mediated lateral inhibition between adjacent cells is an important mechanism for breaking symmetry and creating diversity in a population of homogeneous cells [96, 67, 97]. The regulatory mechanism

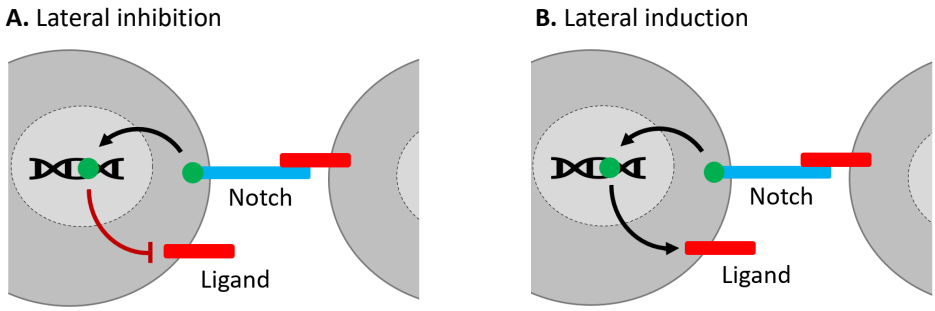

Figure 1.2: Schematic illustration of the regulatory mechanisms of Notch signaling-mediated lateral inhibition and lateral induction between adjacent cells. (A) Notch signaling down-regulates ligands expression in the same cell. (B) Notch signaling up-regulates ligands expression in the same cell.

of Notch-controlled lateral inhibition between cells is that the binding of Notch ligand in one cell to Notch receptor in a neighboring cell triggers Notch signaling which down-regulates the expression of Notch ligand in the receiver (Fig 1.2A). Therefore, a cell with high levels of Notch ligand would prevent neighboring cells from adopting the same cell fate by triggering Notch signaling in its neighbors [74]. Consequently, lateral inhibition generates a 'salt-and-pepper' pattern in a population of adjacent cells. For example, this phenomenon has been observed in the development of *Caenorhabditis elegans* vulva and uterine [98], the formation of *Drosophila* epidermal sensory bristles [99, 100], and the pigmentation pattern of zebrafish [101]. Based on the regulatory mechanism of Notch signaling-mediated lateral inhibition between adjacent cells (see Fig 1.2A), Sprinzak et al. [83] and Boareto et al. [87] extended the model (1.2) to model (1.3):

$$\begin{aligned}
\frac{\mathrm{d}L}{\mathrm{d}t} &= b_L H(S_n;\theta_L,\lambda_L,l_L) - \beta L - k_t N_{ext} L - k_{ci} N L,\\
\frac{\mathrm{d}N}{\mathrm{d}t} &= b_N H(S_n;\theta_N,\lambda_N,l_N) - \beta N - k_t L_{ext} N - k_{ci} L N,\\
\frac{\mathrm{d}S_n}{\mathrm{d}t} &= k_t L_{ext} N - \beta_S S_n.
\end{aligned} \tag{1.3}$$

Here the shifted Hill function $H(X;\theta,\lambda,l) = (\lambda X^l + \theta^l)/(X^l + \theta^l)$ represents positive regulation ($\lambda > 1$) or negative regulation ($\lambda < 1$) mediated by protein X to the production of other proteins (e.g., the regulations in Fig 1.2). To represent the Notch signaling-mediated regulation pathway in Fig 1.2A mathematically, there is $\lambda_L < 1$ in model (1.3). Parameters θ, λ and l control the shape of this Hill function. In

the absence of cis-inhibition of Notch receptor ($k_{ci} = 0$), the mathematical model (1.3) would be a special case of the mathematical model (1.1) for tissue patterning if we assumed that the level of Notch receptor in cell is at a quasi-steady state (i.e., $dN/dt = 0$). Compared to the mathematical model (1.1), Sprinzak et al. [83] find that the model (1.3) with mutual inactivation of Notch and its ligand is more robust for facilitating developmental patterning.

