

Universidade de São Paulo  
Instituto de Física

**Tópicos em sinalização celular e bioinformática:  
Princípios de funcionamento do circuito de  
sinalização Notch e aprendizagem  
supervisionada variacional de relevância**

**Marcelo Boareto do Amaral**

Orientador: Prof. Dr. Nestor Caticha

Tese de doutorado apresentada ao Instituto de Física  
para a obtenção do título de Doutor em Ciências

Banca Examinadora:

Prof. Dr. Nestor Caticha - IF/USP (Orientador)  
Prof. Dr. Said Rabbani - IF/USP  
Profa. Dra. Vera Henriques - IF/USP  
Prof. Dr. Ricardo Z. N. Vêncio - FFCLRP/USP  
Prof. Dr. Roberto Kraenkel - IFT/UNESP

São Paulo  
2015

**FICHA CATALOGRÁFICA**  
**Preparada pelo Serviço de Biblioteca e Informação**  
**do Instituto de Física da Universidade de São Paulo**

Amaral, Marcelo Boareto do

Topics in cell signaling and bioinformatics: Operating Principles of Notch Signaling Pathway and Supervised Variational Relevance Learning. São Paulo, 2015.

Tese (Doutorado) – Universidade de São Paulo. Instituto de Física. Depto. de Física Geral

Orientador: Prof. Dr. Nestor Caticha

Área de Concentração: Física Estatística

Unitermos: 1.Bioinformática; 2. Biofísica; 3. Física Computacional.

USP/IF/SBI-018/2015

University of São Paulo  
Institute of Physics

Topics in cell signaling and bioinformatics:  
Operating Principles of Notch Signaling  
Pathway and Supervised Variational Relevance  
Learning

Marcelo Boareto do Amaral

Supervisor: Prof. Dr. Nestor Caticha

PhD thesis submitted to Institute of Physics in  
fulfillment of the requirements for the Degree of  
Doctor of Science.

Examination Committee:

Prof. Dr. Nestor Caticha - IF/USP (Supervisor)  
Prof. Dr. Said Rabbani - IF/USP  
Profa. Dra. Vera Henriques - IF/USP  
Prof. Dr. Ricardo Z. N. Vêncio - FFCLRP/USP  
Prof. Dr. Roberto Kraenkel - IFT/UNESP

São Paulo  
2015



*To my parents.*

# Acknowledgments

I would like to express my special appreciation and thanks to my advisors Nestor Caticha, Eshel Ben-Jacob and José N. Onuchic for being a tremendous mentor for me, by encouraging my research and helping me to grow as a research scientist. All of you have been a source of inspiration for me. I also would like to thank my co-advisors Vitor B. P. Leite and Cecilia Clementi. Vitor has been more than a co-advisor, his advice on both research as well as on my career have been invaluable.

Most part of this thesis was developed due to a great partnership, that I hope to continue, with Mohit Jolly. I really enjoy both working and eating Indian food with him. I also thanks Jonatas Cesar and Mingyang Lu for the collaboration. It was a pleasure to work with these guys.

I'd like to thank to all of my roommates who I really enjoy living with: Shridhar Jayanthi, Brianna Kuypers, Felix Tavares and Vinicius Calsavara (Cokin). I also want to thanks all my good friends and old companions: Rafael Viegas, Rodrigo Euzébio, Renata Batista and Carolina Silvério. I also want to thanks my friends on Center for Theoretical Biology Physics (CTBP): Ricardo Santos, Weihua Zheng, Dongya Jia; my friends in Universidade de São Paulo (USP): Jonatas Cesar, Edgar Alvarenga, Bruno Pace, Alexandre Patriota; and the members of the Clube de Biologia Sintética, in particular: Andrés Ochoa, Otto Heringer, Chico Camargo, Pedro Medeiros, Pedro Pessoa, Felipe Alves, João Molino, Cauã, Joana, Maca, Lucas, Débora, Cleandro and many others.

A special thanks to my parents, to whom this thesis is dedicated. I could not express in words how grateful I am to them for all the sacrifices that they have made on my behalf. I would also like to thanks my beloved girlfriend Tatiana Amor, for being such a loving and encouraging person. I love all of you.

And lastly but not less important, I'd like to thanks to FAPESP for the finantial support, CTBP/Rice University for partial finantial support and to COSEAS for the opportunity of leaving in CRUSP.

# Abstract

In the first part of this thesis, we studied the operating principles of cell fate decisions mediated by Notch signaling pathway. This pathway have important role in cell fate determination during embryonic development, wound healing and tumorigenesis. Notch signaling is activated by binding of Notch receptor of one cell to either of its ligand- Delta or Jagged- of another cell. Notch-Delta circuit forms an intercellular toggle switch, and two neighboring cells tend to adopt different fates - Sender (high ligand, low receptor) and Receiver (low ligand, high receptor). Here, we present a new tractable theoretical framework that incorporates both Delta and Jagged in Notch signaling, and show that Notch-Delta-Jagged circuit enables an additional fate - hybrid Sender/Receiver (S/R) (medium ligand, medium receptor) and behaves as a three-way switch. Further, we found that production rates of both the ligands and the asymmetric modulation of binding affinity of Notch to its ligands by glycosyltransferase Fringe severely affects the parameter range of the existence of these states and their relative stability - high levels of Jagged, but not that of Fringe or Delta, promote hybrid S/R state and lateral induction. We elucidate the role of Jagged in cell fate determination and discuss its possible implications in understanding tumor-stroma crosstalk, which frequently entails Notch-Jagged communication. We further evaluate the interplay among Notch signaling, inflammation and Cancer Stem Cell population. We show that inflammation can expand the population of Cancer Stem Cells (CSCs) by increasing the levels of Jagged in cells that can further activate Notch signaling pathway in neighboring non-CSCs. Our results suggest that, inhibiting the production of Jagged dampens the effect of inflammation in expanding the CSC population, indicating that inflammatory signal function through Notch-Jagged signaling to increase CSCs. Our results are consistent with observations in basal-like breast cancer, where loss of Fringe and constitutive activation of NF-kB-Jag1 axis

promotes CSC population. Our computational framework can be tailored to include additional signals such as p53 and hypoxia that affect this plasticity to gain stemness, thus providing a platform that can be useful in designing novel therapies.

In the second part of the thesis, we introduce a new method for feature selection: Supervised Variational Relevance Learning (Suvrel). This is a variational method to determine metric tensors to define distance based similarity in pattern classification, inspired in relevance learning. We propose a new methodology where the metric tensor can be calculated analytically. Preprocessing the patterns by doing linear transformations using the metric tensor yields a dataset which can be more efficiently classified. We test our methods using publicly available datasets, for some standard classifiers. We also applied this methodology to study the relationship between global structural parameters and the Enzyme Commission hierarchy. Lastly, we propose a new methodology for identifying the differentially expressed genes using DNA microarray technology. Unlike traditional approaches, our methodology skips intermediate unnecessary preprocessing steps and therefore does not accumulate errors due to these analysis, resulting in a more sensitive and robust method.

# Resumo

Na primeira parte desta tese, estudamos os princípios operacionais das decisões celulares mediadas pelo sistema de sinalização Notch. Este sistema tem papel importante nas decisões celulares que ocorrem durante o desenvolvimento embrionário, cicatrização de feridas e na formação de tumores. O circuito de sinalização é ativado quando o receptor Notch de uma célula interage com um dos ligantes - Delta ou Jagged - de uma célula vizinha. O circuito Notch-Delta forma um comutador intercelular, e duas células vizinhas tendem a adotar estados diferentes - Emissor (muito ligante e pouco receptor) e Recebedor (pouco ligante e muito receptor). Neste manuscrito, apresentamos uma nova abordagem teórica que integra ambos Delta e Jagged no circuito Notch. Mostramos que o circuito Notch-Delta-Jagged permite um novo estado - um híbrido Emissor/Recebedor (E/R) com concentrações intermediárias de receptores e ligantes, e portanto o circuito é age como uma chave de três vias. Em seguida, observamos que a taxa de produção de ambos os ligantes, assim como a modulação assimétrica da afinidade de ligação do Notch com seus ligantes mediada pela glicosiltransferase Fringe, afeta severamente o intervalo de existência dos estados e sua relativa estabilidade - altos níveis de Jagged, mas não de Fringe ou Delta, promovem o estado híbrido E/R e o processo de indução lateral. Nós elucidamos o papel de Jagged na determinação dos estados celulares e discutimos sua possível implicação no entendimento da comunicação entre tumor e estroma, que frequentemente envolve comunicação via interações Notch-Jagged. Posteriormente, avaliamos a interação entre Notch, inflamação e a população de Células Cancerígenas Estaminais (CCE). Mostramos que inflamação pode expandir a população de CCE por meio do aumento dos níveis de produção de Jagged que posteriormente ativa o sistema de sinalização Notch em células vizinhas não-CCE. Nossos resultados sugerem que a inibição da produção de Jagged atenua o efeito da expansão de CCE devido a inflamação, indicando que

inflamação cresce a população de CCE via interações Notch-Jagged. Nossos resultados são consistentes com observações em câncer de mama do subtipo basal (triplo negativo), onde a perda de Fringe e a ativação constitutiva do eixo NF-kB - Jag1 promove a expansão da população de CCE. Nossa abordagem computacional pode ser adaptada para incluir circuitos adicionais tais como p53 e hipóxia, que afetam a plasticidade celular, proporcionando assim uma plataforma útil para a projeção de novas terapias.

Na segunda parte desta tese, introduzimos um novo método para seleção de características: Suvrel. Este é um método variacional, inspirado em aprendizado de relevância, para determinar tensores métricos para definição de distâncias baseadas em similaridades, para utilização em métodos de classificação. Nós introduzimos uma nova metodologia na qual o tensor métrico pode ser calculado analiticamente. O preprocessamento das características por uma transformação linear utilizando o tensor métrico calculado via Suvrel melhora a eficiência dos classificadores. Testamos nosso método para conjuntos de dados públicos, utilizando os classificadores mais comumente utilizados. Nós também aplicamos esta metodologia no estudo da relação entre parâmetros estruturais globais e o sistema de classificação de função enzimática. Por último, introduzimos uma nova metodologia para a identificação de genes diferencialmente expressos utilizando a tecnologia de microarranjos de DNA. Diferentemente das abordagens tradicionais, nossa metodologia evita passos intermediários de preprocessamento que são desnecessários e devido a isto não acumula erros destas análises, o que resulta em um método mais sensível e robusto.

# Contents

Preface	xii
<b>I - Topics in cell signaling</b>	<b>1</b>
Introduction	2
1 Theoretical framework for Notch signaling	7
1.1 Mathematical models for Notch-Delta signaling . . . . .	7
1.1.1 The Collier model . . . . .	7
1.1.2 Model of mutual inactivation of Notch and Delta . . . . .	8
1.2 A model for Notch-Delta-Jagged circuit . . . . .	9
1.3 A model for Notch-Delta-Jagged-Fringe circuit . . . . .	11
1.4 Considering the effect of both soluble and membrane-bound ligands . . . . .	12
1.5 Parameters' values . . . . .	13
1.6 Conclusion . . . . .	14
2 Notch signaling: two ligands, same signal and different behaviours.	16
2.1 Cell fate decisions in Notch-Delta, Notch-Jagged and Notch-Delta-Jagged signaling . . . . .	17
2.1.1 Notch-Delta (ND) circuit: a two-cell toggle switch . . . . .	17
2.1.2 Notch-Jagged (NJ) circuit . . . . .	20
2.1.3 Notch-Delta-Jagged (NDJ) circuit: a three-way switch. . . . .	21
2.2 The effect of Fringe mediated asymmetric Notch-ligand binding . . . . .	25
2.2.1 The effect of Delta/Jagged asymmetry on the cell-cell fate modulation	27

2.2.2	Fringe promotes lateral inhibition patterning . . . . .	29
2.2.3	Fringe alters the relative stability of the steady states of Notch-Delta-Jagged signaling . . . . .	30
2.3	Production rates of the two ligands affect cell-fate decision antagonistically .	33
2.3.1	Ligand production rates control tissue level patterning . . . . .	35
2.4	Soluble ligands affect cell-fate decision in a paracrine manner . . . . .	37
2.5	Discussion . . . . .	39
2.5.1	The crucial role of hybrid S/R state in tumor progression . . . . .	39
<b>3</b>	<b>Inflammation gives rise to mutually reinforcing Cancer Stem Cells through Notch-Jagged signaling</b>	<b>44</b>
3.1	Model Formulation . . . . .	46
3.2	Inflammation promotes the hybrid Sender/Receiver phenotype by increasing Notch-Jagged signaling over Notch-Delta signaling . . . . .	47
3.3	Self-activation of Jagged amplifies the effect of inflammation . . . . .	49
3.4	Fringe acts antagonistic to inflammation and inhibits Notch-Jagged signaling	52
3.5	Inflammation increase CSC population through Notch-Jagged signaling . .	53
3.6	Discussion . . . . .	56
<b>Appendix I</b>		<b>59</b>
I.A -	Derivation of the model with modulation by Fringe . . . . .	59
I.B -	Parameter sensitivity analysis . . . . .	61
<b>II - Topics in bioinformatics</b>		<b>63</b>
<b>4</b>	<b>Supervised Variational Relevance Learning, an analytic geometric feature selection with applications to omic datasets</b>	<b>64</b>
4.1	Methods . . . . .	66
4.2	Results . . . . .	70
4.3	Discussion and Conclusions . . . . .	74
4.3.1	Tables . . . . .	76

<b>5 Relationship between global structural parameters and Enzyme Commission hierarchy: implications for function prediction</b>	<b>82</b>
5.1 Introduction . . . . .	83
5.2 Methods . . . . .	84
5.2.1 Parameter selection . . . . .	86
5.3 Results and Discussion . . . . .	87
5.4 Conclusion . . . . .	91
<b>6 T-test at the probe level: An alternative method to identify the statistically significant genes for microarray data</b>	<b>92</b>
6.1 Introduction . . . . .	93
6.2 Methodology . . . . .	96
6.3 Results . . . . .	97
6.3.1 Sensitivity . . . . .	97
6.3.2 Robustness . . . . .	99
6.4 Discussion . . . . .	100
<b>Appendix II</b>	<b>103</b>
II.A Positivity of the metric tensor . . . . .	103
II.B Off-diagonal terms . . . . .	103
<b>Bibliography</b>	<b>105</b>

# Preface

Multicellular organisms start as a single-cell that divides into hundreds or trillions of cells. Throughout this process, the cells need to communicate and make a number of crucial decisions for the proper development of the organism. The initial decision during embrionic development is the establishment of the major axes of the organism, followed by many other complex decisions that will assembly the tissues and organs. Surprisely, a small set of highly conserved signaling pathways mediates all those decisions.

Understanding the operating principles of those pathways is important not only to answer fundamental questions concerning animal development but also to understand the development of diseases such as cancer. Cancer cells, similar to the cells during embrionic development, use intricate modes of communication to sustain successful collective behavior and growth. These cells communicate among themselves, with the surrounding stromal tissue and even with the immune system in order to coordenate sustained angiogenesis, to evade apoptosis and proliferate. For example, cancer cells use juxtacrine Notch signaling to increase angiogenesis and increase the population of Cancer Stem Cells (CSCs), the paracrine HGF/SF system to regulate cell growth, motility, and morphogenesis, and the autocrine IL-6 system to modulate cellular proliferation. Therefore, breaking the code of cell-to-cell communication may yield cancer-fighting drugs, similar to drugs targeted against bacterial communication that are now used to fight bacteria. For example, drugs that interfere with bacterial quorum sensing is used to target tuberculosis.

The first part of this thesis focus on the study of the mechanisms of Notch signaling pathway. This pathway mediates cell-to-cell communication by contact and plays crucial roles in cell fate determination during embryonic development and cancer progression. We developed a new theoretical framework to elucidate the operating principles of Notch signal-

ing [1, 2]. We further studied the interplay between Notch signaling and inflammation [3]. Our results suggest that inflammation can increase the population of Cancer Stem Cells via Notch signaling. To support our hypothesis, we used bioinformatics techniques to analyze experimental data from public databases. Bioinformatics is the focus of the second part of this thesis. We first introduce a new method for feature selection: Suvrel (Supervised Variational Relevance Learning) [4]. We show that this variational method can be solved exactly analytically and can improve class prediction. We further applied this method to studied an enzyme classification problem [5]. Lastly, we present a new methodology to identify the differential expressed genes using DNA microarrays data [6].

## Part I

# Topics in cell signaling

*Houston, we solve a problem.*

# Introduction

The emergence of multicellular organisms is considered to be one of the major evolutionary events in life on Earth. Multicellular organisms can not be described as a superimposition of many fundamental cells (units of life) but as a complicated interconnected cluster of cells that must communicate, coordinate and organize themselves. The emergence of the organs is a result of high levels of differentiation and organization of cells, that requires sophisticated intercellular communication mechanisms. Therefore, cell-cell communication is crucial for the development and evolution of multicellular organisms. Interestingly, in metazoans, less than 20 signaling pathways are required to generate all diversity of cell types and patterns, and among them only seven control most interactions that happens during the organism development. These highly conserved signaling pathways are: Wnt; TGF-beta; Hedgehog; Receptor Tyrosine Kinase (RTK); Jak/STAT; nuclear hormone receptor; and Notch [7]. During development, these pathways are used to establish polarity and body axes, coordinate pattern formation and choreograph morphogenesis.

Among the most important signaling pathways, Notch stands for its simplicity in design and versatility in function. Despite having a small number of core signaling components, Notch signaling affects a wide spectrum of cell differentiation decisions in metazoan species [8]. Notch have a central role in cell fate decisions and regulate the differentiation of epithelial, neuronal, blood, bone, muscle and endothelial cells. Additionally, Notch signaling is involved in other biological processes such as arterial cell fate determination during embryonic development and the control of angiogenesis by regulating the selection of endothelial tip and stalk cells in sprouting blood vessels. Numerous studies have demonstrated that the aberrant gain or loss of Notch signaling pathway components has been involved in many genetic disorders such as Alagille syndrome. The Notch pathway has also been implicated

in the tumorigenesis of several types of tumor malignancies including, lung adenocarcinoma, melanoma, ovarian carcinoma, breast cancer, among others [9].

This pathway consists of the transmembrane receptor Notch and the transmembrane ligands Delta and Jagged. In mammals, there are four Notch receptors (Notch1-4) and five ligands (Jag1, Jag2, Delta-like 1 (Dll1), Dll3 and Dll4). When Notch (any of the four members) of one cell interacts with Delta or Jagged (any of the five members) of the same cell, it leads to degradation of both of them; this mechanism is known as *cis*-inhibition. Conversely, when Notch of one cell interacts with Delta or Jagged of the neighboring cell (*trans*-interaction), the Notch receptor is cleaved and the Notch Intracellular Domain (NICD) is released into the cytoplasm. NICD then enters the nucleus and activates many downstream target genes - therefore activating the Notch pathway [8] .

Although Delta and Jagged generate the same signal (NICD), the dynamics of Notch-Delta and Notch-Jagged signaling are quite different, because NICD regulates the production of the two ligands asymmetrically - it inhibits Delta but activates Jagged (Figure 1A). Therefore Notch and Delta form an intercellular double negative feedback loop, but Notch and Jagged form an intercellular double positive feedback loop. Consequently, the two cells interacting via Notch-Delta adopt different fates - one cell behaves as a Sender (S) with (high Delta (ligand), low Notch (receptor)) on its surface; and the other behaves as a Receiver (R) with (low Delta (ligand), high Notch (receptor)) on its surface. This cell-fate diversification mechanism is known as lateral inhibition and can lead to checkerboard-like or “salt-and-pepper” patterns (Figure 1B) as observed during bristle patterning in flies, inner ear patterning in vertebrates, and neurogenesis in both flies and vertebrates [10]. On the other hand, two cells interacting via Notch-Jagged can act as both Sender and Receiver as they have (high Jagged (ligand), high Notch (receptor)) on their surface. This cell-fate convergence mechanism is known as lateral induction (Figure 1C) and is crucial during cardiac development, inner ear development, and the formation of a smooth muscle wall around a nascent artery [11, 12, 13, 14].

Another level of asymmetry that modulates the dynamics of Notch signaling via Delta vs. Jagged is the effect of a downstream target of NICD - the glycosyltransferase Fringe [15] that can increase the binding affinity of Notch for Delta, but decrease that for Jagged, both for *cis*- and *trans*-interactions [16, 17, 18, 19]. These two levels of asymmetry in Notch-Delta

vs. Notch-Jagged signaling call for elucidating their different operating principles.

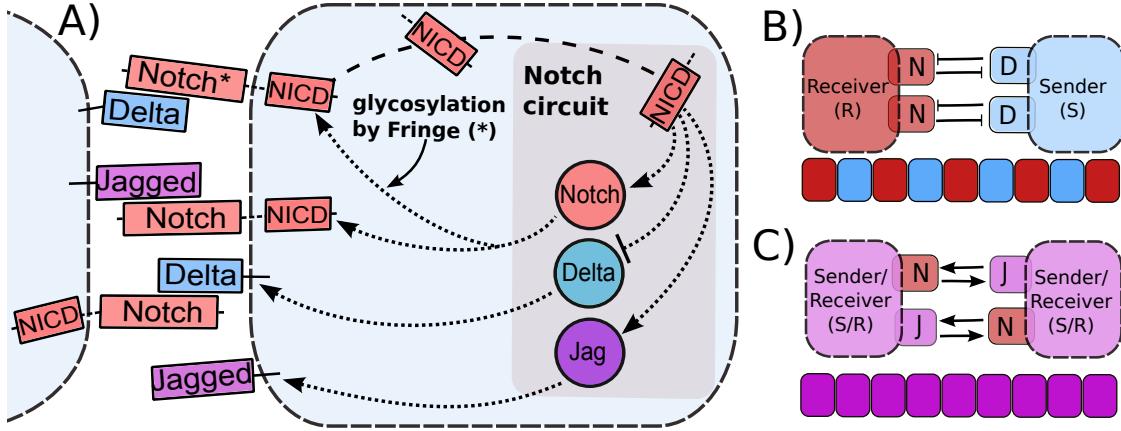


Figure 1: Overview of intra-cellular and inter-cellular Notch signaling pathway, and tissue patterning outcomes. A) Notch, the transmembrane receptor of one cell, binds to Delta or Jagged, the transmembrane ligands belonging to the neighboring cell. This trans-interaction leads to the cleavage and release of Notch Intracellular Domain (NICD) that migrates to the nucleus and modulates the gene expression of many Notch pathway targets. This modulation also indirectly leads to the transcriptional activation of Notch and Jagged and inhibition of Delta. Interaction between Notch receptor and ligands (Delta or Jagged) of the same cell (cis-interaction) leads to the degradation of both the receptor and the ligand. Glycosylation of Notch by Fringe modifies Notch to have higher affinity for binding to Delta and lower affinity for binding to Jagged. Due to this asymmetric regulation of Delta and Jagged by NICD. B) Notch-Delta signaling forms an intercellular toggle switch and alternate cells adopt distinct fates - Sender (low Notch (receptor), high Delta (ligand)) and Receiver (high Notch (receptor), low Delta (ligand)), giving rise to a checkerboard-like pattern (lateral inhibition), whereas C) Notch-Jagged signaling forms an intercellular double positive feedback loop and the cells adopt similar fates (lateral induction - (high Notch (receptor), high Jagged (ligand))).

Much experimental and theoretical research efforts have been directed towards understanding the Notch-Delta-dependent cell fate determination [10, 20, 21, 22, 23, 24]. In contrast, the role of Notch-Jagged signaling has gained limited research attention despite its crucial role during embryonic development, and promoting tumor progression in multiple ways. For example, overexpression of Jagged has been associated with poor prognosis at least in breast cancer and prostate cancer [25], thus highlighting the importance of understanding its role in Notch signaling. Other recent studies have shown that Notch signaling can be activated not only in a juxtacrine manner but also in a paracrine manner, by soluble forms of the ligands Jagged and Delta [26, 27, 28]. The soluble Jagged and Delta have different effects

on tumor progression - soluble Delta inhibits tumor growth [29, 30], while soluble Jagged strongly aggravates the malignant progression of cancer. Also, transmembrane Jagged1 can initiate metastasis by inducing cancer cells to undergo Epithelial to Mesenchymal Transition (EMT) - a dormant embryonic program that allows them to migrate and invade. It also facilitates the colonization of the Circulating Tumor Cells (CTCs) by enabling cell-cell communication between CTCs and the cells of the organ where they settle down and develop metastases [31, 9] that causes more than 90% of cancer-related deaths [32]. Furthermore, Notch-Jagged signaling can also maintain and/or expand the population of Cancer Stem Cells - highly plastic cancer cells that can initiate a new tumor [33, 27], thereby promoting tumor relapse. Notably, Notch-Jagged signaling also plays a crucial role in angiogenesis [34], cancer metastasis [9], and rapid development of cancer chemotherapy and radiation therapy resistance [35]. Not surprisingly, poor survival and cancer recurrence is associated with high Jagged1 levels in patients [31]. Hence, understanding Notch-Jagged signaling is essential to control tumor spread.

We devise a new theoretical framework that incorporates Jagged in addition to Delta and the asymmetric regulation of the proteins by the signal NICD that activates Notch and Jagged but represses Delta. We also included the second asymmetry between the ligands due to the effect of the glycosyltransferase Fringe and show that different outcomes are generated depending on which ligand is dominant. We elucidate the role of Jagged in cell fate determination and discuss its possible implications in understanding tumor-stroma crosstalk, which frequently entails Notch-Jagged communication. We also show that inflammatory signal can expand the CSC population by augmenting Notch-Jagged signaling, presenting the first step towards understanding the molecular interplay of Notch-Jagged signaling, inflammation and CSC population.

## Overview

In Chapter 1, we revise previous theoretical models for Notch signaling and introduce our theoretical framework. A computational model that includes both ligands Delta and Jagged can be crucial to understand the developmental systems where both lateral inhibition and lateral induction are acting simultaneously, such as during angiogenesis and pancreatic cell

differentiation. During angiogenesis, Jagged mediates the number of ‘stalk’ cells that may separate two ‘tip’ cells [36], and the three fates that can be obtained in the context of Notch-Delta-Jagged signaling may underlie the highly dynamic phenotypes that endothelial cells can attain that are neither entirely either tip nor stalk [37]. Both lateral inhibition and lateral induction are also critical for multistability of the pancreatic system, wherein the loss of either of them can lead to lineage switching between various cell types [38].

In Chapter 2, we used our theoretical framework to elucidate the operating principles of Notch signaling. We find that the asymmetry between the modulations of Delta and Jagged leads to the existence of previously unexplored possibility of a new Sender/Receiver phenotype enabling two interacting cells to share similar fate. We show that higher production rate of Jagged, but not that of Delta, expands the range of parameters for which both cells attain the hybrid S/R state. Conversely, glycosyltransferase Fringe reduces this range of conditions, and reduces the relative stability of the hybrid S/R state, thereby promoting cell-fate divergence and consequently lateral inhibition-based patterns. Lastly, soluble Jagged drives the cells to attain the hybrid S/R state, and soluble Delta drives them to be Receivers. We also discuss the absolutely critical role of hybrid S/R state in promoting cancer metastasis by enabling collective cell migration and expanding the Cancer Stem Cell (CSC) population.

In Chapter 3, we incorporate the effect of cytokines in our theoretical framework in order to study the interplay between inflammatory signals and Notch signaling. We found that inflammation expands the hybrid S/R state of cells (extends the range of physiological conditions for which cells can both send and receive signals). This effect of cytokines is amplified by indirect self-activation of Jagged via NF- $\kappa$ B, but inhibited by Fringe that increases the binding affinity of the Notch receptor for Delta and decreases that for Jagged. We further show that inflammation can expand the population of CSCs by increasing the levels of Jagged in cells that can further activate Notch signaling pathway in neighboring non-CSCs. Inhibiting the production of Jagged dampens the effect of inflammation in expanding the CSC population, indicating that these cytokines function through Notch-Jagged signaling to increase CSC population.

# Chapter 1

## Theoretical framework for Notch signaling

Biological experiments are many times very difficult to interpret due to the high complexity of genetic regulatory networks which often involve counter-intuitive feedback mechanisms. Because of that, modeling approaches are fundamental to provide understanding and insights on how different elements operate. *In silico* approaches can offer complementary understanding of experimental studies as well as enable the formulation of new predictions to guide new experiments. In this chapter, we present a brief overview of previous models for Notch signaling pathway and introduce our theoretical framework.

### 1.1 Mathematical models for Notch-Delta signaling

#### 1.1.1 The Collier model

One of the first theoretical model for lateral inhibition was proposed by Collier et al [22] in 1996. Their simplified framework is represented by two equations per cell: one that represents the dynamics of Notch and the other that represents the dynamics of Delta. The basic assumption of the model is that Delta of the neighboring cells activates Notch signaling of the cell which in turn activates the production of Notch and inhibits the production of Delta - therefore creating an intercellular negative feedback loop between neighboring cells. The dynamics of Notch and Delta in a cell is given by:

$$\frac{dN}{dt} = N_0 H^+(D_{ext}) - \gamma N \quad (1.1)$$

$$\frac{dD}{dt} = D_0 H^-(N) - \gamma D \quad (1.2)$$

where  $N_0$  and  $D_0$  represents the production rate of Notch and Delta, respectively and  $\gamma$  represents the degradation of both Notch and Delta. The variable  $D_{ext}$  represents the amount of Delta of the neighboring cells available for binding and the positive hill function  $H^+(D_{ext})$  represents the increase in the production of Notch with the amount of external Delta ( $D_{ext}$ ) and the negative Hill function  $H^-(N)$  represents the inhibition of the production of Delta by Notch. Hill functions are defined as:  $H^+(x) = \frac{x^n}{x_0^n + x^n}$  and  $H^-(x) = \frac{1}{x_0^n + x^n}$ .

This model brings interesting insights about patterning formation by Notch-Delta signaling circuit and many extensions of this model have been developed over the past years [12, 39, 40, 41]. One particularly interesting extension of Collier's model was proposed by Petrovic et al [12]. Their model was the first framework to include both ligands Delta and Jagged, and their simulations was useful to understand some of their experimental results in the context of inner ear development [12].

### 1.1.2 Model of mutual inactivation of Notch and Delta

Experimental studies have suggested that the Notch receptor and the ligand Delta mutually inactivate each other when interacting in the same cell, in a mechanism called *cis*-inhibition [21, 42, 43, 44]. Sprinzak et al [21] measured experimentally the combined *cis-trans* input-output relationship for the Notch-Delta system. Their results revealed that *cis* and *trans* interactions have different responses: *trans* activation generate a graded response while *cis*-interaction is sharp and occurs at a fixed threshold. They also developed a simple mathematical model based on their experimental measurements that takes this feature in account. Within their framework [21], Notch receptor ( $N$ ) belonging to one cell can interact with Delta ( $D$ ) of the same cell - known as *cis*-interaction, or with those of the neighboring cell - Delta ( $D_{ext}$ ) - known as *trans*-interaction. The *cis*-interaction, also referred as *cis*-inhibition, causes the degradation of both the interacting proteins. On the other hand, the *trans*-interaction, also referred as *trans*-activation, leads to the cleavage of Notch receptor which releases Notch Intracellular Domain (NICD; represented as  $I$  in the model). In this

framework, the equation for the dynamics of Notch ( $N$ ), Delta ( $D$ ) and NICD ( $I$ ) are given by:

$$\frac{dN}{dt} = N_0 - k_C ND - k_T ND_{ext} - \gamma N \quad (1.3)$$

$$\frac{dD}{dt} = D_0 - k_C DN - k_T DN_{ext} - \gamma D \quad (1.4)$$

$$\frac{dI}{dt} = k_T ND_{ext} - \gamma_I I. \quad (1.5)$$

where  $\gamma$  represents the degradation rate of both Notch and Delta and  $\gamma_I$  the degradation of NICD.  $k_C$  and  $k_T$  are the strengths of *cis*-inhibition and *trans*-activation respectively; and  $N_0$  and  $D_0$  are the production rates of Notch and Delta respectively.  $N_{ext}$  and  $D_{ext}$  represent the amount of protein available for binding.