In addition to vulva development in *C. elegans* [98] and epidermal sensory bristle formation in *Drosophila* [99, 100], lateral inhibition also works in vertebrate blood vessel development [41, 102, 103, 104]. The lateral inhibition is driven by intracellular vascular growth factor (VEGF) signaling and intercellular Delta-like (Dll)4-Notch signaling in endothelial cells, which occurs in two steps: (1) the binding of VEGF to vascular growth factor receptor 2 (VEGFR2) generates VEGFR2 signal which results in the up-regulation of Dll4 in one cell; (2) Dll4-transactivated Notch activity leads to the down-regulation of VEGFR2 in the neighboring cells (see Fig 1.3). Consequently, stimulated by pro-angiogenic cues such as VEGFs, some endothelial cells will specialize into tip cells to lead the collective motility of neighboring stalk cells that develop the branches of blood vessels (see Fig 1.3; [105]), which is called sprouting angiogenesis. At the early stage of sprouting angiogenesis, the VEGF-Dll4-Notch lateral inhibition

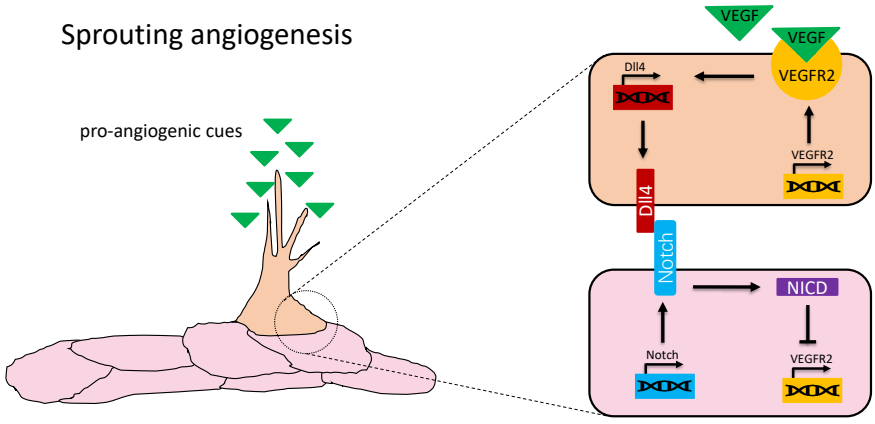

Figure 1.3: Schematic illustration of sprouting angiogensis and the regulatory mechanisms of tip endothelial cell selection. Intracellular VEGF signaling and intercellular Dll4-Notch signaling drives the lateral inhibition between adjacent endothelial cells. In particular, the binding of external VEGF to VEGFR2 in tip cell leads to the up-regulation of Dll4 which triggers Notch signaling in the neighboring cell. Notch signaling down-regulates the expression of VEGFR2 in the same cell.

generates an alternating pattern of cell fates, a 'salt-and-pepper' pattern, along the front of blood vessels, which has been observed in mice [41], zebrafish [106] and *in vitro* [48]. The extent to which the endothelial cells behave as tip cells depends on the level of intracellular VEGFR2 signaling [59, 106] that is generated by the binding of VEGF to the VEGFR2. According to this regulatory mechanism for tip cell selection in sprouting angiogenesis, Stepanova et al. [107] refined the model (1.3) by considering the VEGF-Dll4-Notch signaling:

$$\begin{aligned}
\frac{\mathrm{d}D}{\mathrm{d}t} &= b_D H(S_v;\theta_D,\lambda_D,l_D) - \beta D - k_t N_{ext} D - k_{ci} N D,\\
\frac{\mathrm{d}N}{\mathrm{d}t} &= b_N H(S_n;\theta_N,\lambda_N,l_N) - \beta N - k_t D_{ext} N - k_{ci} D N,\\
\frac{\mathrm{d}S_n}{\mathrm{d}t} &= k_t D_{ext} N - \beta_S S_n,\\
\frac{\mathrm{d}R}{\mathrm{d}t} &= b_R H(S_n;\theta_R,\lambda_R,l_R) - \beta R - k_v V_{ext} R,\\
\frac{\mathrm{d}S_v}{\mathrm{d}t} &= k_v V_{ext} R - \beta_S S_v,
\end{aligned} \tag{1.4}$$

where the binding of external VEGF (V_{ext}) to VEGFR2 (R) generates intracellular VEGF signal (S_v) that leads to the up-regulation ($\lambda_D > 1$) of Dll4 (D). Moreover, the binding of Notch ligand Dll4 (D_{ext}) from neighboring cell to Notch receptor (N) generates Notch signal (S_n) that leads to the down-regulation ($\lambda_R < 1$) of VEGFR2 in the receiving cell. Based on the same theoretical understanding of tip endothelial cell selection in sprouting angiogenesis (Fig 1.3), Bentley et al. [41] developed an agent-based model to investigate tip cell selection in angiogenesis. All of these models recapitulate the 'salt-and-pepper' pattern generated by tip and stalk endothelial cells. The mathematical model (1.4) can be reduced to model (1.3) using the quasi-steady state assumption (i.e., $dR/dt = 0$; $dS_v/dt = 0$).

In contrast to lateral inhibition, Notch signaling also mediates lateral induction where adjacent cells tend to adopt the same cell fate. For example, the development of vascular smooth muscle cells on arteries [108] and the specification of the sensory progenitor cells during *Drosophila* inner ear development [109]. Notch-associated lateral induction occurs because Notch and its ligands form a positive feedback loop between adjacent cells (see Fig 1.2B). Mathematical models have been proposed to investigate the effects of lateral induction mediated by intercellular Notch signaling on cell fate in tissue development [87, 42].

Dll4-Notch signaling has long been thought to mediate a lateral inhibition between endothelial cells (Fig 1.3), which has been discussed in previous theoretical

research [41, 42, 107]. This theoretical understanding is supported by a 'salt-and-pepper' pattern of tip and stalk cells phenotype along the sprouting branch of blood vessels [102, 106]. However, experimental data show that increasing cell-cell contact results in both higher Notch signaling and Dll4 expression [43], while inhibition of Notch signaling suppresses the expression of Dll4 in endothelial cells [43, 62, 48, 110]. These data suggest that intracellular Notch signal may also lead to the upregulation of Dll4 in endothelial cells, which drives a lateral induction between adjacent cells (see Fig 1.2B). Therefore, the intracellular Notch signal (S_n) in endothelial cells could be a paradoxical component [111], that simultaneously has two opposite effects on the same target, Dll4. How this paradoxical component regulates endothelial cell fate during vascular growth is still not fully understood. In Chapter 3 of this thesis, we have refined the Stepanova model (1.4) by including the positive effects of Notch activity on Dll4 expression and investigate the biological consequence of this paradoxical effect in endothelial cell fate during sprouting angiogenesis.

In vertebrates, the Notch pathway is composed of up to four distinct receptor types (Notch1-4) and five different membrane-bound ligands: Jag1, Jag2, Dll1, Dll3 and Dll4 [112]. It has been established that different ligands can elicit unique cell fates [113, 114, 115]. For example, Jag1 stimulates blood vessel production whereas Dll4 limits this process [43]. In order to decipher the rules of the opposite effect of Dll4 and Jag1 on blood vessel growth, Boareto et al. [42] hypothesized that Notch signals in endothelial cells result in the up-regulation of Jag1 expression. Based on this regulatory mechanism, Boareto et al. [42] extended dynamic model (1.4) by adding Jag1 dynamics:

$$\begin{aligned}
\frac{\mathrm{d}D}{\mathrm{d}t} &= b_D H(S_v;\theta_D,\lambda_D,l_D) - \beta D - k_t N_{ext} D - k_{ci} N D,\\
\frac{\mathrm{d}J}{\mathrm{d}t} &= b_J H(S_n;\theta_J,\lambda_J,l_J) - \beta J - k_t N_{ext} J - k_{ci} N J,\\
\frac{\mathrm{d}N}{\mathrm{d}t} &= b_N H(S_n;\theta_N,\lambda_N,l_N) - \beta N - k_t (D_{ext}+J_{ext}) N - k_{ci}(D+J) N,\\
\frac{\mathrm{d}S_n}{\mathrm{d}t} &= k_t (D_{ext}+J_{ext}) N - \beta_S S_n,\\
\frac{\mathrm{d}R}{\mathrm{d}t} &= b_R H(S_n;\theta_R,\lambda_R,l_R) - \beta R - k_v V_{ext} R,\\
\frac{\mathrm{d}S_v}{\mathrm{d}t} &= k_v V_{ext} R - \beta_S S_v,
\end{aligned} \tag{1.5}$$

where the intracellular Notch signal (S_n) mediated by Dll4 (D_{ext}) and Jag1 (J_{ext}) from neighboring cells leads to the up-regulation ($\lambda_J > 1$) of Jag1 (J) and down-regulation