Based on Sprinzak et al results, many theoretical studies have shown that *cis*-inhibition facilitate lateral inhibition patterning, enable faster dynamics and provide greater robustness to pattern formation [20, 45]. Recently, *cis*-inhibition has also been reported for interaction between Notch and Jagged [16], however its role in determining cell-fate dynamics for the combined Notch-Delta-Jagged circuit remains elusive.

## 1.2 A model for Notch-Delta-Jagged circuit

We generalized the theoretical framework devised by Sprinzak et al. [21] by incorporating Jagged in addition to Delta and including the feedback effects of NICD that indirectly activates Notch and Jagged, and represses Delta, thereby creating an asymmetry between Notch-Delta and Notch-Jagged interactions (Figure 1.1). The deterministic equations for the dynamics of Notch ( $N$ ), Delta ( $D$ ), Jagged ( $J$ ) and NICD ( $I$ ) are given by:

$$\frac{dN}{dt} = N_0 H^{S+}(I) - k_C N(D + J) - k_T N(D_{ext} + J_{ext}) - \gamma N \quad (1.6)$$

$$\frac{dD}{dt} = D_0 H^{S-}(I) - k_C DN - k_T DN_{ext} - \gamma D \quad (1.7)$$

$$\frac{dJ}{dt} = J_0 H^{S+}(I) - k_C JN - k_T JN_{ext} - \gamma J \quad (1.8)$$

$$\frac{dI}{dt} = k_T N(D_{ext} + J_{ext}) - \gamma_I I \quad (1.9)$$

where  $\gamma$  represents the degradation rate of all three transmembrane proteins Notch, Jagged and Delta, and  $\gamma_I$  the degradation of NICD.  $k_C$  and  $k_T$  are the strengths of *cis*-inhibition and *trans*-activation respectively; and  $N_0$ ,  $D_0$ , and  $J_0$  are the production rates of Notch, Delta and Jagged respectively.  $N_{ext}$ ,  $D_{ext}$  and  $J_{ext}$  represent the amount of protein available for binding - which can be on the membrane surface of neighboring cells or in a soluble form. Experimental evidence suggest that membrane-bound ligands can generate a stronger signal when compared with soluble forms [46]. The distinction between these two forms of ligand - membrane-bound and soluble - is addressed below (see Eqs 1.16-1.19). We consider shifted Hill functions [47] to represent the effect of NICD ( $I$ ) on the production rates of the proteins. Shifted Hill functions are defined as  $H^S(I, \lambda) = H^-(I) + \lambda H^+(I)$  or in simpler notation:  $H^{S+}(I)$  if  $\lambda > 1$  and  $H^{S-}(I)$  if  $\lambda < 1$ , where the weight factor  $\lambda$ , represents the fold-change in production rate, therefore, for activation,  $\lambda > 1$ ; for repression,  $\lambda < 1$ ; and for no effect,  $\lambda = 1$  ( $\lambda_N, \lambda_J > 1$  and  $\lambda_D < 1$  in our model). For the case of two interacting cells, the variables  $N_{ext}$ ,  $D_{ext}$  and  $J_{ext}$  should be replaced by  $N$ ,  $D$ ,  $J$  of the neighboring cell.

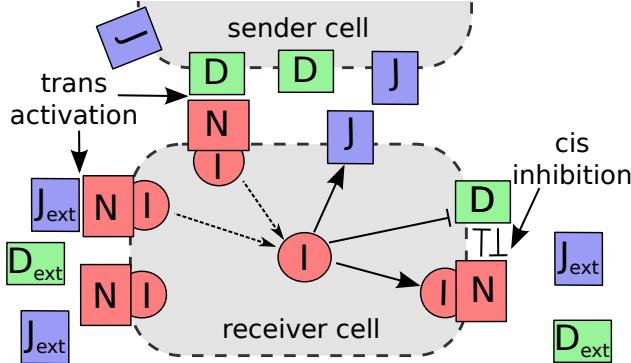


Figure 1.1: Schematic illustration of Notch signaling circuit. NICD ( $I$ ) is released when the receptor ( $N$ ) of the receiver cell interacts with the ligand of the sender cell ( $D$  or  $J$ ) or with external ligands in a soluble form ( $D_{ext}$  or  $J_{ext}$ ) - so called *trans*-activation. The released signal activates the expression of  $N$  and  $J$ , and inhibits the expression of  $D$ . The *cis*-inhibition occurs between the receptor and ligand in the same cell and leads to the degradation of both proteins.

The Notch signal (NICD) degrades rapidly as compared to N, D and J [48]. Therefore, we can assume a quasi-steady approximation for it. In this case, the equations 1.6-1.9 can be reduced to two equations by defining  $L = D + J$  and  $L_{ext} = D_{ext} + J_{ext}$ , where  $L$  represents the total amount of both ligands in the cell, and  $L_{ext}$  denotes the total amount of external

ligands available to bind. Then, the reduced system is:

$$\frac{dN}{dt} = N_0 H^{S+}(I_{eq}) - k_C NL - k_T NL_{ext} - \gamma N \quad (1.10)$$

$$\frac{dL}{dt} = D_0 H^{S-}(I_{eq}) + J_0 H^{S+}(I_{eq}) - k_C LN - k_T LN_{ext} - \gamma L \quad (1.11)$$

where, by quasi steady-state approximation,  $I_{eq} = k_T NL_{ext}/\gamma_I$ .

### 1.3 A model for Notch-Delta-Jagged-Fringe circuit

Glycosylation of Notch by Fringe creates additional asymmetry between Delta and Jagged by modulating the binding affinity of the two ligands to Notch - the glycosylated Notch has a higher affinity to bind to Delta, but lower affinity to bind to Jagged [18, 19]. To incorporate this mechanism within our framework, we considered two distinct sub-populations of Notch - the one modified by Fringe, and the other unmodified. Since NICD, which is represented in the model by  $(I)$ , activates Fringe [15], we have taken the fraction of glycosylated Notch (denoting the effect of Fringe on Notch) to increase with  $(I)$ . This glycosylated Notch has different strengths of *cis*-inhibition and *trans*-activation for Delta and for Jagged [16]. Thus, while representing effective Notch (sum of glycosylated and unglycosylated Notch), we consider the strengths of *cis*-inhibition and *trans*-activation of Notch for Delta and for Jagged to depend on  $(I)$  (see derivation of the model in the Appendix A). The resulting model for one cell is given by:

$$\frac{dN}{dt} = N_0 H^{S+}(I) - N(k_{C_D} D + k_{T_D} D_{ext} + k_{C_J} J + k_{T_J} J_{ext}) - \gamma N \quad (1.12)$$

$$\frac{dD}{dt} = D_0 H^{S-}(I) - k_{C_D} ND - k_{T_D} DN_{ext} - \gamma D \quad (1.13)$$

$$\frac{dJ}{dt} = J_0 H^{S+}(I) - k_{C_J} NJ - k_{T_J} JN_{ext} - \gamma J \quad (1.14)$$

$$\frac{dI}{dt} = N[k_{T_D} D_{ext} + k_{T_J} J_{ext}] - \gamma_I I \quad (1.15)$$

where  $k_C(I)$  and  $k_T(I)$  are now functions of the signal NICD given by:  $k(I) = k[1 + aH^+(I)] = kH^S(I, \lambda^F)$  where  $\lambda_F = 1 + a$ . The shifted Hill function  $H^S(I, \lambda^F)$  represents the increase of the Fringe effect on the binding asymmetry with the increase of  $(I)$  and the

parameter  $\lambda^F$  represents the increase ( $\lambda^F > 1$ ), decrease ( $\lambda^F < 1$ ) of both *trans*-activation and *cis*-inhibition rate due to glycosylation. Experimental evidence suggest that  $\lambda_D^F > 1$  and  $\lambda_J^F < 1$  [18, 19], representing the increase of the binding affinity between Notch and Delta, and the decrease of that between Notch and Jagged.

Our models shows a good robustness with respect to changes in parameter values as discussed in the appendix B. All the codes were developed in python using the PyDSTool [49].

## 1.4 Considering the effect of both soluble and membrane-bound ligands

Experimental evidence suggests that membrane-bound ligands should activate the signal strongly compared to soluble ligands. This happens because soluble ligand does not have enough mechanical pulling force to activating the signal [50]. However, alternative mechanisms such as ligand multimerization can lead to sufficient mechanical force for ligand activation [51]. Another evidence that membrane-bound should activate the signal strongly is that lower lateral mobility of the ligand leads to higher signaling [46]. Therefore, a model that considers both soluble and membrane-bound ligands should have two terms in Notch equation for interaction with both forms of these ligands. For example, the model presented in Eqs. 1.6-1.9 would be:

$$\frac{dN}{dt} = N_0 H^{S+}(I) - k_C N(D + J) - k_T^m N(D_{ext}^m + J_{ext}^m) - k_T^s N(D_{ext}^s + J_{ext}^s) - \gamma N \quad (1.16)$$

$$\frac{dD}{dt} = D_0 H^{S-}(I) - k_C DN - k_T^m DN_{ext} - \gamma D \quad (1.17)$$

$$\frac{dJ}{dt} = J_0 H^{S+}(I) - k_C JN - k_T^m JN_{ext} - \gamma J \quad (1.18)$$

$$\frac{dI}{dt} = N[k_T^m (D_{ext}^m + J_{ext}^m) + k_T^s (D_{ext}^s + J_{ext}^s)] - \gamma_I I \quad (1.19)$$

where  $D_{ext}^m$  and  $J_{ext}^m$  represents membrane-bound Delta and Jagged respectively, and  $D_{ext}^s$  and  $J_{ext}^s$  represents soluble Delta and Jagged respectively.  $k_T^m$  and  $k_T^s$  represents the trans-activation rate for membrane-bound ligands and for soluble ligands respectively, and its expected that  $k_T^m > k_T^s$ .

## 1.5 Parameters' values

Our model is based on the model proposed by Sprinzak et al [21] which well-fitted their experimental results designed to measure the interaction between Notch and Delta, both in cis and trans. Their experimental results suggest a Hill coefficient for Notch activity close to 2 and this value was chosen to represent the NICD activation of Notch ( $n_N$ ) and inhibition of Delta ( $n_D$ ). For the case of NICD modulation of Jagged, a higher Hill coefficient ( $n_J = 5$ ) was chosen in order to represent both direct activation of Jagged by NICD and indirect modulations by miRNA - Jagged is strongly repressed by miR200 [52], however, miR200 is repressed by SNAIL [53] which in turn is activated by NICD [11], resulting in a effective strong activation of Jagged by the signal. A strong activation of Jagged is required for the maintenance of the (S/R). For  $n_J = 4$  the range of existence of the S/R state is significantly decreased and for  $n_J = 3$  this state is no longer observed. Previous work [21], shows that cis-inhibition rate ( $k_C$ ) is approximately ten times higher than trans-activation rate ( $k_T$ ) for Notch-Delta interaction and we assumed the same values for Notch-Jagged interaction when the effect of Fringe is not taken into account. Unlike the previous model [21] in which the rates are in units of Relative Fluorescence Units, our variables represent the number of proteins in the membrane for Notch, Delta and Jagged, and the number of proteins inside the nucleus for the signal (NICD). Because of that, we scaled the values of cis-inhibition and trans-activation rate accordingly.

The rapid degradation of the signal is an important feature in the Notch signaling [48] and because of that, we considered the NICD degradation rate ( $\gamma_I$ ) to be 5 times higher than the typical protein degradation rate  $\gamma = 0.1h^{-1}$ . The values of  $N_0$ ,  $D_0$  and  $J_0$  which represents the production rate of the proteins were chosen in order to keep the maximum number of proteins in the membrane up to approximately 5000 per cell. This value is consistent with experimental results where the concentration of the proteins varies up to a few hundreds  $ng/ml$  [54] - or a few thousands of proteins per cell. Once most proteins are exported to the membrane, we should expect a few thousands of proteins in the membrane when the expression is up-regulated. Similarly, the number of NICD inside the nucleus varies up to a few hundreds and because of that, we select the threshold of the Hill function ( $I_0$ ) to be 200.

For the effect of glycosylation of Notch by Fringe, we assumed that the fraction of glycosylated Notch increases with the signal (NICD), represented by a Hill coefficient ( $n_F = 1$ ). We

assumed that glycosylation of Notch increases its affinity to Delta in a factor of 3 ( $\lambda_D^F = 3.0$ ) and decrease its affinity to Jagged in 70% ( $\lambda_J^F = 0.3$ ). Experimental evidences for modulation of Notch by lunatic Fringe shows a 4.4 fold reduction for Jagged1-mediated signaling and at least two fold increase for Delta1-mediated signaling [18]. Finally,  $\lambda_N = \lambda_J = 2.0$  for Notch and Jagged represents an activation of their production by the signal and  $\lambda_D = 0.0$  for Delta represents its inhibition by the signal.

Table 1.1: *Parameters values used in the simulations, unless indicated otherwise.*

parameter	value	unity
$\gamma$	0.1	$time^{-1} (h^{-1})$
$\gamma_I$	0.5	$time^{-1} (h^{-1})$
$N_0$	from 500 to 1400	# of proteins
$D_0$	from 1000 to 1800	# of proteins
$J_0$	1200	# of proteins
$k_T$	$5e^{-5}$	$time^{-1} (h^{-1})$
$k_C$	$5e^{-4}$	$time^{-1} (h^{-1})$
$I_0$	200	# of proteins
$n_N, n_D$	2.0	dimensionless
$n_J$	5.0	dimensionless
$n_F$	5.0	dimensionless
$\lambda_N, \lambda_J$	2.0	dimensionless
$\lambda_D$	0.0	dimensionless
$\lambda_D^F$	3.0	dimensionless
$\lambda_J^F$	0.3	dimensionless

## 1.6 Conclusion

We present the first step towards including the role of Jagged in cell fate determination. Jagged-mediated signaling indicates an evolutionarily need to implicate different repertoire of responses in cell-cell communication, and has been shown to be critical in mammalian embryonic development as well as tumor progression. A better understanding of Notch-Delta-Jagged signaling, which is affected by various signals in the tumor microenvironment

[36], can provide valuable clues how to target cancer survival by interfering with the tumor-stroma cross-talk.

# Chapter 2

## Notch signaling: two ligands, same signal and different behaviours.

Notch signaling through Jagged, and that through Delta have different dynamics because of two elements of asymmetry in the signaling circuit. First, NICD inhibits Delta through its downstream effector Hes1 [55], but activates Jagged both directly [13] and indirectly through miR-200 (as discussed in the previous chapter). These modes of regulation effectively create an intercellular double negative feedback loop between Notch and Delta [10], but an intercellular double positive feedback loop between Notch and Jagged [13]. Consequently, Notch-Delta signaling between two cells behaves as a two-way switch: one cell has (high Delta (ligand), low Notch (receptor)) expression on its surface, whereas the other cell has (high Notch (receptor), low Delta (ligand)) on its surface. According to common terminology, the first cell behaves as a Sender (S) and the second one as a Receiver (R). In other words, the Notch-Delta signaling stand alone causes the two neighboring cells to acquire opposite fates in a mechanism known as lateral inhibition, is implicated, for example, in control of neurogenesis in Drosophila and vertebrates [56], and in ‘salt-and-pepper’ patterns observed during wing vein formation [10]. On the other hand, for stand alone Notch-Jagged signaling between two cells, Notch and Jagged levels in both cells go hand in hand - (high Notch, high Jagged). Therefore, both cells can act as both Sender (S) and Receiver (R) - or the two cells acquire similar fates. This mechanism, known as lateral induction, is implicated,

for example, in mammalian inner ear development [14, 12], control of epidermal stem cell clusters [57], as well as inner cardiac development [11]. Therefore, Delta and Jagged affect the collective cell fate decisions in a group of cells quite differently.

The second asymmetry between signaling through the ligands Delta and Jagged arises due to post-translational modifications of Notch that modulate the binding of Notch to Delta and to Jagged. Fringe, a glycosyltransferase, can decrease the affinity of Notch to bind to Jagged, but increase the affinity of Notch to bind to Delta [58]. Consequently, Fringe creates two distinct populations of Notch on the cell surface: one that has comparable binding affinity to both Jagged and Delta, and one that strongly prefers binding to Delta.

In this chapter, we investigate the effects of these two elements of asymmetry in Notch signaling. We first evaluate the dynamics of the circuit when only one ligand is available. In this case, the binding affinity of Notch to both ligands are the same and the only difference between Delta and Jagged is the modulation of their production by the signal (NICD), which repress Delta but activates Jagged. We also evaluate the dynamics of the combined circuit, when both ligands are available for binding. Next, we introduce the effect of Fringe, which modifies the binding affinity between Notch and the ligands. In a further analysis, we evaluate the effect of external signals which can modify the production levels of the ligands. Lastly, we evaluate the effect of soluble ligands in the dynamics of interacting cells.

## 2.1 Cell fate decisions in Notch-Delta, Notch-Jagged and Notch-Delta-Jagged signaling

In order to understand the operating principles of Notch signaling, we first start evaluating the dynamics of circuit when only one ligand is available and later we evalate the combined circuit.

### 2.1.1 Notch-Delta (ND) circuit: a two-cell toggle switch

To evaluate the dynamics of Notch-Delta stand alone circuit, we analyzed the reduced model given by equations 1.10-1.11 when  $J_0 = 0$ , i.e., no Jagged is produced. This simplified circuit presents bistability, and the two stable steady states are: (i) Sender (S) - in which the cell has

(high Delta (ligand), low Notch (receptor)), and (ii) Receiver (R) - in which the cell has (low Delta (ligand), high Notch (receptor)) (Figure 2.1). Therefore, Notch-Delta signaling forms an intercellular mutually inhibitory switch, which drives adjacent cells to adopt alternate fates - one cell acts as a Sender and the other as a Receiver, or vice-versa, as noted in previous studies [21, 22, 20, 45]. A canonical example of this phenomenon, also known as lateral inhibition, is the AC/VU differentiation in *C. elegans*, where initially the two cells are identical (i.e. both neighbors have the same levels of ligand and receptor), but then due to a random fluctuation, one cell has increased levels of ligand and/or the other has increased level of receptor; and this stochastic difference is amplified by the mutually inhibitory feedback loop - driving the cell fate diversification [59]. However, lateral inhibition need not be static, it can be dynamic too - i.e. cells can compete against their neighbors continuously, as seen during sprouting angiogenesis, where the migrating ‘stalk’ cells are continuously competing to become ‘tip’ cells [60] . Also, during neural development, lateral inhibition occurs between the dynamic clusters of neuronal precursors, in order to strike a balance between the number of cells differentiating to become neurons, and those staying as precursors for further rounds of neurogenesis [61].

In order to better understand the dynamics of the circuit, we evaluate a bifurcation diagram when the external Delta ( $D_{ext}$ ) acts as a control parameter - the range of existence of the different Notch/Delta states of a single cell as function of ( $D_{ext}$ ). We see that for small  $D_{ext}$ , the cell behaves as a Sender (S); and for large  $D_{ext}$ , the cell behaves as a Receiver (R). We further see the existence of bistability for intermediate levels of  $D_{ext}$ , - the cell can either be a Sender (S) or a Receiver (R) (Figure 2.2A). Next, we present the phase diagram (two-parameter bifurcation diagram) for a single cell driven by two control parameters, the external Notch ( $N_{ext}$ ) and the external Delta ( $D_{ext}$ ) shown in Figure 2.2B. Doing so reveals the existence of three distinct phases - (i) monostable Sender {S} phase, (ii) monostable Receiver {R} phase, and (iii) a bistable phase {S,R} where cells can either be Receiver or Sender. The results indicate that the Notch-Delta circuit behaves as an inter-cellular mutually inhibitory bistable two-way toggle-like switch. As such, this switch drives two Notch-Delta interacting neighboring cells to adopt opposite fates - one cell as a Sender and the other as a Receiver or vice-versa. This result is consistent with the mutual inhibition mechanism commonly associated with Delta-mediated Notch signaling [62], also

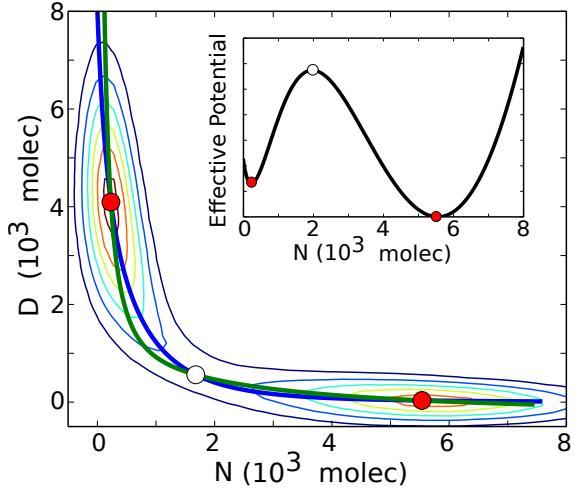


Figure 2.1: Dynamical properties of Notch-Delta signaling circuit.  $N$ - $D$  stand-alone circuit ( $J_0 = 0$ ,  $D_0 = 1000$  molecules/hour,  $N_0 = 500$  molecules/hour,  $N_{ext} = 500$  molecules,  $D_{ext} = 1100$  molecules) presents two stable states: Sender (S) (high Delta, low Notch) and Receiver (R) (low Delta, high Notch). Green nullcline represents  $dN/dt = 0$ , and blue nullcline represents  $dD/dt = 0$ . Stable steady states are represented by red circles and unfilled circles represent unstable steady states. The background contours illustrate the values of the effective potential ( $-\log(P)$ ) where  $P$  is the probability density calculated by solving the differential equations stochastically using the Euler-Maruyama method. Most probable regions are at the red regions and least probable at the blue regions. Insets represent the one-dimensional effective potential computed along the  $dD/dt = 0$  nullcline.

referred to as lateral inhibition. For this reason, the Notch-Delta signaling is critical for generating checkerboard-like patterns as well as sharp boundaries of wing vein formation in the Drosophila wing disc [20], and also in the differentiation of sensory cells [63].

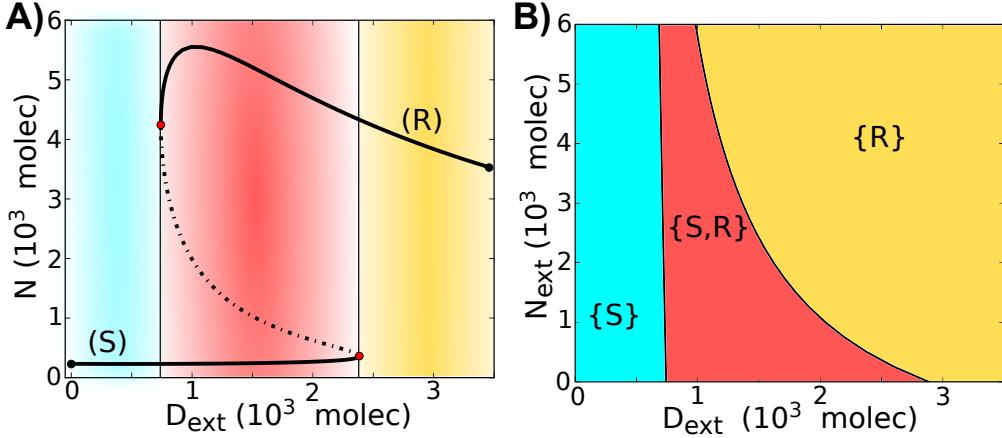


Figure 2.2: Dynamical system characteristics of the Notch-Delta circuit. A) Bifurcation, for the one-cell case, of Notch protein levels on the membrane as a function of the number of external Delta ( $D_{ext}$ ) for fixed  $N_{ext} = 500$  molecules. Starting in the Sender (S) state, i.e.- (low Notch, high Delta) (blue region) and increasing the external Delta ( $D_{ext}$ ) at some threshold the cell undergoes a transition to the Receiver state, i.e. (high Notch, low Delta) (yellow region). The reverse transition occurs at a different number of  $D_{ext}$  proteins that leads to a region of coexistence of both states - Sender and Receiver (red region). Solid curves represent stable steady states, while dotted curves represent unstable steady states. B) Phenotype diagram as a function of external Notch ( $N_{ext}$ ) and external Delta ( $D_{ext}$ ) for one cell model. The monostable phase  $\{S\}$  corresponds to the Sender state (low Notch, high Delta) and monostable phase  $\{R\}$  corresponds to the Receiver state (high Notch, low Delta). The bistable phase  $\{S,R\}$  corresponds to a region of co-existence of both states - Sender and Receiver.

### 2.1.2 Notch-Jagged (NJ) circuit

We next consider the dynamics of Notch-Jagged signaling by analyzing the reduced model given by equations 1.10-1.11 when  $D_0 = 0$ . Notch-Jagged circuit is monostable with the only steady state being when cells have (medium Jagged (ligand), medium Notch (receptor)) (Figure 2.3), thus cells can both send and receive signals, thereby leading to cell fate convergence. This phenomenon, also known as lateral induction, can also play a crucial role in generating spatial patterns with wavelength of a dozen of cells [64], and can thereby facilitate the spatial patterning of different cell types required during pancreatic development [65]. However, since the previous computational models for lateral induction do not specifically consider Jagged, they can offer limited insights into cases where Jagged plays a key role, such as in tumor-stroma crosstalk [33, 27].

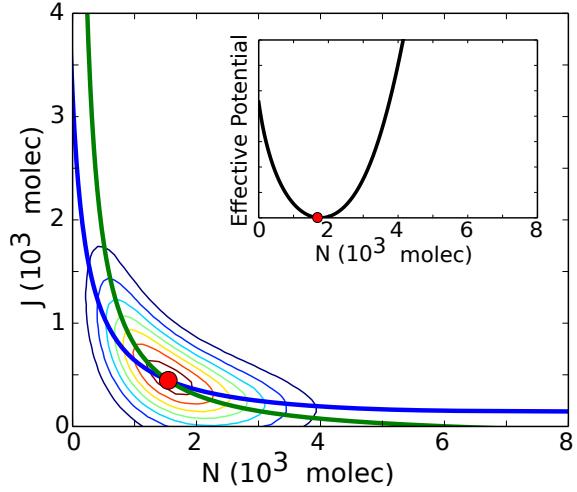


Figure 2.3: Dynamical properties of Notch-Delta signaling circuit. N-D stand-alone circuit ( $J_0 = 400$  molecules/hour,  $D_0 = 0$ ,  $N_0 = 500$  molecules/hour,  $N_{ext} = 500$  molecules,  $L_{ext} = 250$  molecules) presents two stable states: Sender (S) (high Delta, low Notch) and Receiver (R) (low Delta, high Notch). Green nullcline represents  $dN/dt = 0$ , and blue nullcline represents  $dJ/dt = 0$ . Stable steady states are represented by red circles and unfilled circles represent unstable steady states. The background contours illustrate the values of the effective potential ( $-\log(P)$ ) where  $P$  is the probability density calculated by solving the differential equations stochastically using the Euler-Maruyama method. Most probable regions are at the red regions and least probable at the blue regions. Insets represent the one-dimensional effective potential computed along the  $dJ/dt = 0$  nullcline.

### 2.1.3 Notch-Delta-Jagged (NDJ) circuit: a three-way switch.

We next study the dynamics of the combined Notch-Delta-Jagged signaling circuit, i.e., Notch signaling when driven by both its ligands - Delta and Jagged. This system can behave as a three-way switch, allowing for three states: - (i) Sender (S - high ligand, low receptor), (ii) Receiver (R - low ligand, high receptor) and (iii) hybrid Sender/Receiver (S/R - medium ligand, medium receptor) (Figure 2.4).

We further present the phase diagram (two-parameter bifurcation diagram) for a single cell driven by two control parameters, the external Notch ( $N_{ext}$ ) and the external Jagged ( $J_{ext}$ ). Note that, once Notch has the same binding affinity to both Delta and Jagged, the model is symmetric by changing  $J_{ext}$  for  $D_{ext}$ , therefore, the following results also holds when  $D_{ext}$  act as a control parameter. The phase diagram comprises three monostable phases -  $\{S\}$ ,  $\{R\}$  and  $\{S/R\}$ , three phases of coexistence of two phenotypes  $\{S,R\}$ ,  $\{S,S/R\}$  and

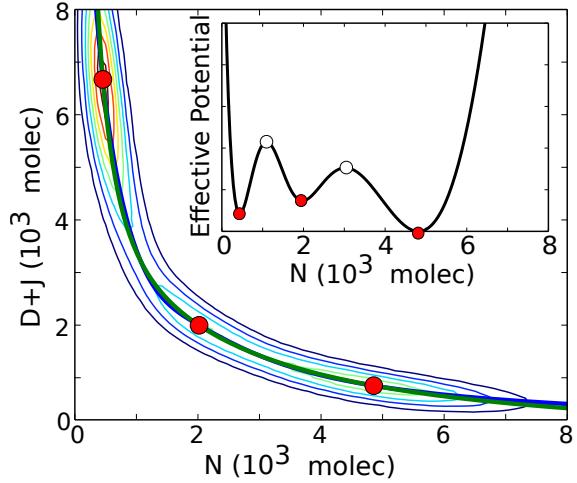


Figure 2.4: Dynamical properties of Notch-Delta signaling circuit. N-D stand-alone circuit ( $J_0 = 1000$  molecules/hour,  $D_0 = 1400$  molecules/hour,  $N_0 = 1500$  molecules/hour,  $N_{ext} = 400$  molecules,  $L_{ext} = 1100$  molecules) presents two stable states: Sender (S) (high Delta, low Notch) and Receiver (R) (low Delta, high Notch). Green nullcline represents  $dN/dt = 0$ , and blue nullcline represents  $dL/dt = 0$ , where  $L = D + J$ . Stable steady states are represented by red circles and unfilled circles represent unstable steady states. The background contours illustrate the values of the effective potential ( $-\log(P)$ ) where  $P$  is the probability density calculated by solving the differential equations stochastically using the Euler-Maruyama method. Most probable regions are at the red regions and least probable at the blue regions. Insets represent the one-dimensional effective potential computed along the  $dL/dt = 0$  nullcline.

$\{S/R, R\}$ , and also a tristable phase showing the coexistence of all three possible states  $\{S, S/R, R\}$  (Figure 2.5A). At larger  $N_{ext}$  values, we see the hybrid S/R state can exist, for some range of  $J_{ext}$ , by itself, i.e. in the monostable  $\{S/R\}$  phase (Figure 2.5B). However, at smaller  $N_{ext}$  values, the hybrid state always co-exists with other states in a bistable phase  $\{S, S/R\}$  and  $\{S/R, R\}$ , or in a tristable phase  $\{S, S/R, R\}$  (Figure 2.5C).