($\lambda_R < 1$) of VEGFR2 (R). Other behaviors of Jag1 are assumed to be similar to those of Dll4. In this model, the intercellular Dll4-Notch signaling drives a lateral inhibition between adjacent cells, but the intercellular Jag1-Notch signaling mediates a lateral induction. Consequently, Boareto et al. recapitulate the 'salt-and-pepper' pattern of tip and stalk cell fates in a line of interacting cells [41, 106, 48] by increasing the production rate of Dll4 [42]. Increasing the production rate of Jag1, they find that the 'salt-and-pepper' pattern of cell phenotypes was destabilized and there is a hybrid tip/stalk phenotype defined by an intermediate state between tip and stalk cell fates [42]. Although the intermediate cell state has been observed in other biological processes such as epithelial-mesenchymal transition and hematopoietic progenitor cells differentiation [116], to date, a hybrid tip/stalk phenotype of endothelial cells has not been identified experimentally. Moreover, how the Jag1-induced hybrid phenotype of endothelial cell [42] is linked to the excessive sprouting promoted by Jag1 in cells [43], i.e., the collective behavior of hybrid endothelial cells, is elusive.

In Chapter 4 of this thesis, through a combination of experimentation and mathematical modeling, we have refined the mathematical model (1.5) by considering the hetero-dimerization of Dll4 and Jag1 in endothelial cells, which is an extension of our work in Chapter 3. The hetero-dimerization of Dll4 and Jag1 is also confirmed experimentally in Chapter 2 of this thesis. In contrast to a hybrid cell phenotype, we predict that endothelial cells tend to adopt a tip cell fate when the level of Jag1 is high in cells, which is consistent with the event that Jag1 expressed in endothelial cell stimulates tip cell selection and vascular sprouting in angiogenesis.

1.2.3 Multicellular network formation

The specialization of endothelial cells into endothelial tip cell or endothelial stalk cell contributes to vascular growth in vertebrates [117, 118, 119] through coordinating sprouting angiogenesis that is the process of forming a new vascular branch from the existing blood vessels [120, 121, 104]. Complementary to angiogenesis, forming the vascular networks *de novo* from dispersed endothelial progenitor cells is called vasculogenesis [104], which can be observed in the development of early embryo [122] and retinal vasculature [123]. Vasculogenesis is easier to mimic *in vitro* than angiogenesis, and a commonly used *in vitro* model is the culturing of dispersed endothelial cells on the top of *Matrigel* which is a popular matrix for vasculogenesis assays [124, 125]. Where, the dispersed endothelial cells would autonomously form a vascular-like network structure [126, 124, 125].

Alongside experimental studies about vasculogenesis, researchers also benefit from mathematical modeling to decipher the mechanisms of this process. So far, the mathematical models for vascular patterning can be mainly categorized in continuous models [127, 124, 128], single-particle models [129, 130] and multi-particle models [53, 54, 131]. The continuous models based on differential equations partially account for the spatial dynamics of endothelial cells density [124], but fail to recapitulate the morphodynamics of individual cells during collective motility of endothelial cells. The single-particle model represents cells as individual particles and recapitulates the collective motility and interactions of endothelial cells based on a Lagrangian equation [129], but does not consider the dynamics of cell shape [53, 132]. The cells in the multi-particle model such as Cellular Potts model [133] have more complex structures that can mimic cell shape changes during vascular patterning [53, 54, 131]. Nevertheless, the current mathematical models are still deficient in two ways when quantifying the vasculogenesis. One is that the model only recapitulates the vessel-like pattern and does not account for the dynamics in cell morphology [132]. The second is that some recently discovered results about cell-cell signaling [47] is not included. On the one hand, cell-cell signaling affects cell motility [62]. On the other hand, cell motility affects cell-cell signaling because the neighbors of any cell could change during vasculogenesis. To recapitulate the experimental data at different biological scales, we have integrated our dynamic models for endothelial cells fate decision into the Cellular Potts model and have investigated the effects of cell-cell signaling on vasculogenesis in Chapter 5 of this thesis.

1.3 Modeling infectious disease spreading

In **Part II** of this thesis we shift our topic from modeling cell-cell signaling to modeling infectious disease spreading. Early dynamic models of infectious disease were relatively simple, usually assuming a closed population and constant biological infectiousness of pathogens. The population is assigned to compartments with different labels representing the disease state (Fig 1.4). According to the development of infectious disease, there are different diagrams of modeling. One of the simplest models is the classic SIR (susceptible-infectious-recovered) model (Fig 1.4A; [17]), in which susceptible individuals were infected and then recovered with immunity. For some infections without long-lasting immunity or with a latency period, there are basic SIS model (Fig 1.4B) and SEIR (or SEIS) model (Fig 1.4C). Over the past decades, these basic models have been extended to varying degrees by taking into account different factors [21, 22, 23] such as pathogen dynamics [26], age structure [25], contact pattern [27], population

mobility [28], meteorological condition [29] and even various interventions [134, 30, 31] against infectious disease.

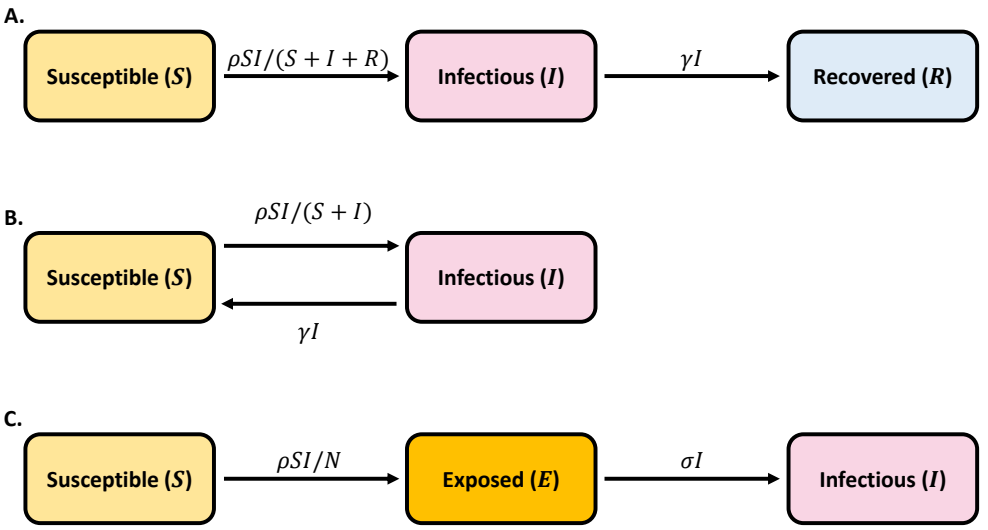

Figure 1.4: Diagram of the basic compartmental models in epidemiology. The transition rates from one compartment to another follow the law of mass action. (A) The SIR model consists of three compartments: the number of susceptible individuals, infectious individuals and recovered individuals, respectively. (B) The SIS model for some infections without long-lasting immunity. (C) An exposed compartment for the infections with a latency period.

In recent years, an increasing amount of data related to these factors has become available, which is helpful for characterizing the disease spreading in a population [46]. Generally, the data related to infectious disease can be categorised into epidemic related data such as the numbers of confirmed cases and deaths [26, 29] and behavior related data such as contact matrix [27] and movement matrix [28]. In **Part II** of this thesis we will model the transmission of an infectious disease based on the information from multi-source data related to infectious disease by considering two applications. In Chapter 6 we investigated the effect of air pollution on respiratory infection, and answered whether and how air pollution and interventions against air pollution in developing countries affect the dynamics of air quality and respiratory infection. In Chapter 7 we focused on the trade-off between mobility restriction and transmission of infectious disease, and discussed when and to what extent the mobility restrictions against severe epidemic outbreaks could be lifted.

1.3.1 Air pollution and respiratory infection

Understanding the driving force of infectious disease spreading is critical to design effective interventions curbing the threat of infectious disease transmission to public health around the world [135]. In tracking driving factors of infectious disease spreading, investigating the relationship between environmental changes and infectious disease transmission is one of the most important topics [136, 137, 138, 139]. Previous works have established that monsoon rains and temperature affect the epidemiology of cholera [140] and the life cycles of vectors such as mosquitoes, and the parasites that they transmit, so they are important environmental drivers of malaria [141], dengue [142] and Ross River fever [143]. Moreover, regional temperature and humidity are related to influenza transmissibility [144, 145]. Air pollution is also an environmental threat to human health. For example, air pollution leads to 3.3 million premature deaths globally each year [146] and has a substantial role in many non-communicable diseases [147] including cancer [148], stroke [3], cardiovascular disease [1, 2] and Alzheimer's disease [149, 150]. Furthermore, available evidence suggests that air pollution can prevent the beneficial cardiopulmonary effects of walking in people with heart or chronic lung disease [151, 152] and results in poor lung function in children [153, 154].