The S/R hybrid state has intermediate levels of both the receptor and the ligands, therefore allowing for bidirectional signaling. In the case of two interacting cells in this hybrid state, indicated by  $(S/R; S/R)$ , the two cells have similar intermediate levels of the ligands in contrast to the additional two opposite states  $((S; R) \text{ and } (R; S))$  (Figure 2.6). In other words, while the Notch-Delta signaling enables only opposite fates,  $(S; R) \text{ and } (R; S)$ , the Notch-Delta-Jagged signaling enables two similar interacting cells to have similar fates of being in a hybrid state  $(S/R; S/R)$ . Such similar fate adoption - also known as lateral induction - is a signature of Notch-Jagged signaling. For example, during inner ear development,

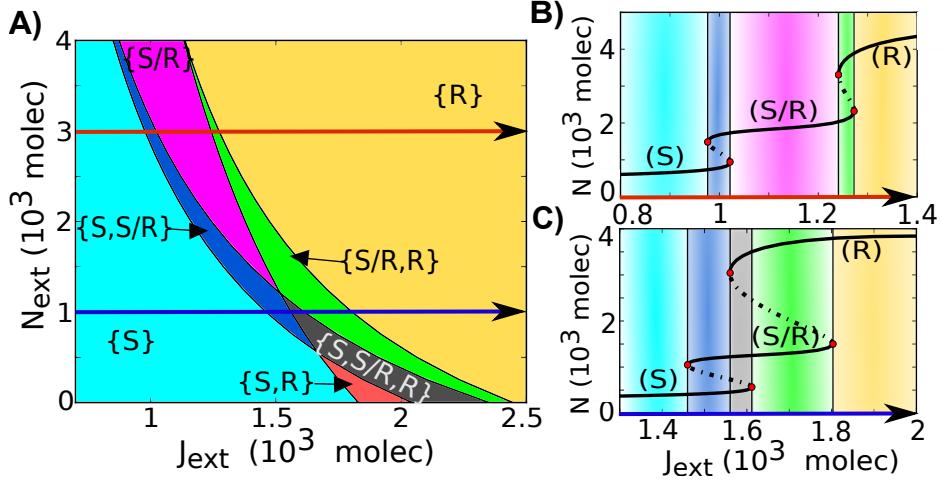


Figure 2.5: Dynamical system characteristics of the Notch-Delta-Jagged circuit. A) Phenotype-diagram when the one-cell Notch-Delta-Jagged circuit is driven by both the external Notch ( $N_{ext}$ ) and external Jagged ( $J_{ext}$ ), for ( $D_{ext} = 0$ ). Each phase, denoted by a different color, corresponds to a different combination of coexisting phases. Same phenotype-diagram is obtained when driven by  $N_{ext}$  and  $D_{ext}$ , for  $J_{ext} = 0$ , once Notch is considered to have the same binding affinity to  $J_{ext}$  and  $D_{ext}$ . B) Bifurcation of Notch protein levels on the membrane when driven by external Jagged for fixed levels of  $N_{ext} = 3000$  proteins. This curve shows the existence of the monostable  $\{S/R\}$  phase (pink region) for a large range of external ligands. C) Same as B) for  $N_{ext} = 1000$  proteins. In this case, the hybrid S/R state co-exists with other states, i.e. seen only in bistable (blue and green regions) and tristable phases (grey region).

lateral induction through Jagged1 in specific regions of the developing otocyst (auditory vesicle) enables the propagation and maintenance of prosensory character in some cells [63]. Also, during cardiac development, lateral induction specifies the cells that undergo epithelial to mesenchymal transition (EMT) to form endocardial cushion and heart valves [11]. Besides, lateral induction has been implicated in vertebrate somite boundary formation [66] as well as in wing margin development [10, 20].

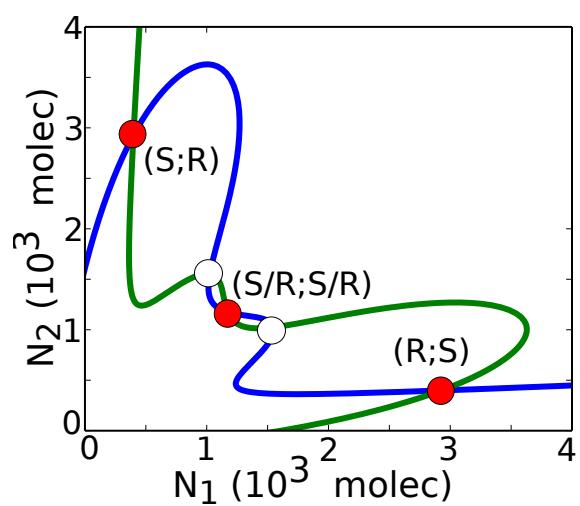


Figure 2.6: Nullclines for the case of two cells interacting with each other through Notch-Delta-Jagged. The blue nullcline is for condition of all ODEs being set to zero except for  $dN_1/dt$  and the green nullcline is for condition of all ODEs being set to zero except for  $dN_2/dt$

## 2.2 The effect of Fringe mediated asymmetric Notch-ligand binding

Glycosylation of Notch by Fringe creates additional asymmetry between Delta and Jagged by modulating the binding affinity of the two ligands to Notch - the glycosylated Notch has a higher affinity to bind to Delta, but lower affinity to bind to Jagged [18, 19]. To incorporate this mechanism within our framework, we considered two distinct sub-populations of Notch - the one modified by Fringe, and the other unmodified. Since NICD, which is represented in the model by  $(I)$ , activates Fringe [15], we have taken the fraction of glycosylated Notch (denoting the effect of Fringe on Notch) to increase with  $(I)$ . This glycosylated Notch has different strengths of *cis*-inhibition and *trans*-activation for Delta and for Jagged [16]. Thus, while representing effective Notch (sum of glycosylated and unglycosylated Notch), we consider the strengths of *cis*-inhibition and *trans*-activation of Notch for Delta and for Jagged to depend on  $(I)$  (see Eq. 1.12-1.15).

Within our framework, different modulations of Notch by Fringe can be represented by different values of the parameters  $\lambda_D^F$  and  $\lambda_J^F$ , which represents either the increase ( $\lambda^F > 1$ ) or decrease ( $\lambda^F < 1$ ) of the *cis*-inhibition and *trans*-activation rates. When  $\lambda_D^F = \lambda_J^F = 1$ , the model is the same as considered earlier without any effect of Fringe (Eqs. 1.6-1.9). In this case, the Notch-ligand binding has equal affinity for external Jagged and external Delta, as reflected in the symmetry of the phenotype diagram (two-parameter phase diagram) for external Jagged and external Delta (Figure 2.7A (center)). The bifurcation diagram for a cell driven by external Delta ( $D_{ext}$ ) and by external Jagged ( $J_{ext}$ ) presents the same behavior - a large range of the intermediate state (S/R) in a monostable phase (Figure 2.7A). However, when the effect of Fringe is incorporated (e.g.  $\lambda_D^F = 3$  and  $\lambda_J^F = 0.3$ ), the circuit behaves differently. The range of the existence of the 4 phases containing the hybrid S/R state, (the phases {S/R}, {S/R, R}, {S, S/R} and {S, S/R, R}), increase with the level of external Jagged (Figure 2.7B (center)). When Notch signaling is mainly mediated by Jagged (high  $J_{ext}$ ), both the forward and backward transitions between (S) and (R) states require a transitions into and from the hybrid S/R state as an intermediary step (Figure 2.7B (right)). Conversely, when the Notch signaling is mainly mediated by Delta (high  $D_{ext}$ ), the forward and backward transitions between (S) and (R) do not undergo through the intermediate state

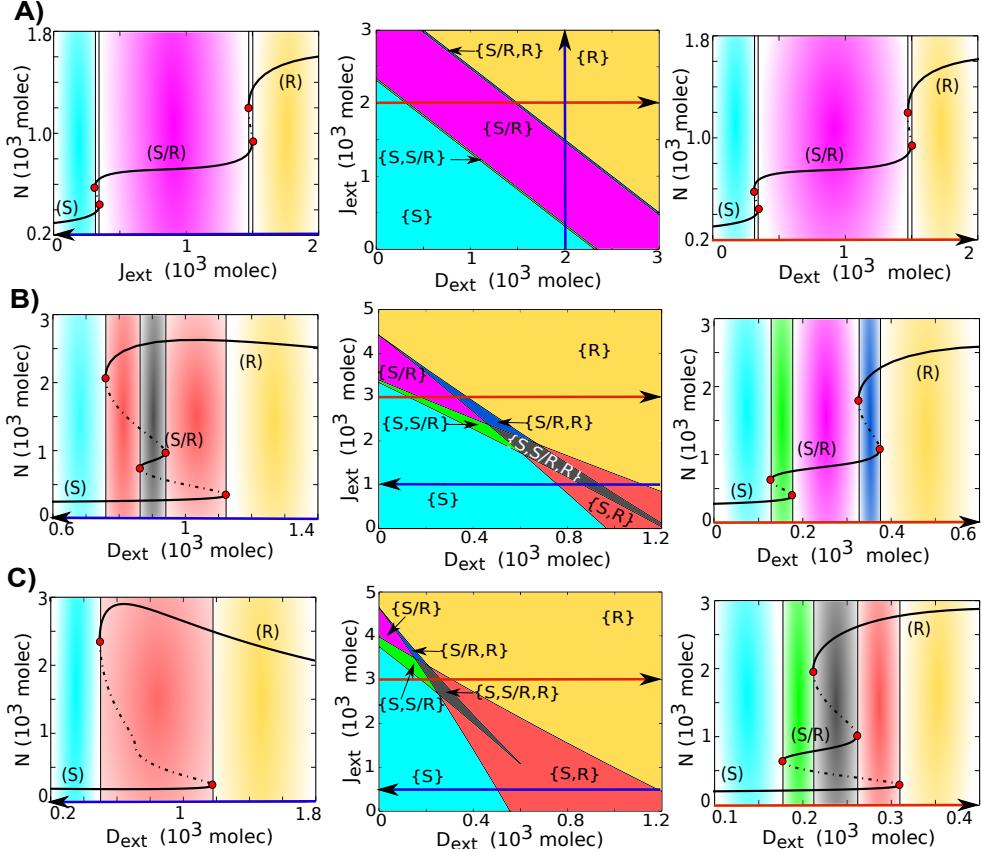


Figure 2.7: Phenotype-diagram and bifurcation curves for the one-cell Notch-Delta-Jagged-Fringe circuit. The phenotype-diagram shows the different possible phases when the circuit is driven by variable levels of both external Jagged and external Delta. A) Phenotype-diagram (center) for  $\lambda_D^F = \lambda_J^F = 1$  (no Fringe effect). In this case, the circuit response to external Jagged and external Delta is symmetric. Bifurcation curve of Notch protein levels with respect to varying external Jagged values (left) for fixed  $D_{ext} = 2000$  and  $N_{ext} = 500$  molecules and (right) bifurcation curve with respect to varying external Delta values for fixed  $J_{ext} = 2000$  and  $N_{ext} = 500$  molecules. B) Phenotype-diagram (center) for  $\lambda_D^F = 3$  and  $\lambda_J^F = 0.3$  (intermediate effect of Fringe). Bifurcation curves of Notch protein levels in response to varying  $D_{ext}$  for  $J_{ext} = 1000$  and  $N_{ext} = 500$  molecules (left), i.e. Notch signaling mainly mediated through Delta and for fixed  $J_{ext} = 3000$  and  $N_{ext} = 500$  molecules (right), i.e. Notch signaling mainly mediated through Jagged. C) Phenotype-diagram (center) for  $\lambda_D^F = 5.0$  and  $\lambda_J^F = 0.2$  (very strong effect of Fringe). Bifurcation curves of Notch protein levels in response to  $D_{ext}$  for fixed  $J_{ext} = 500$  and  $N_{ext} = 500$  molecules (left), in which the hybrid S/R state no longer exists, and the circuit behaves like a bistable toggle switch similar to the circuit considering Notch-Delta only, and for fixed  $J_{ext} = 3000$  and  $N_{ext} = 500$  molecules (right), in which the hybrid S/R state can be observed to co-exist with other states (green and grey regions).

(Figure 2.7B (left)). When the effect of Fringe is considered to be too strong ( $\lambda_D^F = 5$  and  $\lambda_J^F = 0.2$ ), the circuit is mostly bistable and therefore, the response of the circuit becomes similar to the case of stand alone Notch-Delta signaling (Figure 2.7C).

The results above suggest that signaling through Jagged has an important role in maintaining the hybrid Sender/Receiver (S/R) state, and that Jagged makes it much more likely that transition from Sender (S) to Receiver (R) and vice-versa happen through the hybrid (S/R) state.

### 2.2.1 The effect of Delta/Jagged asymmetry on the cell-cell fate modulation

Notch signaling in mammals is mediated through four types of Notch (Notch 1-4) and three types of Fringe (Lunatic, Manic and Radical Fringe) [67]. Experimental evidence suggests that most Fringe proteins act with different types of Notch, possibly leading to different forms of glycosylated Notch; thereby expanding the repertoire of responses that the Notch signaling system can mediate [18, 19]. Most experimental evidence suggests that Fringe increases the signaling mediated by Delta and decreases the signaling mediated by Jagged, resulting in  $\lambda_D^F > 1$  and  $\lambda_J^F < 1$  [18, 19]. The phenotype diagram when the circuit is driven by different values of  $\lambda_D^F$  and  $\lambda_J^F$  presents the response of the circuit for different combinations of Fringe modulations (Figure 2.8).

Because Fringe is activated by NICD [15], its effect is dominant in cells with high number of Notch molecules (Receiver (R) state) that cleave to form NICD. Therefore, to analyze the effect of Fringe on Notch-Delta-Jagged signaling, we choose the external signal to the cell be composed mainly by ligands ( $J_{ext}$ ,  $D_{ext}$ ) and low values of  $N_{ext}$ ; i.e. the external signal can be considered equivalent to a Sender (S) cell. Two such different combinations are chosen - (high  $D_{ext}$ , low  $J_{ext}$ ) and (low  $D_{ext}$ , high  $J_{ext}$ ) (Figure 2.8). In case of (high  $D_{ext}$ , low  $J_{ext}$ ) and at  $\lambda_D^F > 2$  and  $\lambda_J^F < 1$ , i.e. when the external signal is mostly Delta and Fringe increases the affinity of Notch for Delta, and decreases that for Jagged, the cell is mostly in monostable phase of the Receiver (R) state, or, in other words, the cell attains the opposite fate as that of a cell representing the external signals (Figure 2.8A). However, when the external signal is mainly Jagged, i.e. in (low  $D_{ext}$ , high  $J_{ext}$ ), at smaller values of  $\lambda_D^F$  ( $\lambda_D^F < 4$ ) and  $\lambda_J^F$  ( $\lambda_J^F < 1$ ), i.e. the effect of Fringe is not very pronounced, the cell is

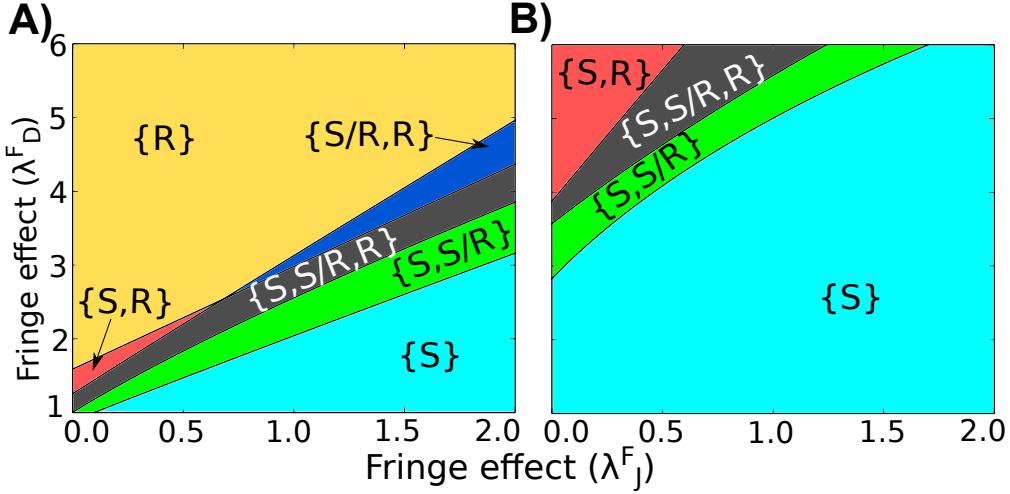


Figure 2.8: Phenotype-diagram of the Notch-Delta-Jagged-Fringe model when the circuit is driven by different values of the Fringe modulation for Notch-Delta interaction  $\lambda_D^F$  and Notch-Jagged interaction  $\lambda_J^F$ . In all curves the external signal represents cells in the Sender (S) state - low concentration of Notch ( $N_{ext} = 500$ ) and high concentration of ligands. Each figure represents a different combinations of the number of external ligands (in number of proteins available to binding). A)  $D_{ext} = 1500$  and  $J_{ext} = 500$  molecules. B)  $D_{ext} = 500$  and  $J_{ext} = 1800$  molecules.

mostly in the monostable phase of the Sender (S) state (Figure 2.8B). Therefore, the cell attains a fate similar to the cell represented by the external signal. These results suggest that while signaling through the Notch-Delta circuit, the two cells attain opposite fates, however, when signaling through the Notch-Jagged circuit, the two cells attain similar fates; thereby suggesting that Jagged helps neighboring cells to maintain similar fates.

We also investigate the Fringe effect when it is driven exclusively by different levels of  $D_{ext}$  and  $J_{ext}$ . We plot a bifurcation diagram for the Fringe effect ( $f = \lambda_D^F = 1/\lambda_J^F$ ) and the amount of external ligand. When the cell is driven mainly by external Delta ( $D_{ext}$ ), in presence of little or no external Jagged ( $J_{ext}$ ), increase in Fringe effect decreases the range of existence of the phases that contain the hybrid S/R state -  $\{S/R\}$ ,  $\{S, S/R\}$ ,  $\{S/R, R\}$  and  $\{S, S/R, R\}$  (Figure 2.9A, C). Similar effect of Fringe levels is observed when the cell is driven by both external Delta and external Jagged (Figure 2.9D). However, when the cell is driven solely by Jagged ( $D_{ext} = 0$ ), an increase in Fringe effect does not significantly decrease the range of existence of the  $\{S/R\}$  phase, but at high levels of Fringe, the levels of external Jagged required to maintain the S/R state are almost doubled as compared to low levels of

Fringe (Figure 2.9B).

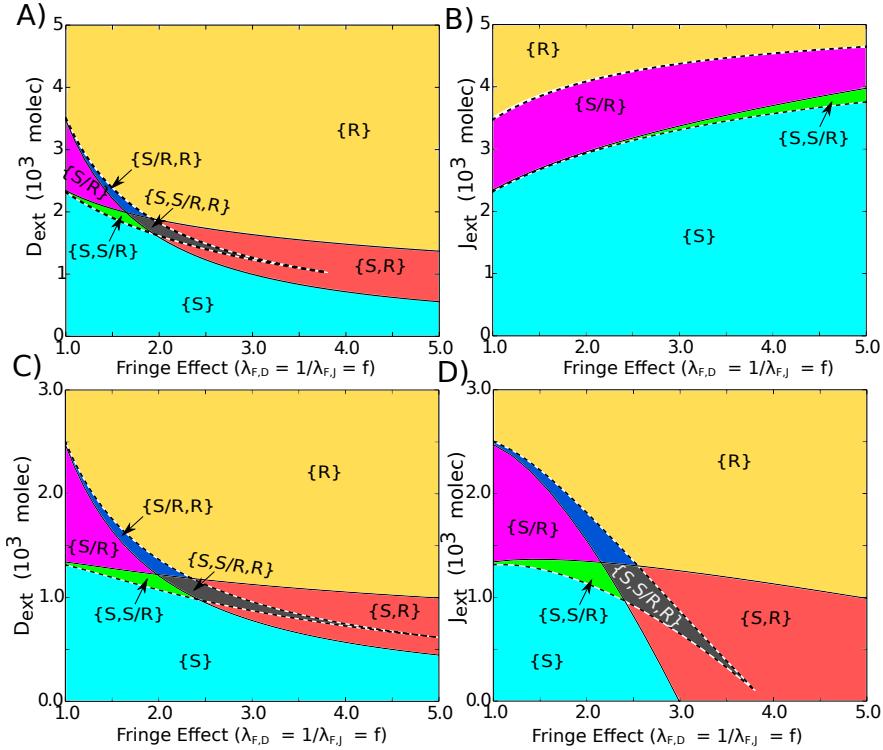


Figure 2.9: Phenotype diagram when the circuit for one-cell is driven by both the effect of Fringe and external ligands. Each phase is represented by a different color and corresponds to a different combination of coexisting states. Each diagram represents different cases for external molecules available to the cell. A) External signal is mediated by Delta ( $J_{ext} = 0$ ,  $N_{ext} = 500$  molecules). B) External signal is mediated by Jagged ( $D_{ext} = 0$ ,  $N_{ext} = 500$  molecules). C) same as A) but for ( $J_{ext} = 1000$  molecules,  $N_{ext} = 500$  molecules). D) same as B) but for ( $D_{ext} = 1000$  molecules,  $N_{ext} = 500$  molecules).

## 2.2.2 Fringe promotes lateral inhibition patterning

As mentioned earlier, Notch-Delta signaling leads to lateral inhibition, where neighboring cells adopt alternate fates; but Notch-Jagged signaling leads to lateral induction, where adjacent cells attain similar fates [10, 12]. As Fringe can promote Notch-Delta signaling, we hereby explore its role on tissue patterning.

To address this issue, we simulate a one-dimensional layer of cells interacting via NDJ pathway, for different values of Fringe effect. Our results show that at high Fringe effect, the

tissue pattern of similar fates is disrupted and the ‘salt-and-pepper’ pattern starts to emerge (Figure 2.10). Our results are consistent with experiments suggesting that Fringe promotes lateral inhibition during neurogenesis [68].

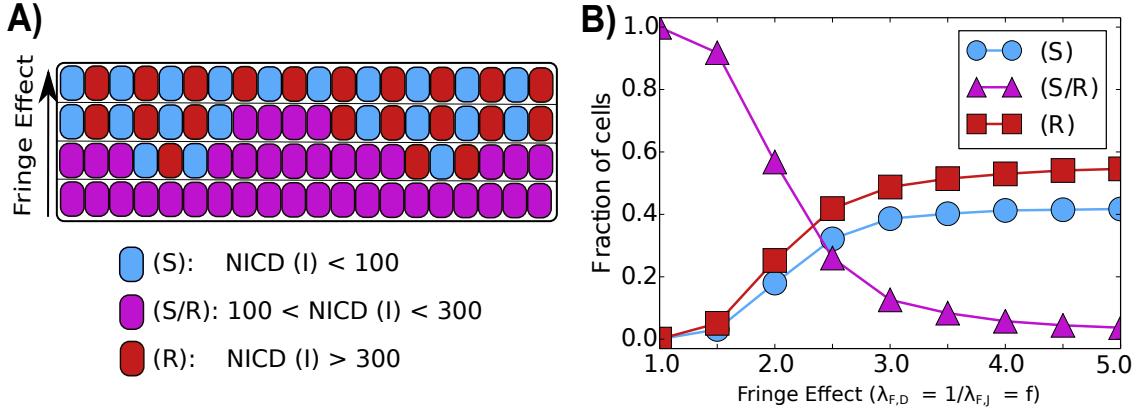


Figure 2.10: Fringe effect on tissue-level patterning. A) Representation of a one-dimensional layer of cells interacting through Notch signaling for different levels of Fringe. B) The average of the fraction of cells in (S), (S/R) or (R) state as a function of the effect of Fringe. Increase in Fringe effect promotes lateral inhibition - formation of ‘salt and pepper’ alternate patterns of cell fate in adjacent cells. The averages were taken over 100 simulations of a one-dimensional layer of 100 interacting cells with periodic boundary condition.

### 2.2.3 Fringe alters the relative stability of the steady states of Notch-Delta-Jagged signaling

In order to evaluate the effect of noise in the N-D-J circuit, we represent the phase space by an effective potential for different values of Fringe effect. We considered two interacting cells and the phase space is presented in terms of the levels of Notch of each cells. The effective potential of the system is defined as  $-\log(P)$ , where  $P = P(N_1, N_2)$  is the probability density at the 2 dimensional phase space ( $N_1 \times N_2$ ). Three states can be observed: first cell as a Sender and the second one as a Receiver (S,R) (cell 1 has low Notch, cell 2 has high Notch), first cell as a Receiver and the second one as a Sender (R,S) (cell 1 has high Notch, cell 2 has low Notch) and both cells as both Sender and Receiver (S/R,S/R) (both with intermediate level of Notch). When Fringe has low or no effect, both cells tend to stay predominately in the intermediate (S/R,S/R) state and alternate the fates only sporadically (Figure 2.11A).

Nevertheless, when the effect of Fringe is considered, the basin of attraction of the (S/R,S/R) becomes shallower and small fluctuations can move the cells to the alternate state (one cell has high Notch and the other low Notch) which now have a deep basin of attraction (Figure 2.11B).

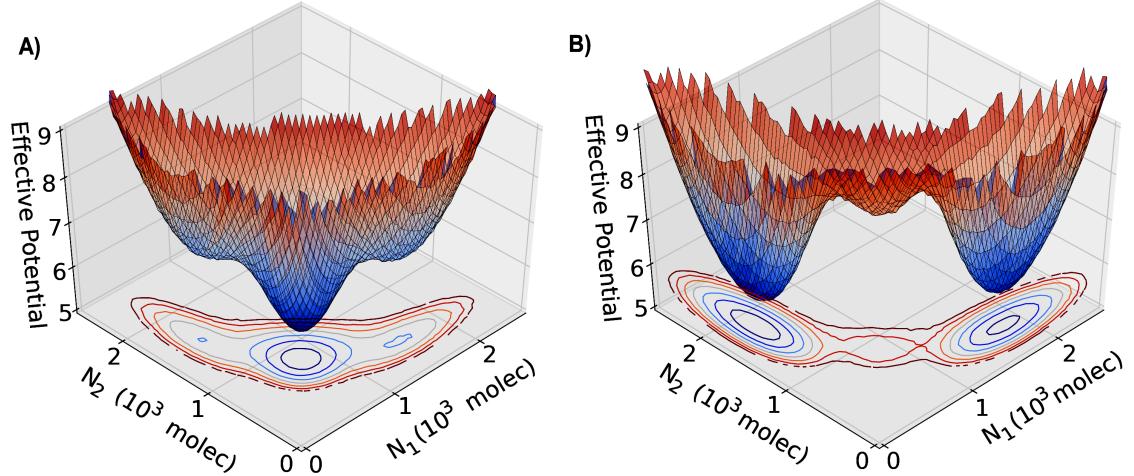


Figure 2.11: Three-dimensional representation of the effective potential as a function of Notch in cell 1 ( $N_1$ ) and in cell 2 ( $N_2$ ). The effective potential is defined as  $-\log(P)$ , where  $P = P(N_1, N_2)$  is the probability density calculated by solving the differential equations stochastically using the Euler-Maruyama method. A) represents the case of no effect of Fringe ( $f = 1.0$ ). B) represents the case considering the effect of Fringe ( $f = 1.5$ ).

These results indicate that Fringe promotes Notch-Delta signaling at the expense of Notch-Jagged signaling. Because Notch-Jagged signaling is more frequently implicated than Notch-Delta signaling in many aspects of tumor progression [15], our results showing the role of Fringe to inhibit Notch-Jagged signaling corroborate with Fringe being reported as a tumor suppressor in multiple cancers such as prostate, lung and basal-like breast cancer (BLBC) [69, 70, 71]. Further, loss of Fringe facilitates Notch-Jagged signaling that is critical in the tumorigenesis of BLBCs [69].

In mammals, Fringe can have three homologues- Lunatic Fringe (Lfng), Manic Fringe (Mfng) and Radical Fringe (Rfng). It may be noted that the results about Fringe promoting Notch-Delta signaling and lateral inhibition are true for those homologues that have similar circuit architecture with Notch signaling as considered here, i.e. they promote Notch-Delta signaling both in cis and in trans, and they are activated by NICD. Among the existing

homologues, only Lfng has been shown to be a direct downstream target of NICD [15]; and both Lfng and Mfng, but not Rfng, promotes Notch-Delta signaling [16]. Consistently, both Lfng and Mfng, but not Rfng, has been reported to be a tumor suppressor [69, 70, 71]. Future models should consider the effect of these different Fringe homologues distinctly instead of collapsing them together.

## 2.3 Production rates of the two ligands affect cell-fate decision antagonistically

The two ligands - Delta and Jagged - can be produced at different rates in the same cell due to stochastic events and/or the influence of external factors such as the inflammatory cytokine TNF- $\alpha$  that can promote the production of Jagged and inhibit that of Delta [36]. Therefore, we next explore the effect of varying the production rates of the ligands on cell-fate decisions mediated by Notch-Delta-Jagged (NDJ) signaling, for both one-cell and two-cell system. For the one-cell system, our results are represented as phenotype diagrams when the circuit is driven by two control parameters - external Delta ( $D_{ext}$ ) and external Jagged ( $J_{ext}$ ) (Figure 2.12). These diagrams consist of multiple phases (sets of co-existing phenotypes for the same parameter set): three monostable phases - {S}, {S/R} and {R}, three bistable phases - {S, R}, {S, S/R} and {S/R, R}, and a tristable phase - {S, S/R, R}.

Different phenotype diagrams denote different combinations of the production rates of Jagged and Delta. At low production rates of Jagged ( $J_0$ ), increase in production rates of Delta ( $D_0$ ) decreases the parameter range for the existence of phases containing the hybrid S/R phenotype (compare the area bounded by dashed lines in Figure 2.12C vs. that in Figure 2.12A). Similar effects of high production rates of Delta ( $D_0$ ) are observed at high production rates of Jagged ( $J_0$ ). Therefore, high production and hence the levels of Delta inhibit Notch-Jagged signaling and promote lateral inhibition or cell fate diversification. Conversely, high production rates of Jagged can significantly increase the parameter range for the existence of phases containing the hybrid S/R phenotype (compare the area bounded by dashed lines in Figure 2.12D vs. that in Figure 2.12B). Hence, the two ligands act antagonistically with respect to their effect on cell-fate decisions in a cell population.

Next, we investigate the effect of different production rates of Delta and Jagged ( $D_0, J_0$ ) in a two-cell system. Both these cells are identical and communicate with each other via the NDJ system. At low production rates of both ligands - Jagged and Delta - both cells act as Receiver ( $R_1, R_2$  -low ligand, high receptor) because Notch (receptor) is still being produced at a fixed rate. However, when Delta is produced faster than Jagged, one cell becomes the Sender (S) and the other Receiver (R) ( $S_1, R_2$ ). Conversely, when Jagged is being produced faster than Delta, both cells attain the hybrid S/R state that enables them to communicate

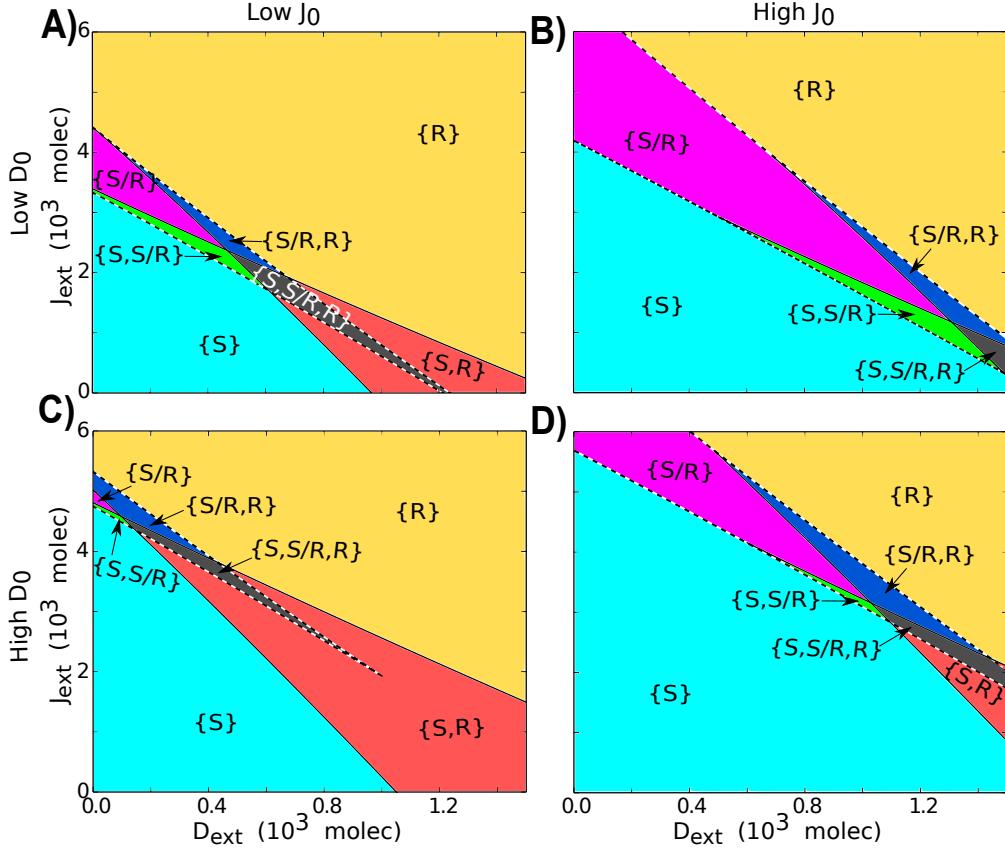


Figure 2.12: Phenotype diagram when the circuit is driven by both external Delta ( $D_{ext}$ ) and external Jagged ( $J_{ext}$ ). Each phase is represented by a different color and corresponds to a different combination of co-existing states. Each diagram represents the case for different production rates of Jagged ( $J_0$ ) and Delta ( $D_0$ )- units in molecules/ hour. A) Low production rate of both Jagged and Delta ( $J_0 = 1200$ ,  $D_0 = 1600$ ). B) High production rate of Jagged ( $J_0 = 1400$ ,  $D_0 = 1600$ ) C) High production rate of Delta ( $J_0 = 1200$ ,  $D_0 = 2000$ ). D) High production rates of both Jagged and Delta ( $J_0 = 1400$ ,  $D_0 = 2000$ ).  $N_{ext} = 500$  molecules.

in a bidirectional manner via the NDJ system ( $S/R_1$ ,  $S/R_2$ ) (Figure 2.12).