In recent times, there is an increasing recognition that poorer air quality particularly in developing countries is synchronized with a higher incidence of respiratory infection [155]. For instance, air pollution is associated with an increased risk of tuberculosis [156, 157], influenza [158, 159, 160], influenza-like illness [161, 29] and COVID-19 [162, 163]. Different from the environmental factors like temperature and precipitation that are difficult to be artificially controlled, air quality in a local area is mainly determined by the persistent emission and removal of air pollutants [164] which can be changed by human activity. For example, using more green energy instead of fossil fuels. Therefore, based on the conclusion that there is an association between air pollution and respiratory diseases, the government should formulate regulatory policies to reduce the levels of air pollutants and remind people to take preventive and control measures to reduce the incidence of respiratory infection. However, the studies of associations between air pollution and respiratory infection have two flaws. First, the correlation does not imply a causation [165]. Second, correlation research alone can not quantify how interventions taken by governments or individuals affect the transmission of respiratory infection. Related to this, in Chapter 6 of this thesis we first established a causal relationship between air pollution and the incidence risk

of influenza-like illness based on real world data, and then proposed a mathematical model linking air pollution and respiratory infection.

Previous studies have modeled population development coupling with environment pollution to investigate the effects of environment pollution on population dynamics [166, 167]. Inspired by this, the dynamic models for infectious disease coupling air pollution can describe the dynamic change of outdoor air pollutants and their impact on the transmission of respiratory disease, and can predict the effects of interventions against air pollution on respiratory disease transmission. In China, the degree of air pollution is quantified using an air quality index (AQI), and high value of AQI means poor air condition outside [168]. By considering the emission and removal of air pollutants, Tang et al. [29] proposed a simple ordinary differential equation (ODE) model for AQI dynamics ($F(t)$) in China:

$$\frac{\mathrm{d}F}{\mathrm{d}t} = c(t) - \mu(t)F(t), \tag{1.6}$$

where the emission rate of air pollutant is a periodic function $c(t)$ and the clearance rate is given by another periodic function $\mu(t) = \mu_0 + \mu_1 \cos(\omega t + \phi)$. They assumed an AQI-dependent incidence $\chi(t)$ of respiratory infection and then integrated AQI dynamics into a compartment model extended from the classic SEIS model (Fig 1.4C) with constant population (N). That is,

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\chi(t) + \gamma_s I_s + \gamma_a I_a, \\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \chi(t) - \sigma E, \\
\frac{\mathrm{d}I_s}{\mathrm{d}t} &= \delta\sigma E - \gamma_s I_s, \\
\frac{\mathrm{d}I_a}{\mathrm{d}t} &= (1-\delta)\sigma E - \gamma_a I_a.
\end{aligned} \tag{1.7}$$

Here the susceptible individuals (S) are infected by infectious individuals (I) and enter the exposed period (E), and then become infectious individuals who recover at a constant rate. The AQI-dependent incidence rate $\chi(t)$ is given by

$$\chi(t) = \rho F(t)\big(1 - \frac{bF(t)}{K + F(t)}\big)\frac{I_s + \theta I_a}{N} S,$$

which assumes that the incidence of respiratory infection is related to air quality $F(t)$ and individuals behaviour change such as wearing masks and reducing outdoor activ-

ities, quantified by $1 - bF(t)/(K + F(t))$. Here the infectious individuals are classified into symptomatic (I_s) and asymptomatic (I_a), and the latter has lower infection force than the symptomatic ($0 < \theta < 1$). For some diseases, recovered individuals are immune to the corresponding pathogens, which can be represented by extending the model (1.7) to include a new compartment for immune individuals [17]. For more epidemic models in air-polluted environment, He et al. [169, 170, 171] modeled the effects of stochastic factors and random diffusion of air pollutants on the dynamics of influenza-like illness; Shi et al. [172] modeled the direct and indirect effects of air pollution on respiratory infection; and Song et al. [173] modeled the spatial–temporal effects of air pollution on the influenza transmission.