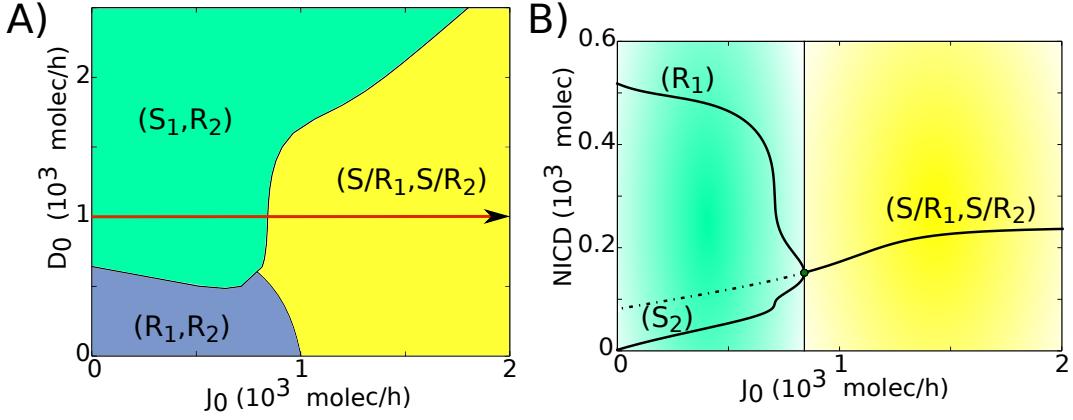


Figure 2.13: Phenotype diagram and bifurcation curve for two interacting cells. A) Phenotype diagram when the two-cell circuit is driven by both the production rates of Jagged ( $J_0$ ) and Delta ( $D_0$ ). Each phase is represented by a different color and corresponds to a different combination of co-existing states B) Bifurcation diagram of NICD levels with respect to the production rate of Jagged ( $J_0$ ) for  $D_0 = 1000$  molecules/hour (red arrow in A)

### 2.3.1 Ligand production rates control tissue level patterning

As mentioned earlier, Notch-Delta interactions lead to lateral inhibition - neighboring cells adopt alternate fates - Sender (S) and Receiver (R) [21, 20, 45, 72]. Notch-Jagged interactions lead to lateral induction [65, 64] - neighboring cells adopt similar fates - both of them can simultaneously send and receive the signal - Sender/Receiver (S/R) state. Although these mechanisms have been well-studied individually, the tissue level patterns that might emerge when both mechanisms act simultaneously have not been explored.

To address this issue, we simulate a one-dimensional layer of cells interacting via Notch pathway, for different values of  $D_0$  and  $J_0$  - Delta and Jagged production rates, respectively. Our results show that when the relative production rate of Delta is increased, the so-called 'salt-and-pepper' pattern (or alternate fate pattern) at a tissue-level begins to emerge (Figure 2.14A,C). On the other hand, when the relative production rate of Jagged is increased, the 'salt-and-pepper' pattern is disrupted, and the cells begin to adopt similar fates where they can both send and receive the signal (Figure 2.14B, D).

Our results are consistent with the patterns observed experimentally when both ligands are produced at different rates. For example, during hypoxia-mediated angiogenesis, increase in the concentration of vascular endothelial growth factor (VEGF) increases the production

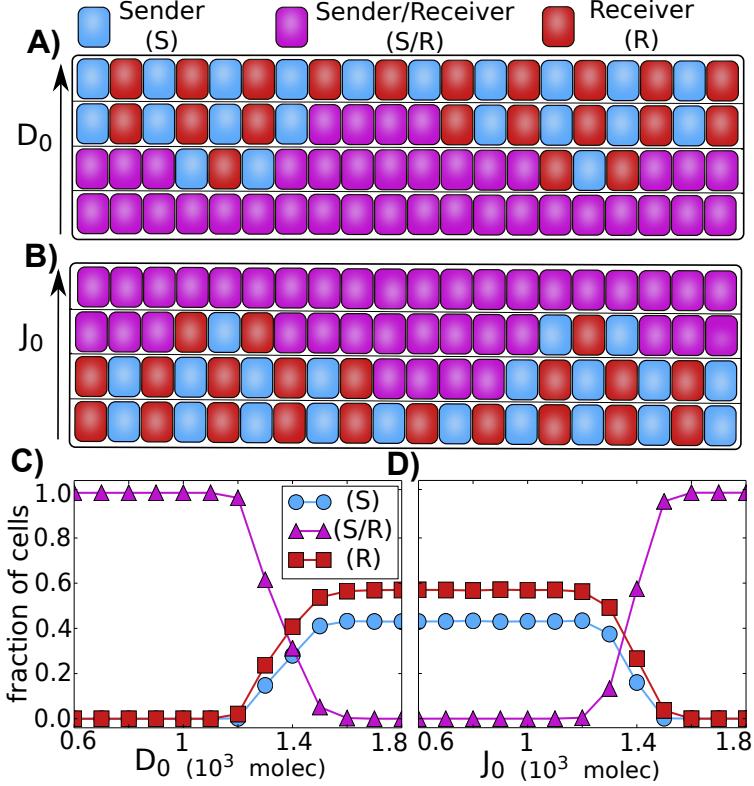


Figure 2.14: Patterning at the tissue level. A) and B) Representation an 1 dimensional layer of 20 interacting cells. A) The increase of the production of Delta ( $D_0$ ) leads to the formation of alternate patterns in which neighbors cells alternate between Sender and Receiver. B) The increase of the production of Jagged leads all the cells to the hybrid (S/R) state, therefore loosing the alternate (S) and (R) pattern. C) and D) Average of the fraction of cells in (S), (S/R) or (R) state as a function of ligand production. The averages were taken over 100 simulations of a 1 dimensional layer of 100 interacting cells with periodic boundary condition. C) Fraction of cells in (S), (S/R) or (R) state as a function of the production of Delta. D) Fraction of cells in (S), (S/R) or (R) state as a function of the production of Jagged. The state of the cells are defined according to the amount of signaling ( $I$ ): Sender ( $I < 100$  molecules), Sender/Receiver ( $100 < I < 300$ ) and Receiver ( $I > 300$ ).

of Delta, thereby causing some cells to adopt a tip fate - those with (high Delta, low Notch). Consequently, the remaining cells adopt a stalk fate - (low Delta, high Notch). It may be noted that in this physiological context, the cells do not necessarily adopt a canonical 'salt-and-pepper' pattern, rather two tip cells might be separated by a few stalk cells, the number of which is determined by Jagged [36]. As another example, the inflammatory factors such as TNF- $\alpha$  can lead to increased Jagged production and decreased Delta production [36],

therefore driving the cells to a similar fate - hybrid Sender/Receiver (S/R) - that can promote bidirectional communication between tumor and stroma, a context where inflammation often plays a key role.

## 2.4 Soluble ligands affect cell-fate decision in a paracrine manner

In addition to transmembrane ligands, Notch signaling can also be activated by soluble forms of the ligands [33, 27, 28].

Soluble Jagged1 for example has been specifically implicated in mediating long-distance communication between tumor cells and stromal cells. Jagged1 can be secreted by endothelial cells that can activate Notch signaling in cancer cells, inducing them to gain migratory and invasive characteristics by undergoing partial or complete Epithelial to Mesenchymal Transition (EMT) [73]. Jagged1 can also induce the expression of NF-kB [74], which can increase the population of CSCs and further increase the secretion of Jagged1 [33], suggesting a wave-like mechanism in the tumor microenvironment to ratchet up the production and maintenance of therapy-resistant CSCs. Future theoretical studies of these circuits hold promise for appreciating the key role of soluble Jagged1 in mediating two interlinked and clinically insuperable facets of cancer - metastasis (as a result of cells undergoing EMT) and tumor relapse (as a result of expanded CSC pool).

Not only soluble Jagged1, but also transmembrane Jagged1 mediates tumor progression in several ways, and has been proposed to be a therapeutic target [31]. Notch-Jagged signaling plays a crucial role in the metastasis of breast cancer cells to bone, where prostate cancer cells expressing Jagged1 communicate with Notch-expressing osteoclasts to “home” in the bone [9]. Also, overexpression of Jagged1 on cancer cells can trigger Notch activation in neighboring endothelial cells (which can possibly secrete more soluble Jagged1), promoting sprouting tumor angiogenesis [75] and thereby tumor growth. Consistent with their pro-tumor roles, high levels of Notch and Jagged1 in cancer cells often correlates with poor patient survival [31].

To evaluate the role of soluble ligand-mediated paracrine signaling in cell-fate decision making, we incorporated the soluble forms of ligands Delta and Jagged in our two-cell system

(two identical cells communicating via NDJ pathway), as described in equations 1.16-1.19 where  $D_{ext}$  and  $J_{ext}$  represents Delta and Jagged of the neighboring cell and  $D_{ext}^s$  and  $J_{ext}^s$  represent soluble Delta and Jagged respectively.

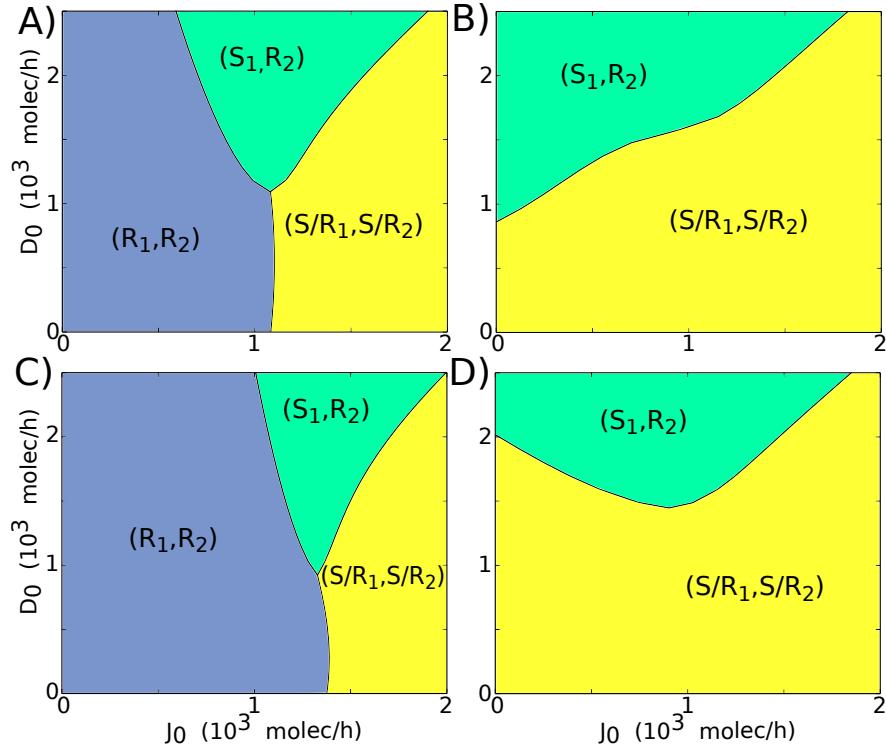


Figure 2.15: Phenotype diagram for two-cell system (two cells interacting via NDJ circuit) when the circuit is driven by the production rates of Jagged ( $J_0$ ) and Delta ( $D_0$ ). Each phase is represented by a different color and corresponds to a different combination of co-existing states. Each diagram represents different levels of soluble ligands Delta ( $D_{ext}^s$ ) and Jagged ( $J_{ext}^s$ ) interacting with these identical cells. A)  $D_{ext}^s = 600$  molecules,  $J_{ext}^s = 0$ . B)  $J_{ext}^s = 600$  molecules,  $D_{ext}^s = 0$ . C)  $D_{ext}^s = 1200$  molecules,  $J_{ext}^s = 0$ . D)  $J_{ext}^s = 1200$  molecules,  $D_{ext}^s = 0$ . The trans-activation rate for soluble ligands is considered to be 3 times lower than for membrane-bound ligands ( $k_T^s = k_T/3$ ).

For the case of paracrine communication via soluble Delta, both the cells tend to behave as Receiver ( $R_1, R_2$ ) for low production rate of Jagged ( $J_0$ ) or both of them behave as hybrid Sender/Receiver ( $S/R_1, S/R_2$ ) for high  $J_0$ . These cells can adopt opposite fates to each other only for intermediate levels of  $J_0$  and high production rate of Delta ( $D_0$ ) (Figure 2.15A, C, compare with Fig. 2.13A). Conversely, when the cells interact with soluble Jagged, the cells tend to attain similar fates - hybrid Sender/Receiver ( $S/R_1, S/R_2$ ) - unless the production

rate of Delta ( $D_0$ ) is high. For high  $D_0$ , these cells attain opposite fates- one behave as Sender, and the other Receiver ( $S_1, R_2$ ) (Figure 2.15B, D). Collectively, our results suggest that the soluble ligands - both Delta and Jagged - increase the range of parameters for which both cells attain same fate.

Our results also offer an explanation into why stromal cells secrete Jagged more often than Delta [33, 27]. Higher levels of soluble Jagged can enable many cells to adopt a hybrid S/R state (Figure 2.15B, D), thereby enabling both-way communication between cancer cells and/or between cancer and stromal cells. Not surprisingly, soluble Delta has been implicated in tumor suppression [30], but soluble Jagged can facilitate cancer cell survival as well as mediates communication between cancer cells and endothelial cells that leads to increased population of therapy-resistant tumor-initiating Cancer Stem Cells (CSCs) [27].

## 2.5 Discussion

### 2.5.1 The crucial role of hybrid S/R state in tumor progression

Being the key cell-cell communication pathway, Notch signaling circuit also couples to other aspects of cell-fate decision. Two of those modules play a crucial role in cancer metastasis - EMT (Epithelial to Mesenchymal Transition) module and stemness module (Figure 2.16A). EMT allows cancer cells to initiate metastasis and travel to different organs in the body, and gaining stemness allows them to initiate a new tumor at those different organs. Here, we discuss the crucial role of hybrid S/R state (that allows bidirectional communication between cancer cells and/or between cancer-stroma) in tumor progression.

#### Hybrid S/R state allows cancer cells to move and circulate collectively

EMT is an embryonic program that is aberrantly activated by cancer cells to invade adjacent tissues and gain access to the blood vessels to initiate metastasis. Cells undergoing EMT shed their epithelial traits (cell-cell adhesion, apico-basal polarity, no motility) and gain mesenchymal traits (no cell-cell adhesion, front-back polarity, and individual motility). However, when cells exit the circulation to start a new tumor at a distant organ, they undergo the reverse of EMT, i.e. MET (Mesenchymal to Epithelial Transition) - to shed their

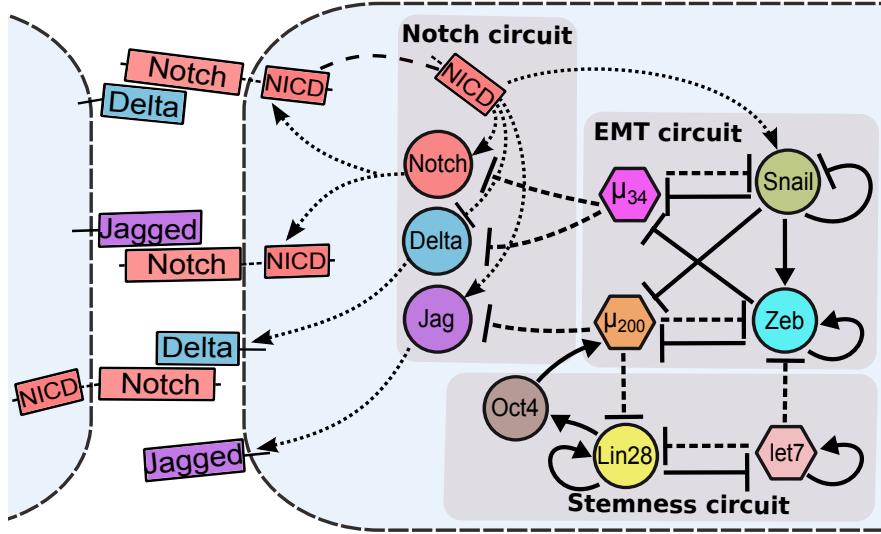


Figure 2.16: Connections between Notch signaling pathway, EMT decision circuit and stemness circuit; and the emergent implications n tumor progression. A) Notch signaling pathway couples to EMT circuit (miR-200/ZEB and miR-34/SNAIL) in multiple ways - miR-34 inhibits Delta and Notch, and miR-200 inhibits Jagged, and NICD transcriptionally activates Snail. EMT and stemness circuits (LIN28/let-7) also couple with each other - let-7 inhibits ZEB, miR-200 inhibits LIN28, and LIN28 activates miR-200 through OCT4. (see [47, 76, 77] and references therein). Dashed lines represent miR-mediated translational repression, and solid lines represent transcriptional regulation – bars for repression, and arrows for activation. B) Proposed interconnections between the hybrid E/M phenotype, Notch-Jagged signaling and maintenance and/or expansion of CSC population. Cell fate convergence achieved via Notch-Jagged signaling can stabilize the ‘metastable’ hybrid E/M phenotype that is associated with maximum stemness. Also, Notch-Jagged signaling can maintain and/or expand the CSC population in a non-cell autonomous manner (see [33]).

migratory traits and start proliferating to grow secondary tumors or colonize [78]. EMT is a two-step process in which epithelial cells first undergo partial EMT and then complete EMT to become mesenchymal. These two steps in decision-making are mediated by a three-way decision-making switch- a double negative feedback loop between microRNA miR-200 and transcription factor ZEB. (miR-200/ ZEB) allows the cells to attain three phenotypes - epithelial (E - high miR-200, low ZEB), mesenchymal (M - low miR-200, high ZEB) and hybrid epithelial/mesenchymal (E/M - medium miR-200, medium ZEB) [47, 77].

Cells in the hybrid E/M or partial EMT phenotype have mixed epithelial (cell-cell adhesion) and mesenchymal (migration) traits, thereby allowing them to migrate collectively, as observed during wound healing and branching mammary morphogenesis. Such collective

migration allows a sheet of cells to migrate by utilizing mechanical forces, and not every individual cell needs to detect and respond to the migration-inducing signal [79, 80, 81]. Notch signaling is instrumental during wound healing [82], but if these cells were to communicate mostly via Notch-Delta, they would have diverse fates from each other (one acts as Sender, and the other Receiver) that can possibly impair wound healing; thus, Notch-Jagged signaling might help maintain similar fates (hybrid S/R) and hence better co-operation among the cells during wound healing. Therefore, the hybrid S/R state facilitates cell-cell cooperation among collectively migrating cells and may play a crucial role in stabilizing the hybrid E/M phenotype that is otherwise considered to be ‘metastable’.

Recent studies show that cancer hijacks the advantages of collective cell migration from wound healing. During metastasis, the hybrid E/M phenotype is manifested as clusters of Circulating Tumor Cells (CTCs). These clusters can be more resistant to apoptosis in blood circulation, as well as up to 50-times more metastatic than individually migrating cells [83, 84], hence posing a much higher metastatic risk than individually migrating CTCs (a manifestation of the complete EMT phenotype). Further, maintaining the partial EMT phenotype might confer higher plasticity to cells to adapt to the changing microenvironment that can be useful for cells to revert to an epithelial phenotype during colonization, as well as continuing subsequent rounds of metastasis [85]. We hypothesize that Notch signaling pathway is critical in maintaining the clusters of CTCs together and therefore targeting it can have important possible therapeutic applications.

To test this hypothesis, future modeling attempts should incorporate the molecular links known between EMT and Notch signaling – miR-200 and miR-34, microRNAs that restrict EMT, can inhibit the translation of Delta, Jagged and Notch [86, 87, 88, 89]. Further, Notch signaling can induce EMT by activating the transcription factor SNAIL [90, 91] that inhibits miR-200 (Figure 2.16A) and is required for cell survival during partial EMT and facilitates ‘re-epithelialization’ of hybrid E/M cells [92, 93].

Activation of EMT in cancer cells also leads to an increase in their tumor-initiating capabilities or transform them to being like Cancer Stem Cells (CSCs) - the therapy-resistant cancer cells that act as a reservoir for new, more differentiated tumor cells, as well as maintain the ability to self-renew for many generations, thereby initiating a tumor [94, 95]. Recent studies show that cells undergoing partial EMT can gain ‘stemness’ that can be compromised

significantly if cells undergo complete EMT and become mesenchymal-like [76, 96, 97, 98]; thereby highlighting another advantage of collectively migrating CTCs during metastatic progression. All these functional advantages of collective cell migration or hybrid E/M phenotype - resistance to cell death during circulation, higher plasticity to both colonize and reinitiate metastasis, and enhanced tumor-initiating abilities - might be attained only in the presence of the stabilizing interactions - Notch-Delta-Jagged signaling (Figure 2.16B).

## Hybrid S/R state increases CSC population

In addition to links through the EMT module, Notch signaling and CSCs or cells with tumor-initiating abilities are interconnected in multiple ways. Notch and Jagged, but not Delta, are overexpressed in CSCs as compared to non-CSCs in multiple cancers- glioblastoma, pancreatic cancer as well as basal-like breast cancer (BLBC) [33, 99, 100, 25]. This overexpression of Notch and Jagged can potentially explain the role of Notch signaling in mediating therapy resistance [35, 101]. It may be noted that CSCs only form a small percentage of total cancer cells, but the way they divide can affect tumor progression drastically - (i) if a CSCs divides to form two non-CSCs, that can lead to depletion of CSC pool; (ii) if a CSC divides to give rise to one CSC and one non-CSC, that can maintain the CSC population; and (iii) if a CSC divides to give rise to both CSCs, that can expand the CSC population. A recent study suggests that Notch/miR-34 axis decides the mechanism of cell division of CSCs – high miR-34 levels lead to low Notch levels and differentiation of CSC to non-CSCs, low miR-34 levels lead to high Notch levels and division of one CSC to two CSCs, and consequently medium miR-34 levels lead to medium Notch levels and hence asymmetric cell division of CSCs [87]. Therefore, activation of Notch signaling, to medium or high levels, can maintain or expand the CSC pool respectively.

Another absolutely critical player that mediates the connection between Notch and CSCs is inflammation. The inflammatory cytokines such as TNF- $\alpha$  and IL-6 can both increase CSC population as well as increase the production rate of Jagged [36, 102]. Recent studies indicate that the increase in CSC population is mediated by high levels of Jagged. Further, Jagged is also secreted by many stromal cells [33, 27], thus the Notch-Jagged signaling might be more frequently activated in tumor microenvironment than the Notch-Delta one. Higher levels of soluble Jagged, but not soluble Delta, can enable a population of cells to adopt a hybrid

S/R state (Figure 2.15B, D) that facilitates bidirectional communication between cancer cells and/or between cancer and stroma, thereby offering a possible explanation of why stromal cells often secrete soluble Jagged, but not Delta. As noted earlier, higher levels of Notch and Jagged can also expand the population of therapy-resistant and tumor-initiating Cancer Stem Cells (CSCs) [33, 27, 103, 104].

In the next chapter, we present an extension of our framework that incorporates interplay between Notch and inflammation. Many models have explored the effect of CSCs on tumor progression in terms of population dynamics, but without considering any cell-cell communication [105, 106, 107, 108]. Since cell-cell communication is fundamental during collective decision-making of many populations, an understanding of tumor progression without considering cell-cell interaction via Notch-Delta-Jagged signaling is incomplete, and cannot be utilized to plan an effective and strategic ‘cyber war’ against them [109].

# Chapter 3

## Inflammation gives rise to mutually reinforcing Cancer Stem Cells through Notch-Jagged signaling

Inflammation is a hallmark of damaged tissue. Acute inflammation plays a crucial role in wound healing, and chronic inflammation predisposes the damaged tissue to the development of cancer - “wounds that do not heal” [110]. The inflammatory microenvironment has increased activation of transcription factors like NF- $\kappa$ B and STAT3 that promote the secretion of cytokines TNF- $\alpha$ , IL-6 and IL-1 $\beta$  that can promote cancer cell proliferation, angiogenesis and metastasis [111], and increase the population of Cancer Stem Cells (CSCs) [112]. Another key pathway that has been implicated in expanding the CSC population is the Notch signaling, however, the interplay among these three crucial pro-tumor factors - inflammation, Notch signaling, and CSC population - has received little attention.

Jagged1 - both in its transmembrane form (present on the cell surface) and soluble form (secreted by stromal endothelial cells) - appear to be the most potent inducer of Notch signaling in maintaining and expanding the CSC population [33, 27], hence enabling the development of macrometastases that claim more than 90% of cancer deaths [32]. Consequently, Jagged1 levels are overexpressed in CSCs as compared to non-CSCs [33, 99, 100, 25], and are associated with poor survival and recurrence [31]. Hence, understanding Notch-Jagged signaling is essential to control the CSC pool and the tumor spread.

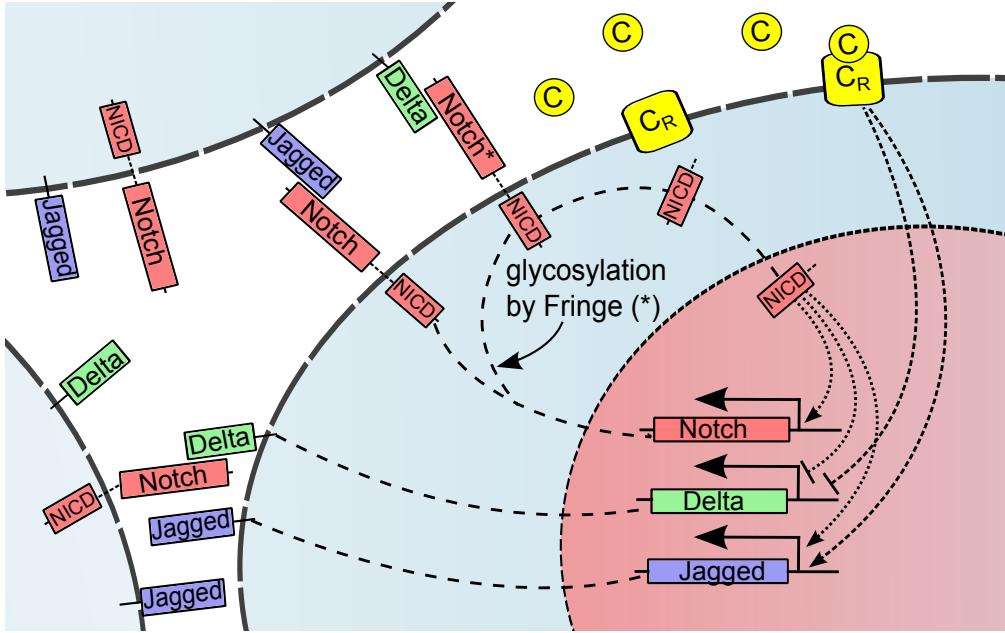


Figure 3.1: Overview of the intra-cellular and inter-cellular Notch signaling pathway. The transmembrane receptor of one cell (Notch), binds to the transmembrane ligand (Delta or Jagged) belonging to the neighboring cell. This trans-interaction leads to the cleavage and release of Notch Intracellular Domain (NICD). NICD migrates to the nucleus and modulates the transcription of many genes. This modulation indirectly leads to the transcriptional activation of Notch and Jagged and inhibition of Delta. Interaction between Notch receptor and ligands (Delta or Jagged) of the same cell (cis-interaction) leads to the degradation of both the receptor and the ligand. Glycosylation of Notch by Fringe modifies Notch to have higher affinity for binding to Delta and lower affinity for binding to Jagged. Inflammatory cytokines ( $C$ ) can bind to their receptor ( $C_R$ ) generating a signal that leads to the transcriptional activation of Jagged and inhibition of Delta.

Multiple mechanisms can regulate this juxtacrine Notch-Jagged signaling. It is inhibited by the glycosylation of Notch by Fringe that increases the affinity of Notch for Delta, but decreases that for Jagged [58] (Figure 3.1). On the other hand, it is promoted by inflammatory cytokines such as TNF- $\alpha$ , IFN- $\gamma$ , IL-6, and IL-1 $\beta$  that can increase the production of Jagged, and some of them can also decrease the production of Delta [36, 113, 114, 115, 116], thereby making more of Jagged than Delta available to bind to Notch (Figure 3.1). Also, Jagged can activate itself via inflammation by promoting the production of NF- $\kappa$ B [117, 74] that mediates the activation of Jagged through TNF- $\alpha$  [113]. Furthermore, autocrine activation of cytokines that induce Jagged such as IL-6 [118] can ramp up this increase in Jagged.

However, the understanding of the effect of the inflammatory cytokines on Notch signaling and consequently the CSC population remains elusive.

Here, we elucidate how inflammation regulates cell-fate decisions mediated by Notch signaling, and consequently impacts the CSC population. We show that inflammation increases, but Fringe decreases, the range of physiological conditions for the existence of the hybrid S/R state, and hence the cell population that communicate via Notch-Jagged signaling. Further, we found that mutual activation between NF-kB and Jagged can amplify this increase in population, and therefore allows CSCs with high levels of Jagged to expand the CSC pool and mutually reinforce the CSC state through Notch-Jagged signaling. Thus, chronic inflammation augments the mutually reinforcing CSC population, as specifically seen in basal-like breast cancer. As the patients undergoing surgery, chemotherapy and radiotherapy often elicit inflammatory response [119, 120], our results indicate how cancer can potentially benefit from current therapies.

### 3.1 Model Formulation

For theoretical investigation of the role of inflammation on Notch-Delta-Jagged (NDJ) signaling, we extended our previously model (Eqs 1.16-1.19) that elucidates the role of Jagged in Notch signaling circuit. We first incorporate the effect of inflammation as an external signal that increase the production of Jagged and decrease that for Delta. This effect of inflammation is represented by shifted Hill functions that can increase ( $\lambda > 1$ ) or decrease ( $\lambda < 1$ ) the production rates of Delta and Jagged. In this case, the dynamics of Notch ( $N$ ), Delta ( $D$ ), Jagged ( $J$ ) and NICD ( $I$ ) are given by:

$$\frac{dN}{dt} = N_0 H^{S+}(I) - k_C N(D + J) - k_T N(D_{ext} + J_{ext}) - \gamma N \quad (3.1)$$

$$\frac{dD}{dt} = D_0 H^{S-}(I) H^{S-}(C) - k_C DN - k_T DN_{ext} - \gamma D \quad (3.2)$$

$$\frac{dJ}{dt} = J_0 H^{S+}(I) H^{S+}(C) - k_C JN - k_T JN_{ext} - \gamma J \quad (3.3)$$

$$\frac{dI}{dt} = k_T N(D_{ext} + J_{ext}) - \gamma_I I \quad (3.4)$$

where  $C$  represents the cytokines (inflammatory signal). The shifted Hill function  $H^{S-}(C)$  and  $H^{S+}(C)$  represents the effect of inflammation that inhibits the production of Delta and increase that of Jagged.

We also consider the case of auto-amplification of Jagged that can occurs via its interaction with NF-kB. In this case, the equation 3.3 becomes:

$$\frac{dJ}{dt} = J_0 H^{S+}(I) H^{S+}(C) H^{S+}(J) - k_C J N - k_T J N_{ext} - \gamma J. \quad (3.5)$$

where the shifted Hill function  $H^{S+}(J)$  represents the effect of Jagged self activation.

### 3.2 Inflammation promotes the hybrid Sender/Receiver phenotype by increasing Notch-Jagged signaling over Notch-Delta signaling

As a first step towards understanding the effect of inflammation on Notch-Delta-Jagged (NDJ) signaling, we analyze the dynamics of the NDJ circuit at two fixed levels of inflammatory cytokines, when the circuit is driven by different levels of external Jagged ( $J_{ext}$ ) available to bind to Notch. This result is presented as a bifurcation diagram where  $J_{ext}$  acts as a control parameter (Figure 3.2). We see that at low levels of inflammation, for small  $J_{ext}$ , the cell has low Notch activity and low levels of Notch signal (NICD) (Figure 3.2A), i.e. it behaves as a Sender (S) - (low receptor, high ligand). As  $J_{ext}$  increases, the cell attains the hybrid (S/R) state where it has medium levels of both the receptor and ligands, therefore acting as both a Sender and Receiver. Further increase in  $J_{ext}$  leads to high levels of Notch signaling (NICD) in a cell, thus it acts as a Receiver (R). Similar behavior is seen during high inflammatory conditions; however, the range of values of  $J_{ext}$  for which the hybrid S/R state exists is much larger for the case of high inflammatory signal than that for low inflammatory signal (compare the length of the red curve in Figure 3.2B vs. that in Figure 3.2A). Further, when the cell is driven by different levels of external Delta ( $D_{ext}$ ), inflammation expands the range of values of  $D_{ext}$  for the existence of the hybrid S/R state, therefore demonstrating that the effect of inflammation on NDJ signaling is independent of the external ligand that activates the Notch signaling in the cell (Figure S2 (*to be added in appendix*)).

These results indicate that inflammation plays an important role in expanding the range of physiological conditions for which cells can adopt this hybrid S/R state - (medium ligand, medium receptor). Such an expression profile can facilitate bidirectional signaling between two cells, and can be crucial in tumor-stroma crosstalk. For instance, in ovarian cancer, both Notch and Jagged1 are overexpressed by ovarian cancer cells and the mesothelial cells [31]. Silencing of Jagged1 in stromal cells as well as in tumor cells impaired the tumor growth, and silencing it both in the tumor and stroma was shown to be synergistic [121]. Furthermore, our results are corroborated by observations specifically in the basal-like breast cancer (BLBC), where Notch, Jagged and IL-6 receptor (IL-R) are overexpressed as compared to other breast cancer subtypes [31, 114], IL-6 mediates communication between cells through Notch3 and Jagged1, Notch3 facilitates the autocrine production of IL-6 [114], and the overexpression of Jagged1 correlates with poor outcome [33].

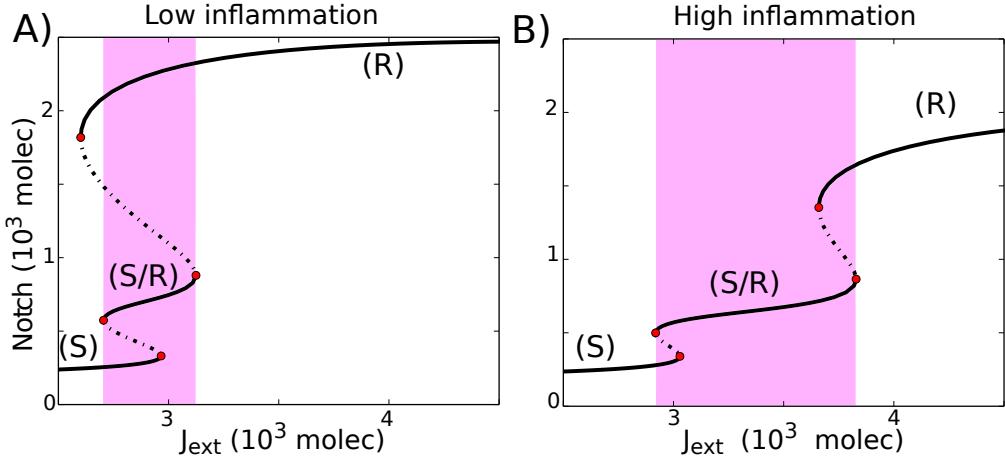


Figure 3.2: Bifurcation curves Notch-Delta-Jagged circuit: Bifurcation curves of Notch protein levels on the membrane as a function of the number of external Jagged ( $J_{ext}$ ) molecules, for different values of the external inflammatory signal ( $C$ ). In both cases, for low  $J_{ext}$  the cell is a Sender (S) and with the increase in  $J_{ext}$  it undergoes a transition to the receiver (R) state through an intermediate hybrid (S/R) state. A) represents the case for low inflammatory signal ( $C = 1000$ ), while B) represents the same for high inflammatory signal ( $C = 3000$ ). Solid lines represent stable steady states, and dotted lines represent unstable steady states. The shaded pink area represents the range of corresponding parameters for the existence of the hybrid S/R state. Bifurcation curves for other proteins as a function of  $J_{ext}$  can be found in the appendix (Figure S3).

To further investigate the role of inflammation in NDJ signaling, a phenotype diagram (two-parameter bifurcation diagram) driven by two control parameters - external inflamma-

tory signal ( $C$ ) and the number of external ligands ( $D_{ext}$  and  $J_{ext}$ ) - was calculated (Figure 3.3). These diagrams demonstrate multiple phases (i.e. sets of co-existing phenotypes in the same physiological conditions): three phases where cells can attain only one phenotype -  $\{S\}$ ,  $\{S/R\}$  and  $\{R\}$ , three phases where cells can attain any of the two phenotypes -  $\{S,R\}$ ,  $\{S, S/R\}$  and  $\{S/R, R\}$ , and one phase where cells can attain any of the three phenotypes -  $\{S, S/R, R\}$ . The selection of the phenotype, when more than one of them is possible, depends on the local microenvironment of the cell, and its individual genetic and epigenetic landscape.

For low levels of the inflammatory cytokines ( $C$ ), the hybrid ( $S/R$ ) state can only exist when other states are possible, i.e. in the  $\{S, S/R, R\}$ ,  $\{S/R, R\}$ , and  $\{S, S/R\}$  phases. However, for high levels of cytokines ( $C$ ), the hybrid  $S/R$  state also exists in a monostable  $\{S/R\}$  phase, i.e. it can be the only possible state for a given range of physiological parameters. Further, as a result of high inflammation, the range of physiological conditions for which the hybrid  $S/R$  state can exist, either alone or as one of the multiple states, increases significantly, irrespective of the external ligand driving the cell (compare the area bounded by dashed lines in Figure 3.3B vs. that in Figure 3.3A).

Because not all cytokines that activate Jagged also inhibit Delta, we plotted these phenotype diagrams for the case when the cytokine ( $C$ ) only activates Jagged, but does not inhibit Delta (to be added in appendix). We found that the increase in the range of physiological conditions for which the hybrid  $S/R$  state can exist, either alone or as one of the multiple states, is larger when both Jagged is activated and Delta is inhibited (Figure 3.3).

### 3.3 Self-activation of Jagged amplifies the effect of inflammation

NF-kB, the inflammatory response transcription factor, mediates the activation of Jagged through TNF- $\alpha$  [113]. Also, Jagged can activate NF-kB [117, 74], thereby leading to its self-activation. The self-activation of Jagged can also happen via autocrine production of IL-6 that activates Jagged [114, 104].

Next, we include this effect of indirect self-activation of Jagged (Equation 3.5). This self-activation is represented by a term ( $\lambda_J^J$ ) denoting the fold-change in expression due to

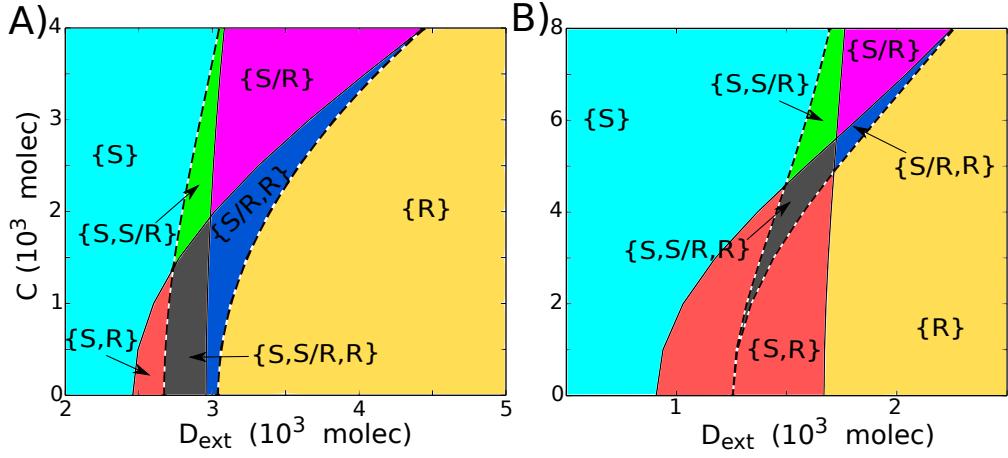


Figure 3.3: Phenotype diagram for the Notch-Delta-Jagged circuit driven by external inflammatory signal ( $C$ ). Phenotype diagram as a function of the external inflammatory signal ( $C$ ) and number of molecules of the external ligands A) for the case of external signal mediated by Jagged ( $J_{ext}$ ), for a fixed level of  $D_{ext} = 500$  molecules. B) for the case of external signal mediated by Delta ( $D_{ext}$ ), for a fixed level of  $J_{ext} = 500$  molecules. Phenotype diagram similar to A) but for fixed level of  $D_{ext} = 1000$  molecules, and one similar to B) but for a fixed level of  $J_{ext} = 1000$  molecules are given in SI section (Figure S4). The area bounded by dashed lines shows all the phases that contain S/R as one of the co-existing states -  $\{S/R\}$  (pink area),  $\{S, S/R, R\}$  (black area),  $\{S, S/R\}$  (green area) and  $\{S/R, R\}$  (blue area).

self-activation. We plot the phenotype-diagram when the circuit is driven by variable levels of external Jagged ( $J_{ext}$ ) and different strengths of self-activation ( $\lambda_J^J$ ). As the levels of  $\lambda_J^J$  increase, we see an increase in the range of physiological conditions for which the hybrid S/R state can exist, either alone or as one of the multiple states, irrespective of the external ligand driving the circuit - Jagged ( $J_{ext}$ ) (Figure 3.4B) or Delta ( $D_{ext}$ ) (Figure 3.4E). Also, when Notch signaling in a cell is majorly activated by external Jagged ( $J_{ext}$ ) rather than Delta ( $D_{ext}$ ), the hybrid S/R state can be the only possible state for a given range of physiological parameters, i.e. the existence of monostable  $\{S/R\}$  phase (the pink colored area exists in Figure 3.4B, but not in Figure 3.4E). This result is consistent with previous work showing that a cell driven mainly by external Delta ( $D_{ext}$ ) can act only as a Sender (S) or a Receiver (R) (i.e.  $\{S\}$ ,  $\{R\}$  and  $\{S, R\}$  phases).

Overall, our results reveal that irrespective of the external ligand activating the Notch signaling in a cell, the self-activation of Jagged amplifies the role of inflammatory cytokines in increasing the range of existence of hybrid (S/R) state and promoting Notch-Jagged sig-

naling. Our results are consistent with experimental observations in long-term immune TLR (Toll-like Receptor) response, where auto-amplification of Jagged1 is crucial in maintaining the high levels of Notch signaling (NICD) that activate the macrophages and generate an adaptive immune response [115].

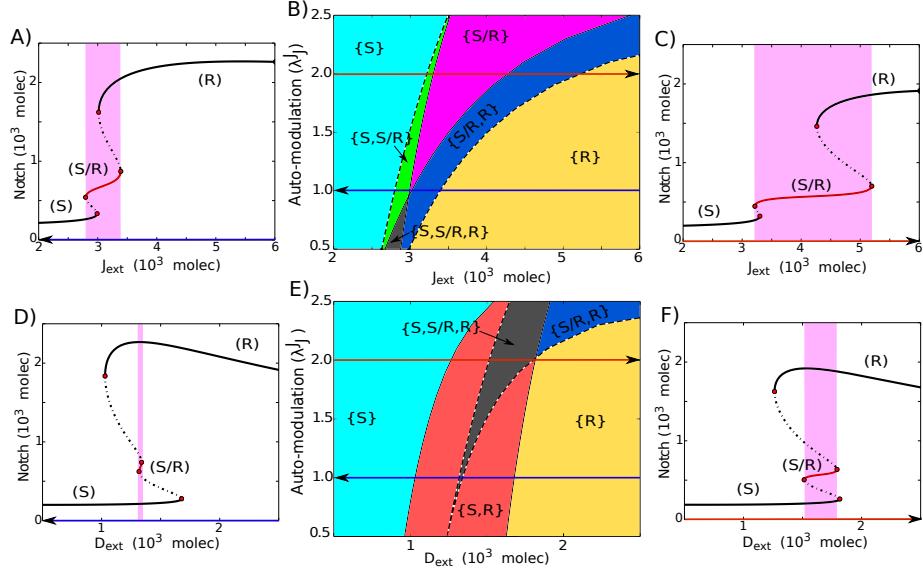


Figure 3.4: Phenotype diagram and bifurcation curves for Notch-Delta-Jagged circuit with the self-activation of Jagged. Phenotype diagram (B, center) as a function of external Jagged ( $J_{ext}$ ) and different levels of Jagged self-activation. A) Bifurcation curves of Notch protein levels on the membrane as a function of the number of external Jagged for no self-modulation of Jagged ( $\lambda_J^J = 1$ ), and C) for self-activation of Jagged ( $\lambda_J^J = 2$ ). Phenotype diagram (E, center) as a function of external Delta ( $D_{ext}$ ) and different levels of Jagged self-activation. D) Bifurcation curves of Notch protein levels on the membrane as a function of the number of external Delta for no self-modulation of Jagged ( $\lambda_J^J = 1$ ), and F) for self-activation of Jagged ( $\lambda_J^J = 2$ ). For all these figures, the level of inflammatory signal is fixed ( $C = 2000$ ). In A), C), D) and F), solid lines represent stable steady states; dotted lines represent unstable steady states; and the pink area represents the range of existence of the hybrid S/R state. In B) and E), the area bounded by dashed lines shows all the phases that contain S/R as one of the co-existing states - {S/R} (pink area), {S, S/R, R} (black area), {S, S/R} (green area) and {S/R, R} (blue area).

### 3.4 Fringe acts antagonistic to inflammation and inhibits Notch-Jagged signaling

Next, we evaluate the role of both the inflammatory signal ( $C$ ) and the glycosyltransferase Fringe on Notch-Delta-Jagged signaling. Fringe increases the binding affinity of Notch for Delta ( $\lambda_J^F > 1$ ), but decreases that for Jagged ( $\lambda_J^F < 1$ ) [58]. Our results are presented in phenotype diagrams when the circuit is driven by variable levels of both external Jagged ( $J_{ext}$ ) and external Delta ( $D_{ext}$ ). In this case, the effect of Jagged auto-activation is not considered and the model is the same as presented in Eqs. 3.1-3.4.

The different phenotype diagrams represent different combinations of the effect of the external inflammatory cytokines ( $C$ ) and that of Fringe (Figure 3.5). First, at low levels of inflammation, the increase in the effect of Fringe decreases the range of physiological conditions for which cells can both send and receive signals (hybrid S/R state), either alone or in co-existence with other states (compare the area bounded by dashed lines in Figure 3.5B vs. that in Figure 3.5A). Consequently, Fringe increases the range of conditions for which the cells can either send or receive signals - ( $\{S\}$ ,  $\{R\}$  and  $\{S, R\}$  phases) (compare the total area of the phenotype diagram minus the area bounded by dotted lines in Figure 5B vs. that in Figure 3.5A). Similar effect of Fringe is also observed at high levels of inflammation ( $C$ ) (Figure 3.5C, D). Thus, high levels of Fringe inhibit Notch-Jagged signaling and hence the existence of the hybrid S/R state. Conversely, high inflammation ( $C$ ) leads to a significant increase of the range of physiological conditions for which cells can both send and receive signals, for both the case of low Fringe effect (compare the area bounded by dotted lines in Figure 3.5C vs. that in Figure 3.5A) and that of high Fringe effect (compare the area bounded by dotted lines in Figure 3.5D vs. that in Figure 3.5B). Because Notch-Jagged signaling is often implicated in many aspects of tumor progression in multiple cancers [31], our results indicating the tendency of Fringe to inhibit Notch-Jagged signaling are consistent with its recently identified role as a tumor suppressor in prostate cancer, lung cancer and basal-like breast cancer (BLBC) [69, 70, 71]. Further, loss of Fringe facilitates Notch-Jagged cell-cell signaling that is instrumental in the origin of basal-like tumors [69].

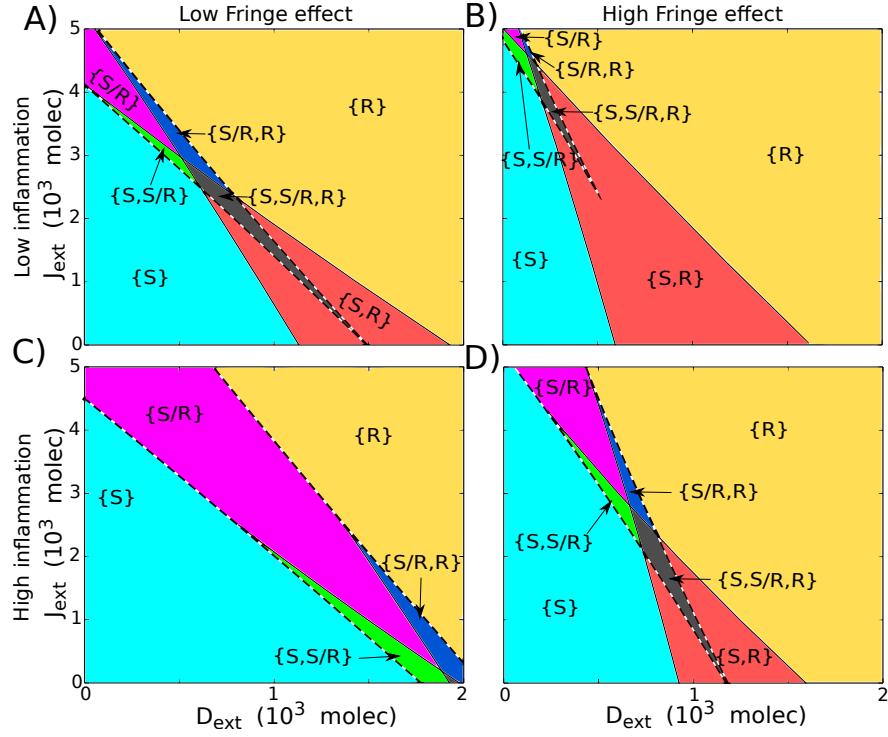


Figure 3.5: Phase-diagrams of Notch-Delta-Jagged circuit as a function of  $J_{ext}$  and  $D_{ext}$ , for different levels of the Fringe and the external inflammatory signal. Phenotype diagram as a function of external Delta and external Jagged. A) For low fringe effect ( $\lambda_D^F = 3$  and  $\lambda_J^F = 0.3$ ) and low inflammatory signal ( $C = 2000$ ). B) For high fringe effect ( $\lambda_D^F = 5$  and  $\lambda_J^F = 0.1$ ) and low inflammatory signal ( $C = 2000$ ). C) For low fringe effect ( $\lambda_D^F = 3$  and  $\lambda_J^F = 0.3$ ) and high inflammatory signal ( $C = 6000$ ). D) For high fringe effect ( $\lambda_D^F = 5$  and  $\lambda_J^F = 0.1$ ) and high inflammatory signal ( $C = 6000$ ). The area bounded by dashed lines shows all the phases that contain S/R as one of the co-existing states - {S/R} (pink area), {S, S/R, R} (black area), {S, S/R} (green area) and {S/R, R} (blue area).

### 3.5 Inflammation increase CSC population through Notch-Jagged signaling

Recent experiments show that various cytokines increase the CSC population, however, the underlying molecular mechanisms remains elusive. We here hypothesize that this effect of inflammation is mediated, at least in part, via the Notch-Jagged signaling. To test this hypothesis, we analyzed the genetic profile of 211 breast cancer patients - 140 of which have luminal-like BC subtype – a subtype where Fringe is not lost and Jag1 is not overexpressed,

and 71 of which belong to triple-negative or basal-like subtype – a subtype in which Fringe is lost, and stromal cells overexpress Jag1 [69, 122].

We evaluate the correlation between the expression levels of Jag1 and inflammation mediators TNF- $\alpha$ , NF-kB and IL-6. Our results suggest a strong correlation between the expression levels of Jagged1 and the levels of these inflammatory proteins for BLBC (Figure 3.6), but not a significant correlation is found for Luminal-like BC, therefore suggesting that the effect of inflammation in increasing the concentration of Jag1 is subtype-specific (Figure 3.6). These results are consistent with the observation that Fringe is lost in BLBC, thus promoting Notch-Jagged signaling. We also observe a significant correlation between Jag1 and the CD44+ CD24- cell surface expression that are biomarkers of CSCs, (Figure 3.6) thereby indicating that high levels of Notch-Jagged signaling, as promoted by inflammation, can promote CSC expansion.

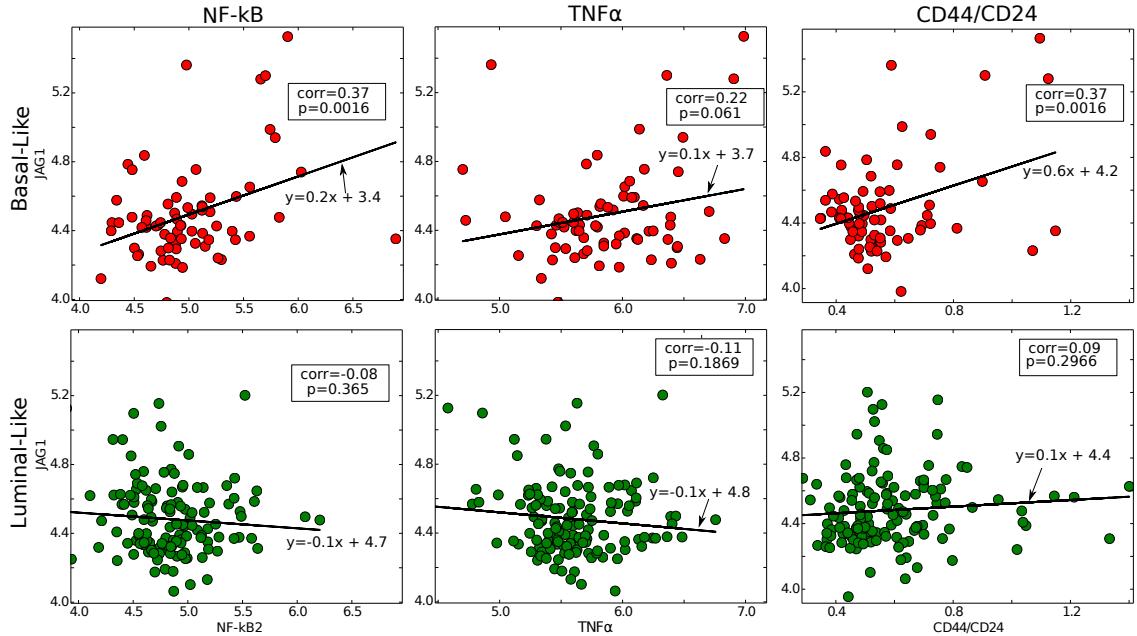


Figure 3.6: Correlation between JAG1 and NF- $\kappa$ B, TNF- $\alpha$ , and CD44/CD24 for Basal-like and Luminal-like Breast Cancer. Gene expression data obtained from 211 patients diagnosed with breast cancer before any therapy. The variable corr represents the Pearson correlation between the variables and the probability of finding a value at least as extreme in an uncorrelated system is represented by  $p$ . Black line represents the linear fit of the data.

To clarify this interplay between inflammation, Notch signaling and CSCs, we simulate

the effect of inflammation in a population of cells that interact via NDJ signaling. We first measured the distribution of cells in terms of their levels of NICD and Jagged. We found that all cells present a Sender/Receiver phenotype, i.e. have high levels of Notch, Jagged and NICD, and low levels of Delta. Among these cells that can both send and receive signals, we defined the cluster of cells with the maximum levels of Notch, Jagged and NICD as Cancer Stem Cells (CSCs), based on observations at least in glioblastoma, pancreatic cancer, colon cancer and basal-like breast cancer that Notch and Jagged are overexpressed in CSCs as compared to non-CSCs [33, 99, 100, 25] (Figure 3.7A).

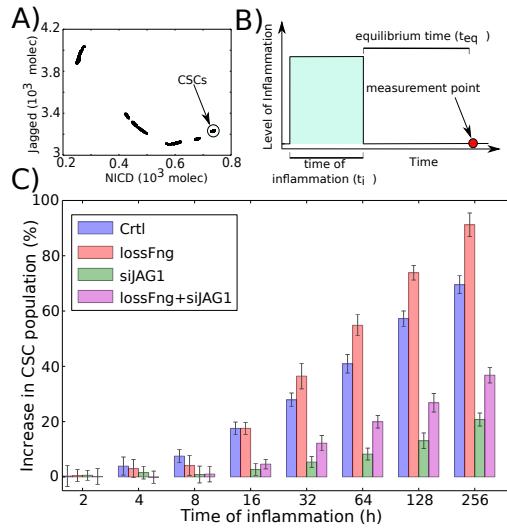


Figure 3.7: Simulating Notch-Inflammation-CSC interplay on a population of cells A) The CSCs are identified as the cells with both high levels of NICD and Jagged, as indicated. B) Schematic illustration of the simulation. A bi-dimensional layer of 100x100 cells interacting through Notch signaling is simulated. The cells start with a random configuration. After reaching the equilibrium, the cluster of cells that present the maximum levels of Notch, Jagged and NICD are defined as CSC population. Then, a high inflammatory signal is induced for different range of duration and the CSC population is measure after the system reaches the equilibrium. B) Percentage of increase in CSC population compared to the case of no inflammatory signal. The bars represent the average of 10 simulations and the color represents different cases: (Ctrl) represent the cells with standard values of the parameters. (lossFng) represent the case where fringe is lost and therefore the binding affinity of Notch to Jagged is similar to that for Delta. (siJag1) represent the case where Jagged has a higher degradation rate representing the effect of treatment with siJag. (lossFng+siJag) represents the case of both effects of loss of Fringe and siJag combined.

To evaluate the effect of inflammation on CSC population, we exposed the cells to in-

flammation for a time ( $t_i$ ) and then measured the CSC population after allowing the system to equilibrate for time ( $t_{eq}$ ) (Figure 3.7B). Our results indicate that acute inflammation does not significantly expand the CSC population, but chronic inflammation can increase the CSC population by more than 60% (Figure 3.7C, blue bars). Besides, loss of Fringe amplifies the increase of CSC by chronic inflammation, but not by acute inflammation (Figure 3.7C, red bars).

We further show this increase of CSCs by inflammation can be decreased by increasing the degradation rate of Jagged (to mimic the effect of a siRNA against Jagged) (Figure 3.7C, green bars), even when fringe is lost (Figure 3.7C, pink bars). However, a similar attenuation in the effect of inflammation was not observed when we increased the degradation rate of Delta (to mimic the effect of a siRNA against Delta) (to be included in appendix). Together, these results suggest that inflammation expands the CSC population by augmenting Notch-Jagged signaling.

## 3.6 Discussion

Inflammatory cytokines such as TNF- $\alpha$ , IL-6 and IL-1 $\beta$  are key messengers enabling paracrine signaling between cancer cells and stromal cells to establish a pro-tumor environment [112]. Here, we propose that in addition to this paracrine role, they can also promote juxtacrine signaling (contact-dependent signaling) through Notch signaling pathway. We find that these cytokines can stabilize the hybrid Sender/Receiver (S/R) state of Notch signaling pathway that enable the cells to both send and receive signals. Specifically, this hybrid S/R state is expected to have a critical role in cell-cell communication between cells that have undergone partial Epithelial to Mesenchymal Transition (EMT) and move collectively [27, 77]. Notch signaling observed during wound healing, a typical case of partial EMT, supports this view [82]. However, direct experimental measurements of both Notch and Jagged are needed to validate our result that the cytokines promote the hybrid S/R state of the cells. It would also be tempting to explore the effect of these cytokines in the maintenance of clusters of Circulating Tumor Cells (CTCs) that are a hallmark of the hybrid E/M phenotype [83] and can be up to 50 times more metastatic than individually migrating CTCs (mesenchymal cells) [84].

We also show that Fringe acts antagonistic to inflammation - it promotes Notch-Delta signaling and inhibits Notch-Jagged signaling. Therefore, at high levels of Fringe, Notch-Delta signaling, but not Notch-Jagged signaling, is prevalent. Consequently, in cases where Notch-Jagged signaling is more prevalent than Notch-Delta signaling, such as during cancer progression, we expect the Fringe effect to be low. Our results are consistent with experiments showing that Fringe acts as a tumor suppressor in multiple cancers and is often lost during tumor progression [69, 70, 71]. Furthermore, our results offer a putative mechanistic understanding of the experimental observation that NF- $\kappa$ B-Jag1 axis could expand the CSC population specifically in breast-like basal cancer (BLBC), but not in other luminal subtypes [33] - loss of Fringe in BLBC but not in others [69] augments the role of Notch-Jagged signaling and consequently expands the CSC population that has higher expression of Notch and Jagged than that in non-CSCs [33, 99, 100, 25].

Further, we show that self-activation of Jagged via NF- $\kappa$ B can amplify the effect of cytokines in promoting the hybrid S/R state. This mutual activation between NF- $\kappa$ B and Jagged can increase the population of Cancer Stem Cells (CSCs) by activating Notch signaling in neighboring cells, as seen specifically in basal-like breast cancer [33]. CSC population can also be augmented by soluble form of Jagged that is secreted by endothelial cells [27], and by other cytokines such as TNF- $\alpha$ , IL-1 $\beta$  and IL-6 [103, 123, 102], that can promote Notch-Jagged signaling. However, the underlying molecular circuitry through which both the inflammatory cytokines and Notch-Jagged signaling increase the CSC pool remains to be largely uncovered. Future experimental studies should elucidate the molecular connections between Notch signaling, inflammatory cytokines and the decision-making circuit for stemness - LIN28/ let-7 [124]. These connections can then be incorporated into the theoretical model for LIN28/let-7 to gain a comprehensive understanding of how inflammation and Notch signaling can regulate stemness [76].

To conclude, we present the first step towards understanding the molecular interplay of Notch-Jagged signaling, inflammation and CSC population. As the attempts to gain a quantitative understanding of cancer signaling networks progress [125, 126], our framework presented here can be expanded to a population-level modeling of cancer cells and stromal cells communicating with each other through the inflammatory cytokines as well as Notch-Jagged pathway, and modulating the plasticity of cancer cells to gain stemness. It may

be noted that these cytokines can also be used by cancer to enslave some stromal cells that can further promote CSC population, for instance, by converting normal fibroblasts to Cancer Associated Fibroblasts [127]. Such analysis is crucial to gain valuable insights into interfering with the tumor-stroma cross-talk and planning a “cyber war” on cancer [109], especially when cancer apparently turns its therapies into opportunities to benefitting from resultant inflammation and be more aggressive.

# Appendix I

## I.A - Derivation of the model with modulation by Fringe

To include the effect of glycosylation of Notch by Fringe we considered two types of Notch, whatever it is modified or not by Fringe. Also, NICD (I) activates Fringe and because of that, we considered that the fraction of glycosylated Notch increases with the increase of NICD (I).

For simplicity, lets first consider the effect of Fringe only in the *trans*-interaction between Notch-Delta. The *cis* interaction between Notch-Delta, and both *cis* and *trans* interaction between Notch and Jagged are omitted, however, the same argument can be used to understand the changes in these interactions due to glycosylation.