By analyzing the dynamics of the model (1.7), Tang et al. evaluated the effects of reducing emission of air pollutants on the prevalence of respiratory infection and suggested persistent control of pollutant emission in China. However, the real intervention such as driving restrictions is state-dependent, which means that control of pollutant emission is triggered only if AQI is bigger than a threshold F_c. Therefore, the model (1.6) for AQI dynamics should be extended to:

$$\frac{\mathrm{d}F}{\mathrm{d}t} = (1 - \epsilon\eta)c(t) - \mu(t)F(t), \tag{1.8}$$

with

$$\epsilon = \begin{cases} 0, & F - F_c < 0, \\ 1, & F - F_c > 0. \end{cases}$$

Here η ($0 \leq \eta \leq 1$) quantifies the strength of control to air pollutants emission. Model (1.8) is a Filippov system. Although the mathematical methods for analyzing autonomous planar Filippov systems have been established [174], analyzing non-autonomous Filippov systems remains challenging. And the effects of this AQI-dependent control on the dynamics of respiratory infection is unclear. In chapter 6 of this thesis, we performed theoretical analysis to system (1.8) and then integrated the dynamics of AQI into a epidemic model based on the results of casual analysis between air pollution and respiratory infection. Finally, we investigated the effects of this intervention against air pollution on respiratory infections by analyzing the AQI-embedded epidemic model of respiratory infection.

1.3.2 Mobility restrictions versus disease spreading

The epidemic data of infectious diseases have shown that the number of infected individuals in a local area will increase exponentially in a short period of time when one infectious individual is introduced and there are no interventions [46]. This exponential growth pattern of the epidemic will follow from one region to another when there is a regional movement of population [175, 176, 177, 178, 179]. Consequently, the population mobility has a strong impact on the persistence and extinction of disease [180, 181]. The mathematical models describing the spatial mobility of a population in epidemic areas can be categorized into spatially continuous model like reaction-diffusion model [182, 183] and spatially discrete model like meta-population model [184, 185]. The mobility of population in reaction-diffusion models is quantified by an isotropic diffusion coefficient of individuals [186, 187, 188, 189]. The meta-population model is more general in describing the heterogeneity of the population [185, 190]. For modeling spatial heterogeneity in a population, the meta-population model is also called patch model [191], where the population is divided into n patches, and there are two matrices representing the migration of susceptible individuals (S) and infectious individuals (I) between these patches. For example, there is an SIS model in each patch i ($i = 1, ..., n$):

$$\begin{aligned} \frac{\mathrm{d}S_i}{\mathrm{d}t} &= \sum_{j=1}^{n}(D_{i,j}S_j - D_{j,i}S_i) - \rho_i \frac{S_i I_i}{S_i + I_i} + \gamma_i I_i, \\ \frac{\mathrm{d}I_i}{\mathrm{d}t} &= \sum_{j=1}^{n}(L_{i,j}I_j - L_{j,i}I_i) + \rho_i \frac{S_i I_i}{S_i + I_i} - \gamma_i I_i, \end{aligned} \tag{1.9}$$

where $D_{i,j}$ represents the rate at which susceptible individuals migrate from patch j to patch i, and $L_{i,j}$ represents the migration rate of infectious individuals. There would be different patch models [192, 193, 194] by extending other basic models such as SIR model or SEIR model (see Fig 1.4).

For infectious disease without effective treatments, to keep national health systems from becoming overburdened and to reduce infection pressure for people with a high risk of severe outcomes, mobility restrictions such as quarantine and lockdown have been used many times in the prevention and control of outbreaks [20, 195, 196, 197]. Although mobility restrictions effectively reduce the spatial spread of infectious disease, they have negative effects on social health and economic development [198, 199]. This has begged a question: is there an optimal strategy for lifting mobility restrictions in epidemic areas?

By analyzing a two-patch model (i.e., $n = 2$ in Eq. 1.9), Wang and Mulone [191] have found that the dispersal rates of susceptible individuals has no effect on the persistence and extinction of the disease. Furthermore, if a disease spreads or becomes extinct in each patch when the patches are isolated, the disease in connected patches would have the same consequences. Particularly interesting, if the dispersal rates of infectious individuals are suitably chosen, the disease can disappear in all patches even if it spreads in one of two patches when they are isolated. However, how the dispersal rates of individuals can be determined is unclear. Related to this, can we give a sufficient condition for the existence of a migration matrix $L = \{L_{i,j}\}$ for mathematical model (1.9) such that the disease disappears in all patches, and find a way to determine the matrix? The solution to the current problem gives a strategy to lift the motility restrictions in epidemic areas.