Lets consider the following reaction: Notch of the cell  $N$  binds to external Delta  $D_{ext}$  creating a complex  $ND_{ext}$ :



The equations for this reaction for both the glycosylated Notch and the non-glycosylated Notch,  $N^*$  and  $N'$ , respectively, are given by:

$$\frac{dN^*}{dt} = N_0^* H^S(I, \lambda_N) - \gamma N^* - (k_{T+}^* N^* D_{ext} - k_{T-}^* [N^* D_{ext}]) \quad (3.7)$$

$$\frac{dN'}{dt} = N_0' H^S(I, \lambda_N) - \gamma N' - (k'_{T+} N' D_{ext} - k'_{T-} [N' D_{ext}]) \quad (3.8)$$

where the  $k_{T+}$  and  $k_{T-}$  represents the binding and unbinding rate respectively. Then, the equations that represent the  $[ND_{ext}]$  complex are:

$$\frac{d[N^* D_{ext}]}{dt} = k_{T+}^* N^* D_{ext} - k_{T-}^* [N^* D_{ext}] - k_I^* [N^* D_{ext}] \quad (3.9)$$

$$\frac{d[N'D_{ext}]}{dt} = k'_{T+}N'D_{ext} - k'_{T-}[N'D_{ext}] - k'_I[N'D_{ext}], \quad (3.10)$$

where the last term represents the release of the signal (NICD) that occurs with a rate  $k_I$  and leads to the degradation of both proteins. Assuming the quasi-steady-state for  $[N^*D_{ext}]$  and  $[N'D_{ext}]$  we obtain:

$$[N^*D_{ext}] = \frac{k^*_{T+}}{k^*_{T-} + k^*_I} N^*D_{ext} \quad (3.11)$$

$$[N'D_{ext}] = \frac{k'_{T+}}{k'_{T-} + k'_I} N'D_{ext}. \quad (3.12)$$

Now, lets define  $N = N' + N^*$ . We can write  $N' = (1 - f)N$  and  $N^* = fN$ , where  $f : [0, 1]$  represents the fraction of Notch that has been modified by Fringe.

Then,

$$\frac{dN}{dt} = N_0 H^S(I, \lambda_N) - \gamma N - k_T(1 + af)D_{ext} \quad (3.13)$$

where,

$$k_T = \frac{k'_{T+}k'_I}{k'_{T-} + k'_I} \quad (3.14)$$

and

$$a = \frac{k^*_{T+}k^*_I}{k'_{T+}k'_I} \frac{(k'_{T-} + k'_I)}{(k^*_{T-} + k^*_I)} - 1. \quad (3.15)$$

Similar procedure can be done for both Notch-Jagged and cis interaction. Then, the model becomes:

$$\frac{dN}{dt} = N_0 H^S(I, \lambda_N) - N[(k_C D + k_T D_{ext})[1 + a_d f(I)] + (k_C J + k_T J_{ext})[1 + a_j f(I)]] - \gamma_N N \quad (3.16)$$

$$\frac{dD}{dt} = D_0 H^S(I, \lambda_D) - k_C D N [1 + a_d f(I)] - k_T D N_t - \gamma_D D \quad (3.17)$$

$$\frac{dJ}{dt} = J_0 H^S(I, \lambda_J) - k_C J N [1 + a_j f(I)] - k_T J N_t - \gamma_J J \quad (3.18)$$

$$\frac{dI}{dt} = k_T N (D_{ext}[1 + a_d f(I)] + J_{ext}[1 + a_j f(I)]) - \gamma_I I. \quad (3.19)$$

We considered that  $f = H^+(I)$  is a positive Hill function that increases when the signal increases. In this case, we can rewrite the term related to the Fringe effect in terms of shifted

Hill functions:  $k(I) = k[1 + aH^+(I)] = kH^S(I, \lambda^F)$  where  $\lambda^F = 1 + a$ . Then, the equations above becomes:

$$\frac{dN}{dt} = N_0 H^S(I, \lambda_N) - N[(k_C D + k_T D_{ext}) H^S(I, \lambda_D^F) + (k_C J + k_T J_{ext}) H^S(I, \lambda_J^F)] - \gamma_N N \quad (3.20)$$

$$\frac{dD}{dt} = D_0 H^S(I, \lambda_D) - k_C H^S(I, \lambda_D^F) ND - k_T DN_t - \gamma_D D \quad (3.21)$$

$$\frac{dJ}{dt} = J_0 H^S(I, \lambda_J) - k_C H^S(I, \lambda_J^F) NJ - k_T J N_t - \gamma_J J \quad (3.22)$$

$$\frac{dI}{dt} = k_T N [D_{ext} H^S(I, \lambda_D^F) + J_{ext} H^S(I, \lambda_J^F)] - \gamma_I I \quad (3.23)$$

## I.B - Parameter sensitivity analysis

We perform the sensitivity of the model applied to the steady state, by quantifying the changes of the signal as function of the changes of the parameters. Each parameter was increased and decreased its value 10% off its values described in table 1.1. The relative change in the steady state of the signal is presented in figure 3.8. These results indicate that the most important parameters - in which a change in its value by 10% generates a higher change in the signal - are the same in the three models, thus showing a good consistency among the models. The most important parameters are: the production and degradation rate of Notch,  $N_0$  and  $\gamma_N$  respectively, the production rate of Delta  $D_0$  and the trans-activation rate  $k_T$ .

For most parameters, changes in its value by 10% leads to very small changes in the signal. On one hand, this means that the results and predictions of the model are quite robust by changes in the parameters. On the other hand, this means that some extra caution must be taken when planning an experimental validation of the values of these parameters. The most simple experiment for validating this model is to measure the intensity of the signal using a report protein. If this kind of experiment is used to fit the parameters, we should expect that many parameters will have a large range of values in which they fit well the

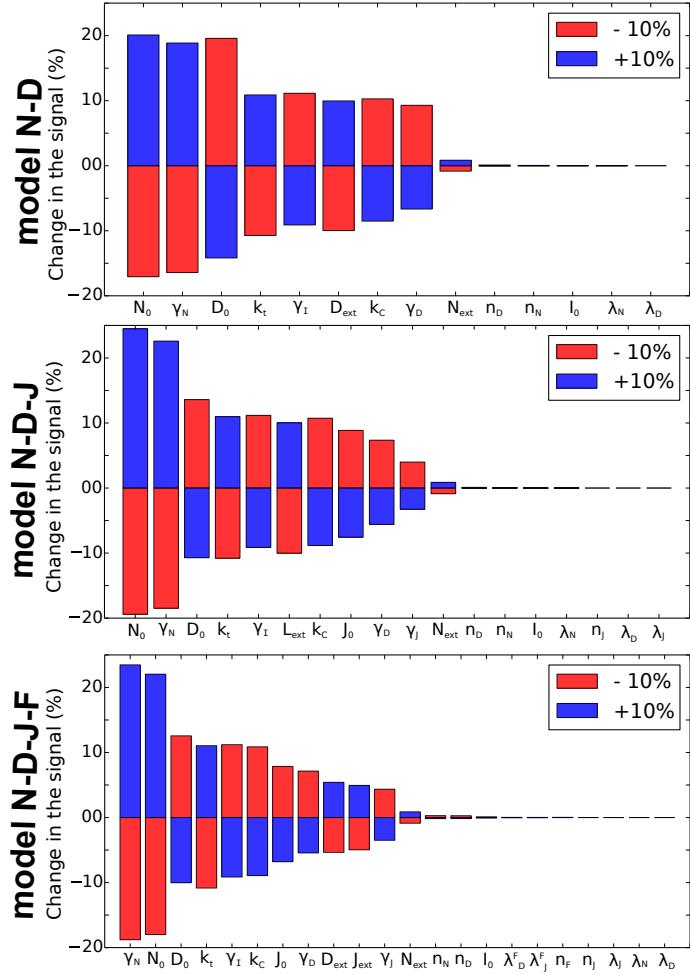


Figure 3.8: *Parameter sensitivity analysis. Relative changes in the steady state of the signal (NICD) with respect to their value in the stable state when each parameter is increased and decreased by 10% over its standard value described in table 1.1.*

experimental data. This should occur once changes in the majority of the parameters leads to small changes in the signal therefore characterizing the model as sloppy [128]. For this reason, experiments to validate the values of the parameters should be carefully designed.

## Part II

# Topics in bioinformatics

*“É que quando eu cheguei por aqui eu nada entendi.”*

*Sampa - Caetano Veloso.*

# **Chapter 4**

## **Supervised Variational Relevance Learning, an analytic geometric feature selection with applications to omic datasets**

In this chapter, we introduce Supervised Variational Relevance Learning (Suvrel), a variational method to determine metric tensors to define distance based similarity in pattern classification, inspired in relevance learning. The variational method is applied to a cost function that penalizes large intraclass distances and favors small interclass distances. We find analytically the metric tensor that minimizes the cost function. Preprocessing the patterns by doing linear transformations using the metric tensor yields a dataset which can be more efficiently classified. We test our methods using publicly available datasets, for some standard classifiers. Among these datasets, two were tested by the MAQC-II project and, even without the use of further preprocessing, our results improve on their performance.

### **Introduction**

Many of the current biomolecular technologies permit measuring a large number of variables, generating high-dimensional data, often involving thousands or even millions of features.

Very often a large fraction of these features are redundant or irrelevant and in this case, feature selection techniques are of fundamental importance for constructing a predictive model. Furthermore, from a biological point of view, inspecting a shorter list of biomarkers candidates can shed light on the biological process involved. Other benefits include shorter training times and the reduction of overfitting.

The archetypal case is the use of feature selection applied to DNA microarrays data, where there are on the order of ten thousand genes, whereas only a small subset of these genes is relevant, and the number of experiments range from a few tens up to a few hundreds. Much effort has been invested in defining a good methodology for outcome prediction using microarray data [129]. It has been shown that feature selection methods have a significant influence in the accuracy of predictive tests using microarray data and surprisingly the simple Student's t-test seems to provide better results than sophisticated methods [130]. In addition, results such as those from the Microarray Quality Control (MAQC)-II study indicate that the choice of the number of features has a higher impact on the method performance than the feature selection method, see figure 5a of [129].

An essential ingredient of classification or clustering by similarity is to define an appropriate definition of similarity or alternatively, distance. Our central problem is to define a metric tensor which incorporates the statistical significance of each feature and hence defines the relevant geometrical space where the data is represented. We present a methodology for data preprocessing, Supervised Variational Relevance Learning (Suvrel), that extends ideas by [131, 132, 133, 134, 135] for relevance learning. This has also been studied under the name distance learning, e.g. see [136, 137, 138]. We suppose a scenario where a classifier will be used after preprocessing for feature selection by Suvrel has been performed. The feature selection preprocessing will hinge in determining an adequate metric tensor. The idea is to choose a metric such that patterns of the same class be represented closer to each other in comparison with elements of different groups. This is done by shrinking the non relevant dimensions and stretching the relevant ones accordingly to a measure of relevance. This new representation of the dataset can be useful to improve the classification accuracy of different methods and also for visualization of the data in low dimensions. The metric tensor can also be used to attribute statistical significant relevance to the features. But this relevance determination does not preclude using a further feature selection method which might ben-

efit from an improved starting point. The new result we present in this manuscript, is a variational method that can be solved exactly analytically. We applied this methodology to several datasets and our results indicate the advantages of representing the data in this new metric. The datasets include two sets obtained from MAQC-II project [129] that can be divided according to two clinical endpoints. For cases of more than two classes, we used a microarray leukemia dataset and a metabolite dataset.

## 4.1 Methods

A datum from an experiment  $i$  is a vector  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iF})$  in the space of features  $\mathbb{R}^F$ , where each feature  $\mu$  represents a dimension (with  $\mu = 1, \dots, F$ ). The learning and validation sets,  $\mathcal{L}$  and  $\mathcal{V}$ , are sets of pairs  $\{(\mathbf{x}_i, c_i)\}$ , where data vectors  $\mathbf{x}_i$  come together with its class label  $c_i$ . The set of class labels values that  $c_i$  can take is  $\mathcal{C} = \{c^1, c^2, \dots, c^K\}$ , where  $K$  is the total number of classes. The number of samples of each class is  $n_{c^1}, n_{c^2}, \dots, n_{c^K}$  and  $n = \sum_{a \in \mathcal{C}} n_a$  is the total number of elements of a set.

The geometry of the space of features is not established until a measure of similarity is specified. We choose to define the underlying geometry by imposing a structure through a metric tensor  $g_{\mu\nu}$ . This idea goes back to Gauss and Riemann and has entered Physics via ideas in gravitation. In the Computer Science and Statistics literature, where more general types of similarity measures are used, the Riemann distance inspired the linear mapping called the Mahalanobis distance [139]. The relevant information to determine the metric structure, to be obtained by a variational process, is the class membership of the vectors of the learning set. Features will be considered relevant if they contribute to decrease intraclass or increase interclass average distances. If we restrict the analysis to the simpler case of a diagonal metric tensor, then each dimension  $\mu$  (e.g. gene in case of microarrays datasets) has a relevance weight  $w_\mu$ , the component of the diagonal metric tensor, still to be determined by minimizing a cost function. The distance between two elements  $i$  and  $j$  is defined as:

$$d_{ij}^2 = \sum_{\mu\nu} g_{\mu\nu} \Delta x_{ij\mu} \Delta x_{ij\nu}, \quad (4.1)$$

where  $\Delta x_{ij\nu} = x_{i\nu} - x_{j\nu}$  is the difference for the  $\nu$ th feature of the patterns  $i$  and  $j$ .

Cost functions that penalizes features that increase intraclass distances while favoring

features that decrease interclass distances have been considered before, giving rise to a large body of literature which has been recently reviewed in [137, 138]. Our main contribution is that the specific form we consider is amenable to analytic manipulation so that the metric tensor can be obtained explicitly and that it can be established that the method renders a positive definite metric tensor. The cost function to be minimized with respect to the metric tensor, is

$$E(\{g_{\mu\nu}\}; \mathcal{L}) = \sum_{a \in \mathcal{C}} \langle d_{ij}^2 \rangle_{a,a} - \gamma \sum_{\langle a \neq b \rangle \in \mathcal{C}} \langle d_{ij}^2 \rangle_{a,b}, \quad (4.2)$$

where  $a$  and  $b$  refer to class index, and  $\gamma$  a parameter that controls the weight of the interclass distance term, with respect to the intraclass distance term. The notation  $\langle a \neq b \rangle$  denotes that different class pairs are included only once. The angle brackets denote an average, given by

$$\langle d_{ij}^2 \rangle_{a,b} = \frac{1}{n_{ab}} \sum_{i \in a, j \in b} d_{ij}^2, \quad (4.3)$$

where  $n_{ab} = n_a n_b$  and  $\sum_{i \in a}$  is the summation over all samples that belong to class  $a$ . Since the distance defined in (4.1) depends on a summation over the features, we can rewrite the cost function (4.2) as:

$$E = \sum_{\mu\nu} g_{\mu\nu} \epsilon_{\mu\nu}, \quad (4.4)$$

where  $\epsilon_{\mu\nu}$  is given by

$$\epsilon_{\mu\nu} = e_{\mu\nu}^{in} - \gamma e_{\mu\nu}^{out} \quad (4.5)$$

and

$$\begin{aligned} e_{\mu\nu}^{in} &= \sum_{a \in \mathcal{C}} \frac{1}{n_{aa}} \sum_{i,j \in a} \Delta x_{ij\mu} \Delta x_{ij\nu}, \\ &= \sum_{a \in \mathcal{C}} \frac{1}{n_a^2} \left( n_a \sum_{i \in a} x_{i\mu} x_{i\nu} + n_a \sum_{j \in a} x_{j\mu} x_{j\nu} \right. \\ &\quad \left. - 2 \sum_{i \in a} x_{i\mu} \sum_{j \in a} x_{j\nu} \right) \\ &= 2 \sum_{a \in \mathcal{C}} \text{cov}(\mathbf{x}_\mu^a; \mathbf{x}_\nu^a) \end{aligned} \quad (4.6)$$

$$\begin{aligned}
e_{\mu\nu}^{out} &= \sum_{<a \neq b> \in \mathcal{C}} \frac{1}{n_{ab}} \sum_{i \in a, j \in b} \Delta x_{ij\mu} \Delta x_{ij\nu}, \\
&= \sum_{<a \neq b> \in \mathcal{C}} \frac{1}{n_a n_b} \left( n_b \sum_{i \in a} x_{i\mu} x_{i\nu} + n_a \sum_{j \in b} x_{j\mu} x_{j\nu} \right. \\
&\quad \left. - \sum_{i \in a} x_{i\mu} \sum_{j \in b} x_{j\nu} - \sum_{i \in a} x_{i\nu} \sum_{j \in b} x_{j\mu} \right) \\
&= \sum_{<a \neq b> \in \mathcal{C}} [\text{cov}(\mathbf{x}_\mu^a; \mathbf{x}_\nu^a) + \text{cov}(\mathbf{x}_\mu^b; \mathbf{x}_\nu^b) \\
&\quad + (m_{a\mu} - m_{b\mu})(m_{a\nu} - m_{b\nu})] . \\
&= (K-1) \sum_{a \in \mathcal{C}} \text{cov}(\mathbf{x}_\mu^a; \mathbf{x}_\nu^a) \\
&\quad + \sum_{<a \neq b> \in \mathcal{C}} (m_{a\mu} - m_{b\mu})(m_{a\nu} - m_{b\nu}). \\
&= (K-1) \sum_{a \in \mathcal{C}} \text{cov}(\mathbf{x}_\mu^a; \mathbf{x}_\nu^a) + K^2 \text{cov}(\mathbf{m}_\mu; \mathbf{m}_\nu), \tag{4.7}
\end{aligned}$$

where  $\mathbf{x}_\mu^a$  is the set of samples of the  $\mu$  feature that belongs to the class  $a$  and  $m_{a\nu} = \frac{1}{n_a} \sum_{i \in a} x_{i\nu}$ . The first term has the sum over classes of the covariance within a class of the features. The second term is the covariance of the vector of dimension  $K$  whose components are the means over a class:  $\mathbf{m}_\mu = \{m_{a\mu}\}$ ,

$$\text{cov}(\mathbf{m}_\mu; \mathbf{m}_\nu) = \frac{1}{K} \sum_{a \in \mathcal{C}} m_{a\mu} m_{a\nu} - \left( \frac{1}{K} \sum_{a \in \mathcal{C}} m_{a\mu} \right) \left( \frac{1}{K} \sum_{a \in \mathcal{C}} m_{a\nu} \right).$$

It follows that

$$\begin{aligned}
\epsilon_{\mu\nu} &= [2 - (K-1)\gamma] \sum_{a \in \mathcal{C}} \text{cov}(\mathbf{x}_\mu^a; \mathbf{x}_\nu^a) \\
&\quad - \gamma K^2 \text{cov}(\mathbf{m}_\mu; \mathbf{m}_\nu), \tag{4.8}
\end{aligned}$$

which only depends on the data vectors and on the supervising information of which experiment  $i$  belongs to each class  $a$ .

Relevance is obtained by imposing that  $E$  is a minimum with respect to the metric tensor variations subject to a scale constraint  $\sum_{\mu\nu} g_{\mu\nu}^2 = 1$ , with  $\gamma$  fixed. The solution, obtained

using the standard method of Lagrange multipliers

$$\frac{\delta}{\delta g_{\mu\nu}} \left( E + \theta \left( \sum_{\mu\nu} g_{\mu\nu}^2 - 1 \right) \right) = 0 \quad (4.9)$$

is

$$g_{\mu\nu} = \frac{-\epsilon_{\mu\nu}}{\sqrt{\sum_{\mu'\nu'} \epsilon_{\mu'\nu'}^2}}. \quad (4.10)$$

Once we obtain the metric tensor  $g_{\mu\nu}$ , we can define a distance between the data points using eq. (4.1) or alternatively apply a coordinate transformation with the square root of the new metric tensor, but for this we have to investigate conditions that lead to a positive definite metric tensor. This is discussed in appendix 6.4, where we prove that there is a  $\gamma^* < 2/(K - 1)$  such that for  $\gamma > \gamma^*$  the metric tensor is positive definite. For the diagonal case, with

$$d_{ij}^2 = \sum_{\nu} w_{\mu} (x_{i\nu} - x_{j\nu})^2, \quad (4.11)$$

the individual feature cost  $\epsilon_{\nu}$  can be written in terms of the averages and variances  $\sigma_{a\nu}^2 = \frac{1}{n_a} \sum_{i \in a} x_{i\nu}^2 - m_{a\nu}^2$ :

$$\epsilon_{\nu} = (2 - (K - 1)\gamma) \sum_a \sigma_{a\nu}^2 - \gamma \sum_{a \neq b} (m_{a\nu} - m_{b\nu})^2, \quad (4.12)$$

which is the form we will use in the numerical applications.

For the case of two classes ( $K = 2$ ) the previous equation reduces to:

$$\epsilon_{\nu} = (2 - \gamma) \sum_a \sigma_{a\nu}^2 - \gamma \sum_{a \neq b} (m_{a\nu} - m_{b\nu})^2. \quad (4.13)$$

Then  $\gamma = 2$  is an acceptable value for classification purposes. If  $\gamma < \gamma^*$  the metric tensor can assume negative values. On the other hand, if  $\gamma > 2$  the term involving the intraclass variance becomes negative, which means that the increase of intraclass variance decreases the cost, hampering distinguishability among the classes.

For  $\gamma = 2$  we have:

$$\epsilon_{\nu} = -2 \sum_{a \neq b} (m_{a\nu} - m_{b\nu})^2, \quad (4.14)$$

so that the metric component for feature  $\nu$  is proportional to the difference of the averages between the conditions. For the case of two classes, a suitable normalization leads to the following relation:

$$\epsilon_\nu = -2t_\nu^2, \quad (4.15)$$

where  $t_\nu$  is the Student's t-variable. Then, up to an overall scale each component of the relevance vector  $\mathbf{w}$  is proportional to the square of the t-variable and consequently to the p-value of a t-test, given by,

$$\omega_\nu = \frac{t_\nu^2}{\sqrt{\sum_{\nu'} t_{\nu'}^4}}. \quad (4.16)$$

We now show results for the diagonal case and address in appendix ?? how to treat off-diagonal cases.

## 4.2 Results

An important application of this method is to use it in conjunction with a classification procedure, by preprocessing the training and validation datasets. A measure of its usefulness comes from comparing the performances of a classifier with and without the preprocessing.

For classification into two classes, where  $\chi = a$  or  $b$  is the prediction and  $\tau = a$  or  $b$  is the true class, a measure of the prediction performance is given by Matthews correlation coefficient, which we write in a convenient form as

$$MCC = \sqrt{\frac{n_a^\tau n_b^\tau}{n_a^\chi n_b^\chi} \det \mathbf{P}} \quad (4.17)$$

where the matrix  $\mathbf{P}$  has elements  $P_{\chi\tau} = P(\chi|\tau)$ ; and  $n_a^\tau$  and  $n_a^\chi$  are respectively the correct and the predicted number of vectors in class  $a$ . This is a measure that takes into account true and false positives/negatives in a balanced way [140, 141]. When MCC is equal to one it means that the method classified correctly all samples, conversely, if MCC is equal to minus one it means that classification was completely incorrect. The main reason for using this figure of merit is that our results can be compared with the MicroArray Quality Control (MAQC) consortium. We studied the performance of several variations of k-Nearest Neighbor (KNN) and Support Vector Machines (SVM). For the datasets where there are

more than two classes, leukemia (3 classes) and cow diet (4 classes), we used a generalization of the MCC discussed in [141] and compared our methodology with feature selection using ANOVA. All raw microarray data were preprocessed using the RMA method [142].

## MAQC consortium datasets

We first applied our methodology to two microarray datasets obtained from the MicroArray Quality Control (MAQC) consortium [129]. Each dataset consists of a training and validation set generated by different experimental groups. The Breast Cancer dataset can be divided according to two endpoints: pre-operative treatment response (pCR) and estrogen receptor (ER). The Multiple Myeloma dataset can be divided according to overall survival milestone outcome (OS-MO), event free survival milestone outcome (EFS-MO) and a “negative” endpoint where the classes are chosen randomly (CPR1). For more details about the class distribution see table 4.1. For these cases of a two-class classification, we choose  $\gamma = 2$  and a suitable normalization so that the metric tensor is proportional to the Student t-variable.

We compared our approach with different feature selection methods based on the t-test, which are: the top 10 and top 100 genes ranked according to its p-value, and using the leave one out (LOO) procedure in the training dataset in order to select the best number of genes. The case where all genes are considered to have the same relevance is also compared. In the majority of the cases, the best performance was obtained using the Suvrel preprocessing, see tables 4.3, 4.4, 4.5 and 4.6. We also compared our results with 17 MAQC data analysis teams that nominated their best models for microarray datasets [129]. In the majority of the cases our approach is superior to the median of the teams performance and frequently is higher than the best MAQC candidate. These results are summarized in figures 4.3 and 4.4 which also present a comparison of the prediction performance for the case without preprocessing and with Suvrel, and a comparison of Suvrel and other algorithms, respectively.

We also used the negative control generated by MAQC for the multiple myeloma dataset. In this case, the classes are assigned randomly and as expected all methodologies have a MCC close to zero, see table 4.7.

## Leukemia dataset

We also applied our methodology to a leukemia dataset that consists of 24 experiments with patients with Acute Lymphoblastic Leukemia (ALL), 28 experiments with patients with Acute Myelogenous Leukemia (AML) and 20 experiments with patients with Mixed-Lineage Leukemia (MLL) [143]. The dataset was divided in training and validation sets according to the methodology described in the original manuscript [143]. The data was normalized by subtracting the average and dividing by the standard deviation of each gene, zero mean and unit variance normalization.

Using the training dataset, we applied our methodology in order to determine the metric vector. Then, the distribution of the points for the validation dataset was represented in two dimensions using Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) [144], figure 4.1. Visual inspection shows a much better result, that can be quantified by noticing that for the first two PCA eigenvectors the explained variance ratio goes from [0.346, 0.078] without preprocessing to [0.493, 0.096] with Suvrel. In order to quantify the change in prediction performance, we compared our approach with a feature selection procedure based on ANOVA, using genes with a p-value lower than 0.05. Suvrel presents a superior performance for all classification methods, see figures 4.3, 4.4 and table 4.8.

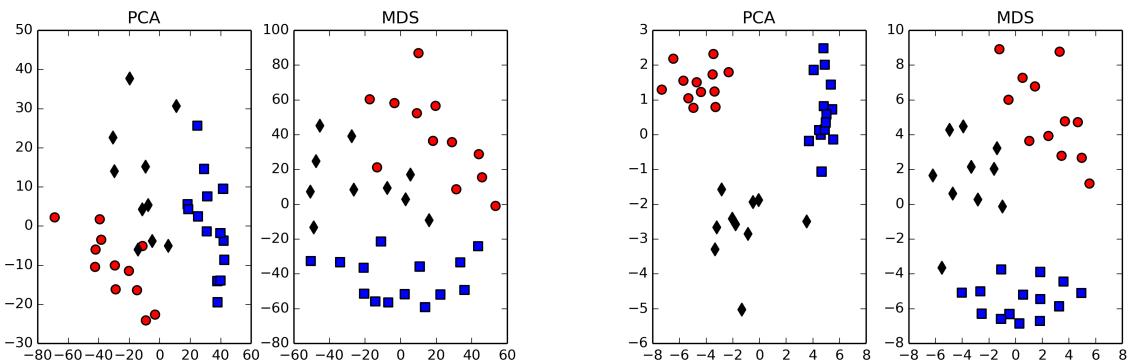


Figure 4.1: *Two dimensional representation of leukemia dataset using Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) (Left) Same relevance assigned to all genes. (Right) Relevance assigned by Suvrel. The symbols for the ALL class are red circles, for AML are blue squares and for MLL are black diamonds.*

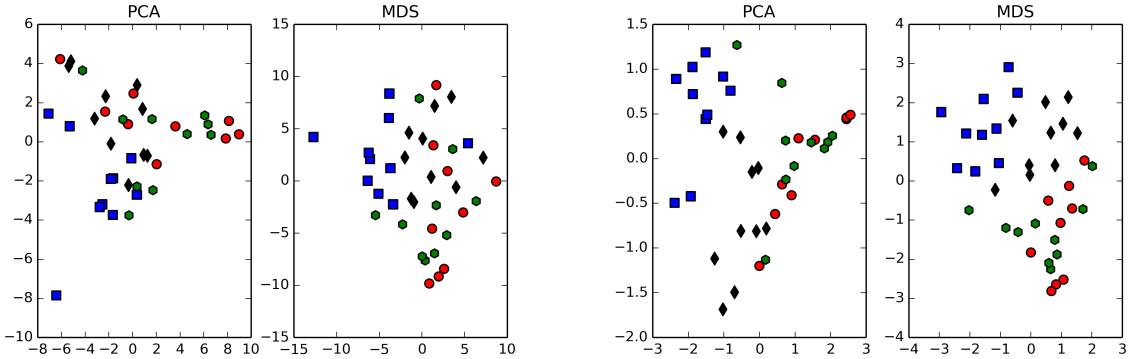


Figure 4.2: *Two dimensional representation of cow diet metabolites dataset using Principal Component Analysis (PCA) and Multidimensional Scaling (MDS). (Left) Same relevance assigned to all genes. (Right) Relevance assigned by Suvrel. (red circles) 0 %, (green hexagons) 15 %, (black diamonds) 30 %, (blue squares) 45 % of grain in cow diet.*

## Cow diet metabolites dataset

The last set consists of data from a metabolite experiment where the concentrations of 39 rumen samples was measured by proton NMR from dairy cows fed with different proportions of barley grain. The groups are divided in 4 classes: 0, 15, 30, or 45% indicating the percentage of grain in the diet [145].

Again using PCA and MDS the dataset is represented in two dimensions for both cases with no feature selection and Suvrel (Figure 4.2). A better class distinction is found in the new metric, mainly for the experiments with 30 and 45 percent of grain in diet. For the first two PCA eigenvectors the explained variance ratio goes from [0.358, 0.134] without preprocessing to [0.469, 0.132] with Suvrel. We also observed a good improvement in class prediction, table 4.2. In this case, since there is a small number of experiments, we used the leave one out methodology to determine the accuracy of the methods. See the comparison in figure 4.3 and figure 4.4.

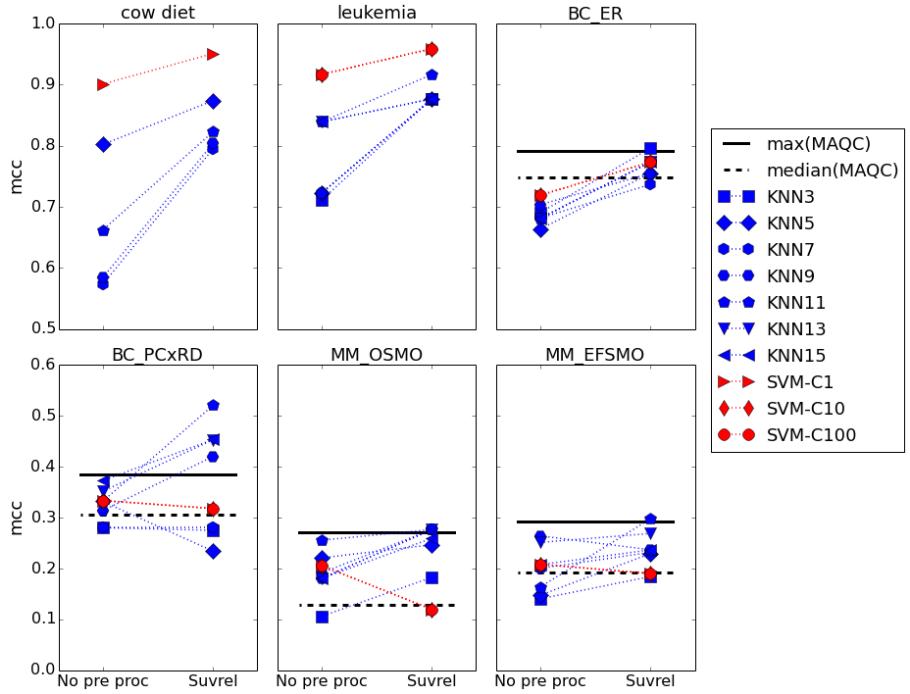


Figure 4.3: Each panel shows the results for each of the datasets of the prediction accuracy for (left) no data preprocessing, e.g. same relevance assigned to all genes, and (right) the case where the data is preprocessed by Suvrel. The horizontal lines indicate the median and maximum performance of MAQC teams.