To investigate the trade-off between lifting mobility restrictions and controlling infectious disease spreading, previous studies mainly focused on various 'what if' question [192, 193, 194]. For instance, what would be the peak time, peak value and final size of infection if 50% of migration of infectious individuals was blocked at a specific moment? They first assumed a specific migration matrix in model (1.9) and then ran this model to investigate the dependence of epidemic development on population mobility and to assess the effectiveness of a certain mobility in curbing disease spreading. Although these studies provide some perspectives for disease prevention and control, the results based on testing predictions do not suggest an optimal strategy for lifting the mobility restrictions in epidemic areas. In Chapter 7 of this thesis, we have developed a novel method which can be used to estimate the travel fluxes $L = \{L_{i,j}\}$ in epidemic areas through integrating multi-source real data.

1.4 Outline of the thesis

With the final aim of providing new insights into the treatment and control of complex disease and infectious disease, we investigate cell-cell signaling and infectious disease spreading in this thesis. In **Part I** we focus on uncovering the molecular mechanisms of Notch signaling from experimental data and dynamic modeling. By performing numerical simulations under different scenarios and comparing with data, we have verified our models with extensive experimental data.

In Chapter 2, we explore the molecular mechanisms of Notch receptor cis-inhibition, trans-activation and recently reported cis-activation. Our biochemical data confirm an previously unreported self-oligomerization of Notch ligands in the cell. To under-

stand the biological consequence of Notch ligands oligomerization, we have proposed a new model of Notch signaling, where the ligand monomer mediates Notch receptor trans-activation dampened by ligand dimer-mediated cis-inhibition of Notch. The applicability and dependability of our model is supported by our data and published data about Notch signaling and Notch-related tissue patterning.

In Chapter 3, we aim to understand the mechanisms of reported complex observations about endothelial cells behaviour in angiogenesis. Experimental data have shown that Notch activity contains a paradoxical component that has both negative [102] and positive [43, 62, 48, 110] effects on Dll4 expression in endothelial cells. Consequently, we develop a new dynamic model where the opposing effects of Notch activity on Dll4 expression is state-dependent. We have validated our model by recapitulating multi-source data related to endothelial cells in angiogenesis, which is difficult to explain using previous models.

In Chapter 4, we dive a little deeper into Notch signaling-related tip endothelial cells specialization in angiogensis. Here the research goal is to decipher the molecular mechanisms by which Notch ligand Jag1 stimulates sprouting angiogensis while Notch ligand Dll4 limits this process. Through a combination of experimentation and dynamic modeling, we favor a mechanism where Dll4-Jag1 heterodimerization and asymmetric affinity of Notch for the Notch ligands Dll4 and Jag1 yield the observed phenomena in angiogensis.

In Chapter 5, we have developed a multi-scale dynamic model of vasculogenesis by integrating an ordinary differential equation (ODE) model for cell-cell signaling into the Cellular Potts model (CPM) for cell motility. Based on the experimental data, cell-cell signaling affects cell behavior by mediating cell polarization. This model reproduces the vascular patterning, which is consistent with previous models. Better than previous models, we recapitulate the morphological dynamics of endothelial cells and the effects of Notch signaling on vasculogenesis.

In **Part II** of this thesis we shift our topic from modeling cell-cell signaling to modeling infectious disease spreading with the aim of disease prevention and control. Here we focus on making prospective predictions and investigating the necessary conditions for infectious disease extinction.

In Chapter 6, we first obtain a causal relationship between air pollution and respiratory infection using published data. Based on this, we develop a compartment model for respiratory infection where the transmission rate of disease is an increasing function of air quality index whose dynamics follow a Filippov system describing the state-dependant interventions against air pollutants emission. Theoretical analysis of

these models predicts that the interventions against air pollution could reduce the basic reproduction number of respiratory infection.

In Chapter 7, we explore the trade-off between lifting mobility restrictions and controlling infectious disease spreading with the aim of designing a migration matrix allowing the movement of individuals between different regions in an epidemic area. To this end, we first give a theoretical condition for the existence of a migration matrix between different regions and then develop a novel method to estimate the migration matrix. We have tested our method by retrospectively investigating the travel restrictions against COVID-19 in China.

Finally, in Chapter 8, we discuss the results of this thesis and propose ideas for future work.