### 4.3 Discussion and Conclusions

We have introduced a new analytic method for feature selection that is based on a variational method applied to cost functions that include intra and interclass information and that can be applied to any number of classes. The variational procedure is simple enough that an analytical solution is available. A parameter  $\gamma$  weights the relative importance given to the cost associated to intraclass and interclass distances. We have shown that the metric tensor is positive definite in a region around  $\gamma = 2/(K - 1)$  where  $K$  represents the number of classes. Then, in the case of two-class classification ( $K = 2$ ), the diagonal of the metric tensor, which can be interpreted as measuring feature relevance, is proportional to the square of Student's t-variable. Ranking the features according to a statistical test and then selecting a restricted number of features is a standard approach of feature selection. However, choosing the number

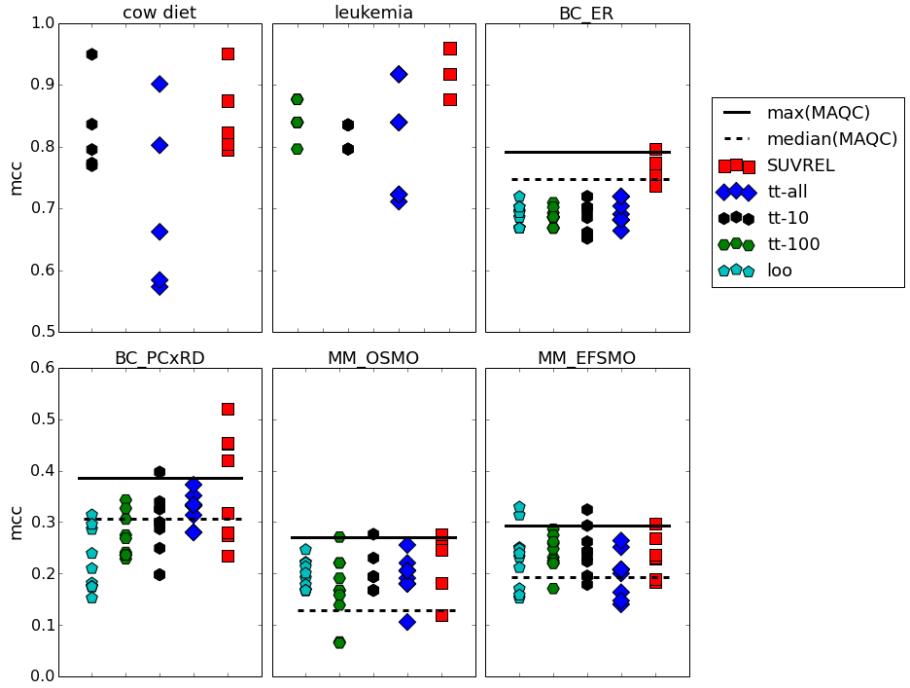


Figure 4.4: *Each panel shows the results for each of the datasets of the prediction accuracy for different preprocessing methods using different versions of KNN and SVM for classification. The horizontal lines indicate the median and maximum performance of MAQC teams.*

of predictive features is a question that has to be carefully addressed and it is usually done by maximizing the classification performance on the training dataset, with respect to the number of features. This procedure is very reasonable if the training and validation datasets are consistent, in the sense that interchanging learning and validations sets is of no great consequence. This is not the case of microarrays where the lack of robustness between the predictive lists of different datasets is a well known issue [146, 147]. In this case, the prediction model becomes very sensitive to the number of features chosen, as highlighted by the MAQC-II consortium [129]. Our method does not impose selection of a fixed number of features, which might explain why it showed a good improvement in the accuracy when compared to the standard methodology despite being based on the same ranking measure, which is the Student's t-variable. Combining Suvrel with standard classifiers (KNN and SVM) turns out to be, often, superior to the well elaborated protocols used by the MAQC-II teams (figure 4.3), therefore suggesting that this is a very useful methodology for feature

section.

Over the past years, metabolomics analysis has become an important tool for understanding biological processes of interest. However, there is a lack of methodologies for selecting the lists of relevant metabolites. Usually, the data points are represented in a lower space by PCA and if this representation shows a good separation between the classes, the metabolites that represent the dimensions that capture the higher variance are considered relevant. Furthermore, we point out that this might not be the best procedure and to chose the dimensions that contribute the most to the clusterization of the groups should be a better choice. Then, Suvrel can be a useful tool for selecting the relevant metabolites.

Due to the high dimensional nature of many bioinformatics data, the use of feature selection techniques has become an indispensable tool for biological model building. We point out that, unlike the majority of metric learning algorithms - which require a numerical minimization algorithm - Suvrel offers an analytical solution to find the relevant set of features. This provides an automatic and computational cheap method that can be useful for large datasets. Furthermore, the underlining assumptions of the method - presented in the form of a cost function - are straightforward to understand and the resulting analytical solution of the metric vector can be interpreted in terms of differences in averages and variances and even, for a particular value of  $\gamma$ , in terms of the well known t-Student variable. Among the extensions to the method we used in the numerical part of this paper we mention two. First we can consider the interactions among the features, which is done by taking the full non-diagonal metric tensor into account. This seems to be an immediate logical extension, since correlation of features is an intrinsic aspect of many biological systems. For instance, in gene expression often the determination of a given class is associated with a poll of features. In this case, it is expected an occurrence of significant (non-diagonal) cross-correlation terms. Second, a very simple online method can be used to update the full metric tensor, to accommodate the classification of new patterns.

### 4.3.1 Tables

Table 4.1: *Class distribution of each endpoint for the Breast Cancer and Multiple Myeloma dataset. Both datasets were generated by the MAQC-II consortium [129]. Columns are : Name of Data set, NST: Number of samples in the training set, PT: number of positive labels in training set, NT: number of negative labels in training set, NSV: number of samples in validation set, PV: number of positive labels in validation set, NV: number of negative labels in validation set*

Data set	Training set			Validation set		
	NST	PT	NT	NSV	PV	NV
Breast Cancer (pCR)	130	33	97	100	15	85
Breast Cancer (ER)	130	80	50	100	61	39
Multiple Myeloma (OS-MO)	340	51	289	214	27	187
Multiple Myeloma (EFS-MO)	340	84	256	214	34	180
Multiple Myeloma (CPR1)	340	200	140	214	122	92

Table 4.2: *Results for the cow diet metabolites dataset. Performance measured by the MCC for classification using several KNN and SVM classifier with data preprocessed using feature selection methods: (i) Suvrel, (ii) no selection, all features are used, (iii) feature selection by Anova. Our methodology shows a superior performance for all combination of prediction methods. The best performance is highlighted in bold.*

Method	Suvrel	no selection	Anova (p <0.05)
SVM	<b>0.951</b>	0.901	0.949
KNN3	0.874	0.802	0.836
KNN5	0.795	0.573	0.769
KNN7	0.804	0.584	0.773
KNN9	0.823	0.662	0.795

Table 4.3: *Results for the Breast Cancer dataset, divided according to estrogen receptor (ER) status. Performance as measured by the MCC for classification using several KNN and SVM classifiers with data preprocessed using feature selection methods: (i) Suvrel, (ii) all features, (iii) top 10 (iv) top 100 features selected by the t-test and (v) the top features chosen using the leave one out procedure (MCC, # features). MAQC teams best performance MCC = 0.792 and median 0.748. Performances superior to MAQC teams are highlighted in bold. Notation: KNN3 = KNN with 3 neighbors. SVM-C1 = SVM with C = 1*

	Suvrel	all	top-10	top-100	LOO
KNN11	0.7735	0.6812	0.6955	0.6681	(0.695, 10)
KNN13	0.7735	0.6812	0.7191	0.6681	(0.719, 20)
KNN15	0.7735	0.6812	0.7191	0.6681	(0.685, 50)
KNN3	<b>0.7964</b>	0.6904	0.6517	0.6927	(0.702, 20)
KNN5	0.7550	0.6637	0.7018	0.7091	(0.693, 150)
KNN7	0.7369	0.6817	0.6607	0.6848	(0.702, 350)
KNN9	0.7550	0.7036	0.6607	0.7018	(0.668, 130)
SVM-C1	0.7735	0.7191	0.6517	0.6861	(0.668, 280)
SVM-C10	0.7735	0.7191	0.6848	0.6861	(0.668, 280)
SVM-C100	0.7735	0.7191	0.7018	0.6861	(0.668, 280)

Table 4.4: *Breast Cancer* dataset divided according to pre-operative treatment response (*pCR*). Columns are the same as in table 4.3. MAQC teams best performance  $MCC = 0.385$  and median 0.306. Performances superior to MAQC teams are highlighted in bold.

	Suvrel	all	tt-10	tt-100	LOO
KNN11	<b>0.5206</b>	0.3318	0.2994	0.3430	(0.3140, 600)
KNN13	<b>0.4530</b>	0.3514	0.3256	0.2287	(0.2392, 150)
KNN15	<b>0.4533</b>	0.3726	0.2994	0.3058	(0.2858, 150)
KNN3	0.2750	0.2801	0.2492	0.2412	(0.1817, 200)
KNN5	0.2341	0.3318	0.3256	0.2750	(0.1528, 670)
KNN7	0.2801	0.2801	0.3396	0.2684	(0.2100, 550)
KNN9	<b>0.4201</b>	0.3135	<b>0.3972</b>	0.3267	(0.2957, 630)
SVM-C1	0.3178	0.3335	0.2870	0.2362	(0.1746, 420)
SVM-C10	0.3178	0.3335	0.1979	0.2362	(0.1746, 420)
SVM-C100	0.3178	0.3335	0.1979	0.2362	(0.1746, 420)

Table 4.5: *Multiple Myeloma* dataset divided according to overall survival milestone outcome (OS-MO). Columns are the same as in table 4.3. MAQC teams best performance  $MCC = 0.271$  and median 0.129. Performances superior to MAQC teams are highlighted in bold.

	Suvrel	tt-all	tt-10	tt-100	loo
KNN11	<b>0.2765</b>	0.2556	0.1941	0.2207	(0.2219, 380)
KNN13	<b>0.2765</b>	0.1803	0.1941	0.2207	(0.2466, 30)
KNN15	0.2593	0.1803	0.1941	0.1912	(0.2207, 60)
KNN3	0.1826	0.1060	0.2302	0.1390	(0.1665, 20)
KNN5	0.2466	0.2207	<b>0.2765</b>	0.2710	(0.2132, 30)
KNN7	<b>0.2765</b>	0.1912	0.1941	0.1675	(0.1912, 120)
KNN9	<b>0.2765</b>	0.1803	0.1941	0.1675	(0.2008, 30)
SVM-C1	0.1197	0.2058	0.1675	0.1585	(0.1783, 60)
SVM-C10	0.1197	0.2058	0.1675	0.0675	(0.1675, 10)
SVM-C100	0.1197	0.2058	0.1675	0.0653	(0.1675, 10)

Table 4.6: *Multiple Myeloma* dataset divided according to event free survival milestone outcome (EFS-MO). Columns are the same as in table 4.3. MAQC teams best performance  $MCC = 0.293$  and median 0.193. Performances superior to MAQC teams are highlighted in bold.

	Suvrel	all	tt-10	tt-100	loo
KNN11	<b>0.2978</b>	0.1637	0.2434	0.2206	(0.2317, 170)
KNN13	0.2688	0.2513	0.2239	0.2863	( <b>0.3137</b> , 130)
KNN15	0.2367	0.2075	0.1789	0.2317	(0.2505, 150)
KNN3	0.1843	0.1404	0.2933	0.2584	(0.2479, 160)
KNN5	0.2295	0.1475	<b>0.3241</b>	0.2746	(0.2392, 60)
KNN7	0.2317	0.2004	0.2619	0.2483	( <b>0.3293</b> , 400)
KNN9	0.2367	0.2641	0.2434	0.2619	(0.2118, 790)
SVM-C1	0.1901	0.2082	0.1952	0.1710	(0.1523, 50)
SVM-C10	0.1901	0.2082	0.2312	0.2250	(0.1708, 50)
SVM-C100	0.1901	0.2082	0.1952	0.2198	(0.1584, 50)

Table 4.7: *Multiple Myeloma* dataset divided according to a random outcome (CPR1). Columns are the same as in table 4.3. In this case, since the classes are chosen randomly it is expected a  $MCC$  close to zero.

	Suvrel	tt-all	tt-10	tt-100	loo
KNN11	0.0897	-0.0161	0.0716	0.0775	(0.0891, 50)
KNN13	0.0077	0.0302	0.0467	0.1166	(0.0758, 40)
KNN15	0.0677	0.0876	0.0276	0.0549	(0.0677, 40)
KNN3	-0.0068	-0.1008	0.0677	-0.0449	(0.0220, 130)
KNN5	-0.0545	-0.0765	0.0667	0.0020	(-0.0262, 70)
KNN7	-0.0424	-0.0575	0.0005	0.0188	(0.0465, 140)
KNN9	-0.0484	-0.0136	0.0194	0.0610	(0.1260, 130)
SVM-C1	-0.1138	-0.1015	-0.0025	-0.0065	(-0.0138, 30)
SVM-C10	-0.1138	-0.1015	0.0083	-0.0373	(-0.0141, 30)
SVM-C100	-0.1138	-0.1015	0.0110	-0.1228	(-0.0059, 30)

Table 4.8: *Performance, measured by MCC, for the leukemia dataset. The best result is highlighted in bold.*

	Suvrel	all	top10	top100
KNN3	0.9168	0.8390	0.8349	0.7958
KNN5	0.8769	0.8390	0.8349	0.8390
KNN7	0.8769	0.8390	0.8349	0.8390
KNN9	0.8769	0.7108	0.8349	0.8390
KNN11	0.8769	0.7224	0.7958	0.8390
KNN13	0.8769	0.7224	0.7958	0.8390
KNN15	0.8769	0.7224	0.7958	0.8390
SVM-C1	<b>0.9589</b>	0.9168	0.8349	0.8761
SVM-C10	<b>0.9589</b>	0.9168	0.8349	0.8761
SVM-C100	<b>0.9589</b>	0.9168	0.8349	0.8761

# **Chapter 5**

## **Relationship between global structural parameters and Enzyme Commission hierarchy: implications for function prediction**

In protein databases there is a substantial number of proteins structurally determined but without function annotation. Understanding the relationship between function and structure can be useful to predict function on a large scale. In this chapter, we analyzed the similarities in global physicochemical parameters for a set of enzymes which were classified according to the four Enzyme Commission (EC) hierarchical levels. We define a distance between proteins in the space of physicochemical characteristics by using our method (Suvrel) to minimize a cost function of the metric tensor built to reflect the EC classification system. Using an unsupervised clustering method on a set of 1025 enzymes, we obtained no relevant clustering formation compatible with EC classification. The distance distributions between enzymes from the same EC group and from different EC groups were compared by histograms. Such analysis was also performed using sequence alignment similarity as a distance. Our results suggest that global structure parameters are not sufficient to segregate enzymes according to EC hierarchy. This indicates that features essential for function are rather local than global. Consequently, methods for predicting function based on global attributes should not

obtain high accuracy in main EC classes prediction without relying on similarities between enzymes from training and validation datasets. Furthermore, these results are consistent with a substantial number of studies suggesting that function evolves fundamentally by recruitment, *i.e.*, a same protein motif or fold can be used to perform different enzymatic functions and a few number of specific amino acids are actually responsible for enzyme activity. These essential amino acids should belong to active sites and an effective method for predicting function should be able to recognize them.

## 5.1 Introduction

The number of proteins structurally determined but without function annotation has grown over the past years. Experimental function characterization is not only time-consuming but also expensive and therefore it is impractical on a large scale. The most commonly used methods to predict function rely on the transfer of annotation from a characterized homologue using sequential or structural alignment. This approach has a number of known limitations; for example, many proteins have no similar sequence or structure counterpart in databases. In addition, methods based in similarities or homology have been considered one of the main sources of error in function annotation, since the divergence of function can generate proteins structurally or sequence related but functionally different.

Instead of traditional alignment techniques, several algorithms that rely on an alternative representation of proteins have been proposed. Some approaches are based on global parameters like physicochemical properties or amino acid (AA) composition [148, 149, 150]. Cai and Chou [151] used a set of features based on AA composition to predict main enzyme class, which follows the Enzyme Classification (EC) hierarchy [152]. They used a Nearest Neighbor predictor and obtained 85% of accuracy including identification of non-enzymes. A Nearest Neighbor approach relies on the function based on nearest enzyme in a predetermined dataset. This methodology is based on similarity between pairs of enzymes and therefore has limitations similar to homology approaches.

In order to avoid these kinds of limitations, other studies assume that proteins performing similar functions must share similar structural characteristics - so called structure-based methods. In this context, Dobson and Doig [153] used a set of global attributes that can be

calculated from structure such as AA composition, secondary structure and surface properties. They selected subsets of different global attributes, and obtained 35% of accuracy in main enzyme class prediction using Support Vector Machines. However, the subsets of relevant attributes depend on which pair of EC classes are being compared. A predictive scheme based on global parameters can not in principle depend on the result that is expected. Borro *et al*[154], in a similar study, obtained 45% of accuracy using a Bayesian Classifier. In a systematic study, Bray et al described many structural features that showed significant differences between EC classes. These attributes were used to predict the main EC classes and obtained 33.1% accuracy [155]. These approaches are based on the underlying assumption that structural properties are intrinsically correlated with function. However, this assumption has not been convincingly demonstrated because the correlations were found using biased selected features.

A better performance in predicting function can be obtained taking into account a wide range of biological data [156]. However, here we focus on the relationship between global structural parameters and the protein function according to EC hierarchy. Our goal is to assess whether previously reported methods [153, 155, 154] have advanced in the direction of establishing methods of prediction of function from structural attributes. The enzyme features analyzed include secondary-structure content, AA frequency, surface area and volume. Using a supervised learning method (Suvrel) [4], we defined a distance between pairs of enzymes and studied the geometry of different groups of enzymes. Then, the distances between enzymes from the same EC group and from different EC groups are compared using histograms. In the next section, we discuss how the database was set and how the distance in the space of characteristics between enzymes is constructed. In the results section, we discuss the distributions of these distances and the clustering structures associated to them. These are also compared with results obtained using the distances based on sequence alignment.

## 5.2 Methods

The enzyme data set was chosen based on two criteria: no more than two enzymes share the same four EC numbers (to avoid function redundancy), and have enzyme structure resolution below 2 Å(good structure resolution). According to these criteria, 1025 enzymes were chosen

Table 5.1: Enzyme data sets according to the main EC class.

EC class	(# of elements)
oxidoreductase	226
transferase	264
hydrolase	300
lyase	118
isomerase	67
ligase	50
total	1025

in SCOP database [157], see Table 1. Each protein was represented by 9 physicochemical parameters and AA composition. The STING database [158] was used to calculate these parameters:

- alpha helix frequency (Alpha),
- beta sheets frequency (Beta),
- area exposed to solvent (Area),
- volume (Vol),
- amino acid frequency on surface (% AA<sub>surf</sub>),
- area exposed to solvent by volume (Area/Vol),
- charges on surface (Q<sub>surf</sub>),
- hydrophobicity (Hyd),
- mean energy of interaction between amino acids (IntEn),
- amino acid frequency.

These parameters are normalized to zero mean and unit variance in order to make them comparable.

### 5.2.1 Parameter selection

Once an enzyme is characterized by a set of parameters, it can be represented by a point in the  $N$  dimensional space of characteristics and the problem can be approached in a geometric fashion. The underlying geometry, still to be defined, will depend on the metric structure we might choose to impose. This freedom should be used to reflect any additional information at our disposal. Not all of the characteristics we have considered have the same relevance in predicting function and this has to be investigated, because some parameters may even be just noise and compromise the analysis. Thus, we used our proposed method Suvrel [4] (see previous chapter) to determine the metric tensor, which is inspired in relevance learning [159] and permits joining the advantage of supervised and unsupervised learning methods. The idea behind the metric's dynamics is simple and it reflects that enzymes we know to be similar in function, from any additional piece of information, should end up at a close distance, while enzymes known to be different in function should end up far apart.

Each parameter  $k$  has a relevance weight factor  $w_k$ , the component of the diagonal metric tensor, still to be determined, and the distance between two enzymes  $i$  and  $j$  is defined as:

$$D_{ij}^2 = \sum_k w_k (x_{ik} - x_{jk})^2, \quad (5.1)$$

where  $x_{ik}$  is the  $k$  parameter's value of the enzyme  $i$ . This methodology permits, by pruning those that become irrelevant, a dimensional reduction of the original space of characteristics.

Consider a selected set of pairs of enzymes divided into two groups. The first group  $C$  is formed by  $n_c$  pairs of enzymes that should be close and the second set  $F$  is formed by  $n_f$  pairs of enzymes that we know should be far from each other. For the *close* set we consider pairs that share the first EC number and for the far apart set, pairs that do not share the first EC. Thus, a suitable cost function is

$$E = \frac{1}{n_c} \sum_{ij \in C} D_{ij}^2 - \frac{1}{n_f} \sum_{ij \in F} D_{ij}^2, \quad (5.2)$$

which decreases when the metric is such that pairs of enzymes in  $C$  are indeed close and pairs in  $F$  are far. We have used the enzyme data set described in Table 1 and the vector  $w$  was obtained by using the Suvrel methodology.

Once the weighting vector  $w$  is obtained, we can define a distance between the enzymes using eq. 5.1 and apply several methods to study the geometry of the cloud of enzymes. In

the following section, we present results of two types. We compared the intra-class and extra-classes distance distributions among the EC levels using histograms and we also analyzed the cluster structure of the data using the standard method Super Paramagnetic Clustering (SPC) [160]. This is a hierarchical clustering process, inspired in Statistical Mechanics which relies on collective properties of the set, to attribute clusters. We consider this method to be an interesting technical choice since it does not require the imposition of a determined *a priori* number of classes or clusters. In this method, a parameter called the temperature, due to the origin of the method, permits probing the geometric structure of the set at different scales. At low temperatures, all patterns belong to the same cluster, and as temperature increases, a series of phase transitions might occur whenever a class of patterns separates into one or more subclasses. Furthermore, it can be described as a Bayesian algorithm, where the hypothesis under which the clustering occurs is clear and manifest.

### 5.3 Results and Discussion

A list of ranked attributes and weighting factors  $w$  obtained by Suvrel is presented in Table 2. Given the vector  $w$ , the distance between each pair of 1025 enzymes in the dataset was computed using Eq. 5.1. Then, we analyze the cluster structure of the enzyme set using the SPC method, described in the previous section. We have found no relevant formation of clusters compatible with EC function. The interesting feature we found is that under increase of the probing scale, the set pulverized into a structureless cloud, without going through a sequence of paramagnetic phase transitions that would have indicated a hierarchical structure in the space of physicochemical characteristics.

These results are better understood when they are presented as histograms of enzyme distance distributions. We compared the distribution of pairs of enzymes belonging to the same EC level and of pairs of enzymes belonging to different EC levels; they are plotted in Fig. 5.1. Both curves are roughly identical for the first three hierarchical levels (Fig. 5.1A-5.1C). This indicates that there is no manifest relationship between global attributes and EC function. In contrast, enzymes that share the four EC numbers, *i.e.*, performing the same specific reaction, have very similar global parameters (Fig. 5.1D). Similar results are found using sequence alignment similarity (Fig. 5.2) where the proportion of AAs at

Table 5.2: Rank of parameters obtained by relevance theory analysis.

rank	parameter	$w_k$	rank	parameter	$w_k$
1	LEU	0.1290	16	GLN	0.0262
2	alpha	0.0976	17	PRO	0.0218
3	GLU	0.0818	18	ASN	0.0173
4	area	0.0766	19	HIS	0.0089
5	% AA <sub>surf</sub>	0.0743	20	EnInt	0.0084
6	vol	0.0688	21	SER	0.0040
7	ALA	0.0540	22	GLY	0.0040
8	CYS	0.0501	23	beta	0.0000
9	ILE	0.0468	24	area/vol	0.0000
10	MET	0.0434	25	hyd	0.0000
11	ARG	0.0430	26	PHE	0.0000
12	LIS	0.0395	27	THR	0.0000
13	Q <sub>surf</sub>	0.0385	28	TRP	0.0000
14	VAL	0.0374	29	TYR	0.0000
15	ASP	0.0286	-	-	-

which the two sequences are different (according to BLOSUM50 scoring matrix) was used as sequence similarity distance. Comparing the results presented in Fig. 5.1 and 5.2, we can relate the high structural similarity seen in forth EC level (Fig. 5.1D) to the high sequence similarity between the enzymes (Fig. 5.2D), *i.e.*, the enzymes sharing the four EC number have similar global features because they are very similar in sequence. Thus, a method based on Nearest Neighbor and global structural parameters should obtain high accuracy if there is some redundancy of sequence between enzymes in the training and validation databases. Meanwhile, structure-based methods should not obtain high accuracy, since enzymes in the same first EC level do not share similar global features. This observation explains and, in the same token, is in agreement with the cited literature.

We must also keep in mind that the EC classification does not intended to reflect the structural features of the enzymes and do not take into account evolutionary events such as divergence or convergent evolution. These events are not rare and are reported in various

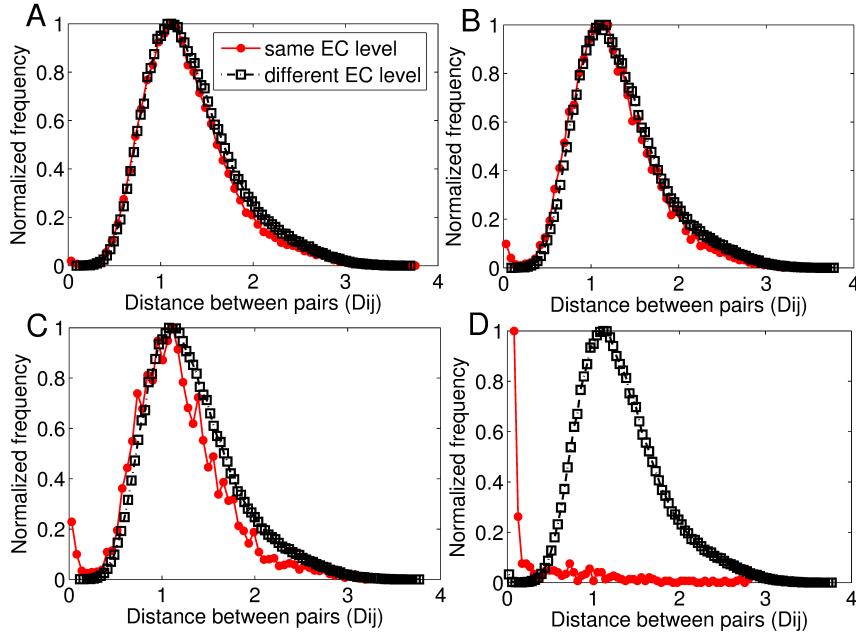


Figure 5.1: Normalized frequency histogram of the distance (in the space of characteristics) between pair of enzymes sharing the same EC level (circle) and between enzymes with different EC level (square). (A) first EC level (B) second level (C) third level (D) fourth level.

studies [161, 162, 163]. For instance, Martin *et al*[164] have concluded that function described by top-level EC numbers is not related to fold, since a few specific residues are responsible for enzyme function. If a few numbers of AA are essential for the function, it should not be distinguished by global physicochemical parameters. In the case of enzymes, these essentials AA should belong to active site, while in the case of non-enzymes, the functionally important regions are involved in ligand-binding or protein-protein interactions. Identifying these AA most likely to be involved in protein activity is a current challenging in protein analysis [165, 166, 167, 168, 169].

Furthermore, these results can be addressed from a protein folding perspective [170, 171]. It is accepted that the protein folding happens in an energy landscape which funnels to the native state, a state with a conformation consistent with the principle of minimal frustration [172, 173]. In order to fold and to display stability under physiological conditions, the interactions present in the native structure are not in conflict, as in a random heteropolymer,

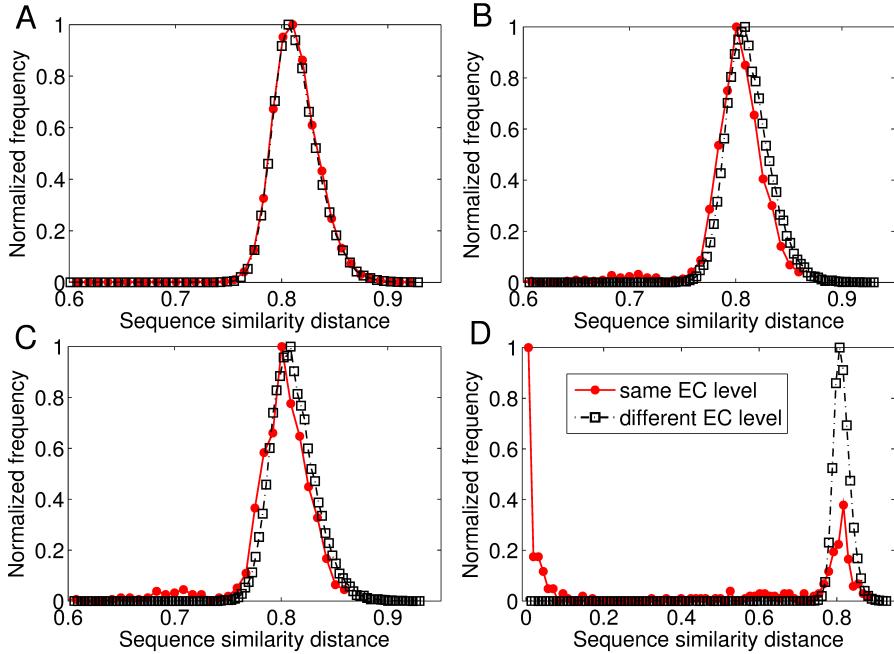


Figure 5.2: Normalized frequency histogram of sequence similarity distances between pairs of enzymes sharing the same EC level (circle) and between enzymes with different EC level (square). (A) first EC level (B) second level (C) third level (D) fourth level, analogous to Figure 2.

but instead are mutually supportive and cooperatively lead to a low-energy structure. These interactions are said “minimally frustrated” or “consistent” [170]. These ideas are valid for most of the AA in a protein, however it has been observed frustrated AAs that, in principle, seems to conflict with the robust folding of the protein. Such frustrated interactions seem to reflect evolutionary constraints other than folding and often correspond to physiologically relevant sites, such as active sites [174]. Thus, one can roughly divide the AAs in a given protein into two groups. In one group are AAs that ensure the protein fold and its stability, these are minimally frustrated AAs. The other group consists of a small group of AAs associated with the protein specific function. The amino acids involved in function are likely to be frustrated. The degree of frustration in AA can be measured [174] and can, in principle, be used to predict enzyme functions. Following this idea, it is also plausible to conceive the recruitment principle, in which a given protein fold is used for different functions. It is “cheaper” for nature to obtain a new enzyme by changing a few AAs than to modify a

protein fold from a given enzyme. Enzymes with different functions which present very similar structures are evidence of this conjecture [175].

## 5.4 Conclusion

In the present manuscript, we evaluated the relationship between global parameters and function defined by the EC hierarchy. We defined a distance between enzymes based on global features and compared the distance distributions among enzymes in the same EC level and in different EC levels. This analysis showed that global features can not distinguish the function defined by EC hierarchy.

Recently, various approaches [153, 155, 154] were developed based on global parameters to predict protein functions motivated by a specific hypothesis on a structure/function relationship. In opposition to these methods, we claim that these structural protein parameters are not sufficient for prediction. In our opinion, the set of attributes extracted from structure has to be the same for all pairs of proteins, independent of their yet to be determined class, as is the case of [153]. A method cannot depend on the answer (i.e. *this protein belongs to that class*) in order to choose the attributes that will be used to *prove* that assertion. The fact that accuracies of up to 60% have been reported is due to a case to case choice of what constitutes a good set of attributes. The statistical significance for prediction has not been established and our results indicate that this is unlikely to be achieved from structure alone.

While function prediction remains a useful and worthwhile task, we point out that completion of this program still needs new ideas. Our results point to the lack of a clear relationship between structure and function, and if one hopes to identify the function, one should focus on specific attributes from active sites or disordered residues [176, 177]. The later has been considered one of the most powerful features in determining the correct biological process for a given protein [178]. However, even though specific attributes prediction and enzyme function prediction are correlated, the later is probably more complicated. The reason is because the enzyme classification (EC code) has predetermined assumptions, which may not necessarily be the best classification framework.

# **Chapter 6**

## **T-test at the probe level: An alternative method to identify the statistically significant genes for microarray data**

Microarray data analysis typically consists in identifying a list of differentially expressed genes (DEG), i.e., the genes that are differentially expressed between two experimental conditions. Variance shrinkage methods have been considered a better choice than the standard t-test for selecting the DEG because they correct the dependence of the error with the expression level. This dependence is mainly caused by errors in background correction which affects more severely genes with low expression values. Here, we propose a new method for identifying the DEG that overcome this issue and does not require background correction or variance shrinkage. Unlike current methods, our methodology is easy to understand and to implement. It consists of applying the standard t-test directly on the normalized intensity data, which is possible because the probe intensity is linear dependent with the gene expression level and because the t-test is scale- and location-invariant. This methodology considerably improves the sensitivity and robustness of the list of DEG when compared to the t-test applied to preprocessed data and to the most widely used shrinkage methods, SAM and LIMMA. Our approach is useful especially when the genes of interest have small

differences in expression and therefore get ignored by standard variance shrinkage methods.

## 6.1 Introduction

Microarrays are a widely used methodology for measuring the expression of thousands of genes simultaneously. A common application of microarrays is to compare the expression levels of genes in samples drawn from two different experimental conditions in order to determine which genes are differentially expressed. Typically, a microarray data set has on the order of ten thousand genes, whereas only a small subset of these genes is relevant, and the number of samples range from a few tens up to a few hundreds. Multiple processing steps are required in order to identify the genes of interest and errors in these steps compromise the reliability of the analysis. As a result, the comparison of the lists of differentially expressed genes (DEG) reported by different groups have revealed very small overlap [147, 146]. Therefore, understanding the sources of this lack of robustness and the development of more reliable methods for the identification of the DEG remains a crucial issue.

The preprocessing of the raw probe intensity data constitutes the initial step in microarray data analysis and its goal is to infer a variable that represents the gene concentration. Usually, the preprocessing analysis is performed in three steps: normalization, background correction and summarization. First, the normalization is performed to reduce sources of variation of non-biological origin among the arrays, in order to make them comparable. Next, in background correction, the background intensity due to non specific hybridization and optical noise is inferred and subtracted from the normalized intensity. Finally, in the summarization step, the multiple probe intensities for each probe set is combined into a single gene expression value.

The probe intensity measure  $Y$  can be modeled as a combination of background  $B$  (due to optical noise and non-specific hybridization) and specific hybridization  $S$  [179, 142]:

$$Y_{ijg} = B_{jg} + S_{ijg} = B_{jg} + \phi_{jg}\theta_{ig}, \quad (6.1)$$

where  $\theta_{ig}$  denotes the expression measure for the gene  $g$  in the  $i$ th sample. This model assumes that the intensity value  $Y_{ijg}$  increases linearly as  $\theta_{ig}$  increases, but the rate of increase  $\phi_{jg}$  depends on the probe, represented by the index  $j$ . It is also assumed that after

normalization, the background  $B_{jg}$  is independent on the sample [180].

Background correction is the most critical step because errors associated with the inference of this variable strongly affect the genes with low expression values - therefore decreasing the statistical power of further analysis. In the Affymetrix platform it is common to include an extra probe - referred to as the mismatch - created by changing the middle (13th) base with the intention of directly measuring the effect of background noise. The methods dChip [179] as well as the Affymetrix methods MAS5.0 and PLIER [181] use the mismatch intensity as an estimate of the background. Therefore, Irizarry *et al.* [142] showed that subtracting the mismatch intensity from the perfect match intensity ( $Y$ ) results in expression estimates with an exaggerated error, mainly for low expression values. So, they proposed a background adjustment step that ignores the mismatch intensities, named RMA [142], which uses the posterior mean  $E[S|Y]$  as a background adjustment.

Once a gene expression estimation is obtained, it is common to take the logarithmic transformation of this variable because the difference in transformed data, the fold change, is considered easier to manipulate and interpret. Some preprocessing algorithms return the expression concentration in log scale (like RMA and Plier) while others do not (like MAS5.0). It seems that there is no consensus about what scale should be chosen when using t-test or others ranking methods. Nevertheless, this choice is important because the t-test, for example, is not invariant under monotone transformations, i.e., the results of the test is different if  $x$  is replaced by  $\log_2 x$ .

After preprocessing, several types of statistical tests can be applied in order to find the genes differentially expressed in two conditions. The independent t-test is the most popular statistical approach to select differentially expressed genes presumably due to its simplicity to implement and interpret. In this test, it is assumed that the data follows a normal distribution and, under the null hypothesis  $H_0$ , the average of the gene expression in both experimental conditions are the same. A discordance of the data from what is specified in  $H_0$  can be quantified as the probability of observing a value for the test statistic that is at least as extreme as the value that was actually observed. This probability  $p$  is referred to as p-value and a threshold value  $\alpha$  can be chosen so that the hypothesis  $H_0$  is rejected if  $p \leq \alpha$ . Then, a statistical significance can be assigned, which is a statistical assessment of whether observations reflect a pattern rather than just chance. The  $t$  variable, which follows

a Student's t distribution when the normality of the data holds, is defined as:

$$t_g = \frac{\langle \hat{\theta}_{gi} \rangle_{i \in A} - \langle \hat{\theta}_{gi} \rangle_{i \in B}}{s_g}, \quad (6.2)$$

where  $\hat{\theta}_{ig}$  is the inferred expression level by a preprocessing method and  $\langle \hat{\theta}_{gi} \rangle_{i \in A}$  and  $\langle \hat{\theta}_{gi} \rangle_{i \in B}$  are the expression averages over the samples  $i$  in the conditions  $A$  and  $B$ , respectively.

The empirical standard deviation is defined as

$$s_g = \sqrt{a \left( \sum_{i \in A} [\hat{\theta}_{gi} - \langle \hat{\theta}_{gi} \rangle_{i \in A}]^2 + \sum_{i \in B} [\hat{\theta}_{gi} - \langle \hat{\theta}_{gi} \rangle_{i \in B}]^2 \right)}, \quad (6.3)$$

where  $a = (1/n_A + 1/n_B)/(n_A + n_B - 2)$  and the constants  $n_A$  and  $n_B$  represent the number of samples under the experimental conditions  $A$  and  $B$ , respectively.

Despite being widely used, the t-test has been subject of criticism in the literature since the error in preprocessed data tend to be asymmetric [182, 183, 184, 185]. As a consequence, the variance estimation is dependent on the expression level because the genes with low expression levels are more affected by errors in background correction. Because of that, modified versions of the standard t-test have been developed as alternative approaches [186, 187, 188, 189, 190]. Those approaches modify the t-test by using a procedure called variance shrinkage which consists in modifying the denominator of the  $t$  variable by combining the gene-specific variance and a predictive variance.

The method Significance Analysis of Microarrays (SAM) [186] is the most popular alternative to t-test and it consists in a modification of the standard t-test by the inclusion of an extra variable  $s_0$ , added to the pooled variance  $s_g$ . The extra variable  $s_0$  is chosen in order to minimize the dependency of  $t$  on the expression level  $\hat{\theta}_g$  and the significant genes are identified comparing the modified  $t$  variable with a similar variable obtained under random permutations among the samples. Modifications of the t-test based on Empirical Bayes [191, 188, 189] have also been widely used and have been considered a good choice [184], and the method LIMMA [189] is the most widely used.

There is no shortage of more sophisticated alternatives to the t-test. However, given the widespread tendency to use the standard t-test, understanding the reason of its poor performance remains a crucial issue. Here, we show that the standard t-test can be used at the probe-level, skipping background correction by using the normalized intensity data

directly in the t-test. Several methods have approached this problem at the probe-level [192, 193, 194, 195], however, all those approaches use the background-corrected data. Our methodology outperforms the standard t-test using preprocessing data and the most used shrinkage methods SAM and LIMMA - therefore suggesting that background correction is a major source of error in microarray analysis. The methods were compared in sensitivity by using the Affymetrix spike-in data set, a commonly used benchmark, and in robustness by using leukemia, breast cancer and multiple myeloma data sets.

## 6.2 Methodology

The standard t-test is location and scale invariant, i.e., the results of the test do not change if  $x$  is replaced by  $ax + b$ , when  $a$  and  $b$  are constants. Because of that, according to the model described in equation 6.1, applying the t-test to the normalized intensity data is equivalent to applying the test on the concentration variable  $\theta$ , as illustrated in the following manipulations:

$$\begin{aligned}
t_{jg} &= \frac{\langle Y_{ijg} \rangle_{i \in A} - \langle Y_{ijg} \rangle_{i \in B}}{s(Y_{ijg})}, \\
&= \frac{\langle B_{jg} \rangle_{i \in A} + \phi_{jg} \langle \theta_{ig} \rangle_{i \in A} - \langle B_{jg} \rangle_{i \in B} - \phi_{jg} \langle \theta_{ig} \rangle_{i \in B}}{\phi_{jg} s(\theta_{ig})}, \\
&= \frac{\phi_{jg} [\langle \theta_{ig} \rangle_{i \in A} - \langle \theta_{ig} \rangle_{i \in B}]}{\phi_{jg} s(\theta_{ig})}, \\
&= \left[ \frac{\langle \theta_{ig} \rangle_{i \in A} - \langle \theta_{ig} \rangle_{i \in B}}{s(\theta_{ig})} \right]_j.
\end{aligned} \tag{6.4}$$

Note that no log transform was performed in the data and that for each gene  $g$  a set of  $t$  variables (one for each probe) is obtained. In order to select the differentially expressed genes, we propose to take the median t-values since the median is more robust than the average in the presence of outliers. Then, the genes can be ranked according to its relevance by the median t-values, which often is enough for selecting a subset of genes as biomarkers candidates.

To estimate the statistical significance of the median t-values of each gene, we define  $F(t)$  and  $f(t)$  to be respectively the cdf and pdf distribution of a t-variable. Then, for the case

of a number of probes equal to  $n$ , the pdf distribution of the median t-value  $g(t)$  is given by [196]:

$$g(t) = \begin{cases} C_n [F(t)]^{n/2} [1 - F(t)]^{n/2} f(t) & \text{if } n \text{ is even and } C_n = \frac{n!}{(n/2)!(n/2)!} \\ C_n [F(t)]^{(n-1)/2} [1 - F(t)]^{(n-1)/2} f(t) & \text{if } n \text{ is odd and } C_n = \frac{n!}{(n-1)/2!(n-1)/2!} \end{cases} \quad (6.5)$$

Now, for a given t-median value  $t_m$ , a p-value can be estimated by integrating the pdf  $g(t)$  function for  $t < t_m$ .

## 6.3 Results

### 6.3.1 Sensitivity

The sensitivity in identifying the genes differentially expressed was evaluated by using spike-in experiments. This data set was obtained from measurements on specifically constructed and controlled DNA microarrays experiments using human genome HG-U133. These experiments were designed by Affymetrix for the purpose of developing and validating the Affymetrix Microarray Suite (MAS) 5.0 expression algorithm [197]. The samples follow a Latin Square design consisting of 42 genes in 14 different concentrations (with 3 repetition each), see table 1 in the appendix.

To simulate an analysis in which two conditions are compared, we rearranged the data as follows: First, the genes from samples 1 to 3 were attributed to condition *A* and those from samples 4 to 6 were attributed to condition *B*. The exceptions were the genes from samples 40 to 42 because we want to probe the effect of small differences in expression, i.e., we only evaluate differences in the fold change equal to 2. We kept the procedure using the replications 4-6 as condition *A* and 7-9 as condition *B*, again excluding the 0 and 512 pM differences. We ended up with 14 “experiments” in the two conditions and 23,000 genes of which 39 are differentially expressed. This procedure is similar to the one done in Affycomp [198] to obtain a ROC curve with fold change equals to 2.

The results are presented as a Receiver Operating Characteristic (ROC) curve whose definition is given by a true positive (TP) rate against a false positive (FP) rate obtained at different threshold values. To plot a single average ROC curve, we calculated the average TP and FP over the 14 experiments obtained in the Affymetrix spike-in data rearrangement.

The proposed approach showed a significant improvement on the sensitivity in recognizing the differentially expressed genes when the t-test was used as the ranking method, figure 1A. Its performance is similar to the SAM and LIMMA best performance (when using RMA as a preprocessing algorithm), figure 1B. In all applied tests the data were normalized using the quantile normalization method [199].

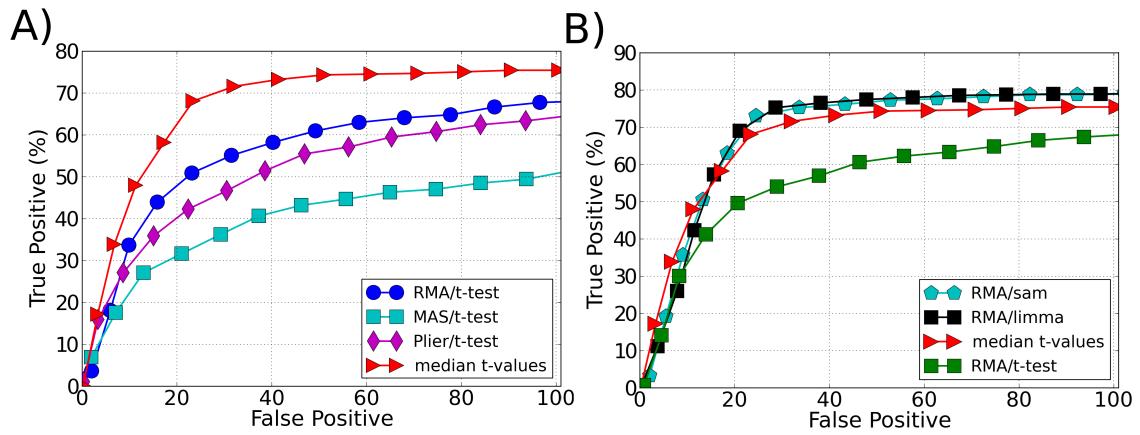


Figure 6.1: *ROC curve comparing the performance (sensitivity) of different ranking methodologies in identifying differently expressed genes in spiked-in data. A) Comparison of our approach (median t-values) with different preprocessing methods used with t-test as ranking method to select the differentially expressed genes. B) Comparison of our approach (median t-values) with the best performance of different ranking methods: t-test, SAM and LIMMA. The best performance of t-test, SAM and LIMMA is obtained when using RMA method as a preprocessing algorithm.*

### 6.3.2 Robustness

In addition to having a good sensitivity, a good method for selecting the differentially expressed genes (DEG) should be robust, i.e, the lists of DEG generated by different samples should share a good fraction of genes. The lack of agreement between those lists is a well known issue in microarray analysis, mainly in cancer studies [147, 146]. Furthermore, a good reliability of the selected DEG correlates positively with the class predictability [200]. In order to assess the robustness of a method, we used a natural similarity measure introduced by Ein-Dor *et al* [146] which is the fraction of genes shared by two lists of DEG obtained from different samples using a given method. More specifically, in order to estimate the robustness of the methods, we generate 100 training samples by taking a subset of  $n$  experiments chosen randomly. For each training sample we chose the  $N_{top} = 100$  most significant genes obtained by the given method. Then, we compare the fraction of shared genes ( $f_{a,b}$ ) between the training samples  $a$  and  $b$ . The average of the fraction of shared genes over all the combinations of two training samples ( $f = \langle f_{a,b} \rangle_{a \neq b}$ ) is the figure of merit we use to represent the robustness. Therefore, the closer  $f$  is to 1, the more robust is the method.

We calculated the average overlap for leukemia, breast cancer and multiple myeloma data sets for different sample sizes. The leukemia data set consists in 24 samples of acute lymphoblastic leukemia patients (ALL), 28 samples of leukemia acute myelogenous leukemia patients (AML) and 20 samples of mixed-lineage leukemia patients (MLL) [143]. We chose only the samples of leukemia type ALL and AML because these two types can be clearly distinguish based solely on gene expression profiles [201]. The Breast Cancer and Multiple Myeloma data sets were obtained from the MicroArray Quality Control (MAQC) consortium [200]. The Breast Cancer data set can be divided according to two endpoints, pre-operative treatment response (pCR) and estrogen receptor (ER). The Multiple Myeloma data set can be divided according to overall survival milestone outcome (OS-MO) and event free survival milestone outcome (EFS-MO), see table 2 in the appendix.

We compared the robustness of our methodology with different ranking methods: t-test, SAM and LIMMA, figures 2-4. The best performance of t-test, SAM and LIMMA was obtained when using RMA as a preprocessing algorithm. Our approach (median t-value) shows a significantly higher overlap for the Leukemia and Multiple Myeloma data sets, figures 2-3, respectively. In the case of the Breast Cancer data set, our approach shows

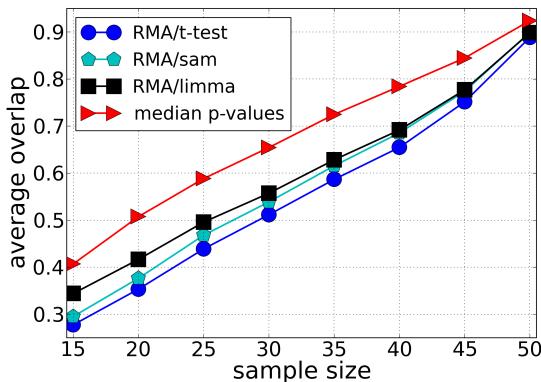


Figure 6.2: *Average of the fraction of genes shared by two list of differentially expressed genes (overlap) as a function of the sample size using the Leukemia data set. Each list of differentially expressed genes is composed by the top 100 genes chosen according to different ranking methods: t-test, SAM and LIMMA (preprocessed by the RMA method) and our approach (median t-value) which does not require a preprocessing algorithm. The average value of the overlap between the lists is calculated over 100 lists chosen randomly.*

a superior performance to the pre-operative treatment response (pCR) endpoint, but an inferior performance to the estrogen receptor (ER) endpoint, figure 4. These results suggest that part of the lack of robustness in microarrays analysis is due to errors incorporated in the preprocessing steps, therefore, explaining the significant gain of robustness of our methodology. However, we point out that for Breast Cancer and Multiple Mieloma, the levels of robustness for some analysis is still very low, suggesting a large biological heterogeneity among the samples.

## 6.4 Discussion

Over the years, a large number of preprocessing algorithms has been suggested. Many of them are based on underlying manipulations and assumptions that are difficult to understand. For example, the Plier method, suggested by Affymetrix, has been regarded as a good choice [202] despite being considered based on biologically implausible assumptions [203]. Among the steps required for preprocessing, background correction is probably the most important since

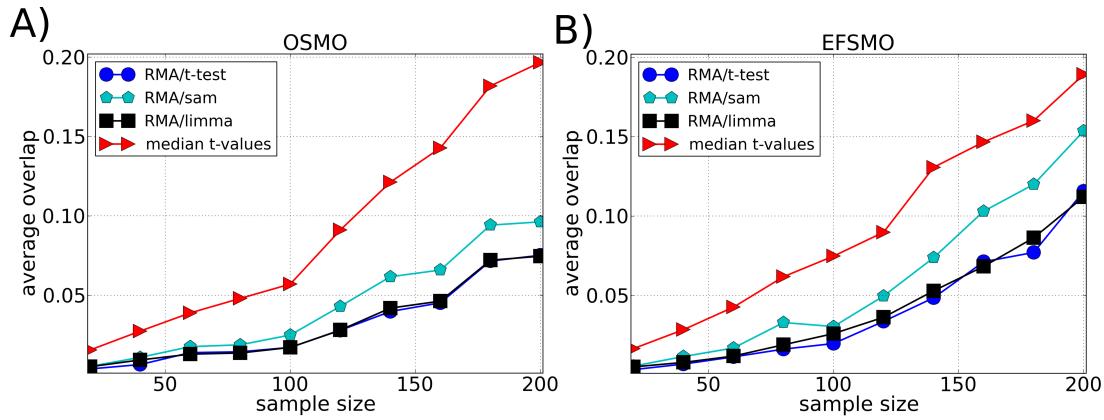


Figure 6.3: *Average of the fraction of genes shared by two list of differentially expressed genes (overlap) as a function of the sample size using the Multiple Myeloma data set divided according to A) Overall Survival Milestone Outcome (OS-MO) and B) Event Free Survival Milestone Outcome (EFS-MO). Each list of differentially expressed genes is composed by the top 100 genes chosen according to different ranking methods: t-test, SAM and LIMMA (preprocessed by the RMA method) and our approach (median t-value) which does not require a preprocessing algorithm. The average value of the overlap between the lists is calculated over 100 lists chosen randomly.*

errors due to this step affect more severely the genes with low expression values. Because of that, the standard deviation becomes dependent on the gene expression level and to overcome this issue, t-test modifications like SAM, LIMMA and others shrinkage methods, have been developed and considered better choices. In fact, these methodologies present a good improvement in the task of identifying the differentially expressed genes when compared with standard t-test, as we showed using spiked-in experiments. However, we highlight that by modifying the pooled variance, these methods tend to ignore the genes with low differences in expression and also, although improving the sensitivity when compared with standard t-test, these strategies does not show a significantly improvement in the task of selecting a robust predictive list.

Here, we introduce an alternative approach for statistical analysis of microarray data that skips background correction step, leading to a more powerful and robust test. Our procedure relies on a well established model that relates the probe intensity level with the

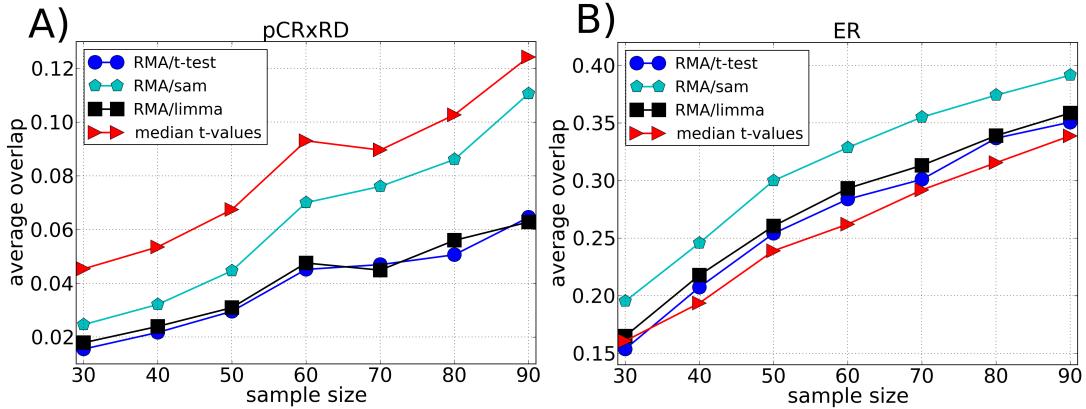


Figure 6.4: *Average of the fraction of genes shared by two list of differentially expressed genes (overlap) as a function of the sample size using the Breast Cancer data set divided according to A) pre-operative treatment response (pCR) and B) estrogen receptor (ER) endpoint. Each list of differentially expressed genes is composed by the top 100 genes chosen according to different ranking methods: t-test, SAM and LIMMA (preprocessed by the RMA method) and our approach (median t-value) which does not require a preprocessing algorithm. The average value of the overlap between the lists is calculated over 100 lists chosen randomly.*

gene expression level, and the standard t-test location- and scale-invariance property. Our method is easy to understand and to implement, however it does not offer an estimative of the expression level for each gene. We highlight that if the question under consideration is the identification of differentially expressed genes (DEG) or the predictive gene lists (PGLs), then intermediate estimation of expression levels is an unnecessary detour, and our method is useful. We also point out that our methodology is useful when the important genes are expected to have a small difference in expression. In this case, shrinkage methodologies are not recommended since they tend to ignore genes with small fold change.

# Appendix II

## II.A Positivity of the metric tensor

Let  $u, v$  be arbitrary unit length vectors in the space of features. Define  $Z^2(u) = \sum_a Z_a^2$  where  $Z_a = \sum_{\mu, i \in a} (x_{i\mu} - m_{a\mu}) u_\mu$  and  $W^2(u) = \sum_{a \neq b} W_{ab}^2$  where  $W_{ab} = \sum_\mu (m_{a\mu} - m_{b\mu}) u_\mu$ . Introduce

$$Q(u, v) = (2 - (K - 1)\gamma)Z^2(u) - \gamma W^2(v). \quad (6.6)$$

Note that for  $u = v$ ,  $Q(u, u) = \sum_{\mu\nu} \epsilon_{\mu\nu} u_\mu u_\nu$  is a quadratic form. The condition for positivity of the metric tensor is that  $Q(u, u) < 0$  for any unit length  $u$ . For  $\gamma > 2/(K - 1)$  this is trivial. We now show that the proof of positivity can be extended to smaller values of  $\gamma$ . Call  $v_*$  the vector for which  $W^2(v_*) = \min W^2 > 0$ . Also call  $u^*$  the vector such that  $Z^2(u^*) = \max Z^2 < \infty$ . Then, if  $\gamma < 2/(K - 1)$  we have  $Q(u, u) \leq Q(u^*, v_*)$ . We also have that  $Q(u^*, v_*) < 0$  if

$$\gamma > \gamma_* = \frac{2}{\frac{W^2(v_*)}{Z^2(u^*)} + (K - 1)}. \quad (6.7)$$

and so  $\gamma_* < 2/(K - 1)$ . Hence the metric tensor is positive definite if  $\gamma > \gamma^*$ . There is a range of values  $\gamma^* < \gamma \leq 2/(K - 1)$  where positive definiteness is preserved. The only condition we assumed necessary is that  $W^2(v_*) > 0$ , which is trivially satisfied if at least two classes have different means.

## II.B Off-diagonal terms

For the diagonal metric tensor case where  $g_{\mu\nu} = \omega_\mu \delta_{\mu\nu}$  and with  $\mathbf{h}^{(0)} = \text{diag}(\sqrt{\omega_1}, \dots, \sqrt{\omega_F})$  we can apply the rescaling

$$\mathbf{x}' = \mathbf{h}^{(0)} \mathbf{x} \quad (6.8)$$

to the entire dataset. Now, standard methods can be applied to study the geometry of experiments and other statistical properties.

If the off diagonal terms are small but we want to still include them, we can proceed perturbatively in a simple manner as long as the metric tensor is a positive-semidefinite matrix. Call  $\mathbf{h} = \{h_{\mu\nu}\}$  the principal square root, i.e.  $g_{\mu\nu} = \sum_\rho h_{\mu\rho} h_{\rho\nu}$ , and we separate into the diagonal part and the off diagonal  $g_{\mu\nu} = \omega_\mu \delta_{\mu\nu} + \epsilon a_{\mu\nu}$ , where  $\epsilon$  is small. Then expanding  $h_{\mu\nu}$ :

$$h_{\mu\nu} = h_{\mu\nu}^{(0)} + \epsilon h_{\mu\nu}^{(1)} + \epsilon^2 h_{\mu\nu}^{(2)},$$

we obtain up to second order in  $\epsilon$

$$h_{\mu\nu}^{(0)} = \omega_\mu^{1/2} \delta_{\mu\nu} \quad (6.9)$$

$$h_{\mu\nu}^{(1)} = \frac{1}{\omega_\mu^{1/2} + \omega_\nu^{1/2}} a_{\mu\nu} \quad (6.10)$$

$$h_{\mu\nu}^{(2)} = -\frac{1}{\omega_\mu^{1/2} + \omega_\nu^{1/2}} \sum_\rho h_{\mu\rho}^{(1)} h_{\rho\nu}^{(1)} \quad (6.11)$$

and the rescaling is

$$\mathbf{x}' = \mathbf{h} \mathbf{x}. \quad (6.12)$$

The complexity of calculating this off diagonal rescaling grows with the cube of the number of features plus the inevitable quadratic dependence on the number of samples for an offline method. For lower dimensional problems  $\mathbf{h}$  can be obtained via SVD. Note that if there is a negative eigenvalue it can only occur if  $\gamma < \gamma^*$ . From continuity of the eigenvalues in  $\gamma$  it will be the smallest in absolute value. SVD can be used to throw away non-relevant features combinations associated to negative eigenvalues and obtain a pseudo-square root matrix in the spirit of the Moore-Penrose pseudo-inverse for values of  $\gamma$  even lower than  $\gamma^*$ .

# Bibliography

- [1] Boareto M, Jolly M, Lu M, Onuchic J, Clementi C, et al. (2015) Jagged-Delta Asymmetry in Notch Signaling Can Give Rise to a Sender/Receiver Hybrid Phenotype. *Proceedings of the National Academy of Sciences of the United States of America* .
- [2] Jolly\* M, Boareto\* M, Lu M, Onuchic J, Clementi C, et al. (under review) Operating Principles of Notch-Delta-Jagged module of cell-cell communication. *New Journal of Physics* .
- [3] Boareto\* M, Jolly\* M, Onuchic J, Ben-Jacob E (to be submitted) Inflammation gives rise to mutually reinforcing Cancer Stem Cells through Notch-Jagged signaling .
- [4] Boareto M, Cesar J, Leite V, Caticha N (2014) Supervised Variational Relevance Learning, an analytic geometric feature selection with applications to omic data sets. *IEEE Transactions on Computational Biology and Bioinformatics* .
- [5] Boareto M, Yamagishi MEB, Caticha N, Leite VBP (2012) Relationship between global structural parameters and Enzyme Commission hierarchy : Implications for function prediction. *Computational Biology and Chemistry* 40: 15–19.
- [6] Boareto M, Caticha N (2014) T-test at the probe level: An alternative method to identify the statistically significant genes for microarray data. *Microarrays* .
- [7] Barolo S, Posakony JW (2002) Three habits of highly effective signaling pathways: Principles of transcriptional control by developmental cell signaling. *Genes and Development* 16: 1167–1181.

- [8] Andersson ER, Sandberg R, Lendahl U (2011) Notch signaling: simplicity in design, versatility in function. *Development* (Cambridge, England) 138: 3593–612.
- [9] Sethi N, Kang Y (2011) Notch signalling in cancer progression and bone metastasis. *British journal of cancer* 105: 1805–10.
- [10] Shaya O, Sprinzak D (2011) From Notch signaling to fine-grained patterning: Modeling meets experiments. *Current opinion in genetics & development* 21: 732–9.
- [11] Timmerman La, Grego-Bessa J, Raya A, Bertrán E, Pérez-Pomares JM, et al. (2004) Notch promotes epithelial-mesenchymal transition during cardiac development and oncogenic transformation. *Genes & development* 18: 99–115.
- [12] Petrovic J, Formosa-Jordan P, Luna-Escalante JC, Abelló G, Ibañes M, et al. (2014) Ligand-dependent Notch signaling strength orchestrates lateral induction and lateral inhibition in the developing inner ear. *Development* (Cambridge, England) 141: 2313–24.
- [13] Manderfield LJ, High Fa, Engleka Ka, Liu F, Li L, et al. (2012) Notch activation of Jagged1 contributes to the assembly of the arterial wall. *Circulation* 125: 314–23.
- [14] Hartman BH, Reh Ta, Bermingham-McDonogh O (2010) Notch signaling specifies prosensory domains via lateral induction in the developing mammalian inner ear. *Proceedings of the National Academy of Sciences of the United States of America* 107: 15792–15797.
- [15] Morales A, Yasuda Y, Ish-Horowicz D (2002) Periodic Lunatic fringe Expression Is Controlled during Segmentation by a Cyclic Transcriptional Enhancer Responsive to Notch Signaling. *Developmental cell* 3: 63–74.
- [16] LeBon L, Lee T, Jafar-Nejad H (2014) Fringe proteins modulate Notch-ligand cis and trans interactions to specify signaling states. *eLife* .
- [17] Panin VM, Papayannopoulos V, Wilson R, Irvine KD (1997) Fringe modulates Notch-ligand interactions. *Nature* 387: 908–12.

- [18] Hicks C, Johnston SH, DiSibio G, Collazo a, Vogt TF, et al. (2000) Fringe differentially modulates Jagged1 and Delta1 signalling through Notch1 and Notch2. *Nature cell biology* 2: 515–20.
- [19] Shimizu K, Chiba S, Saito T, Kumano K, Takahashi T, et al. (2001) Manic fringe and lunatic fringe modify different sites of the Notch2 extracellular region, resulting in different signaling modulation. *The Journal of biological chemistry* 276: 25753–8.
- [20] Sprinzak D, Lakhanpal A (2011) Mutual inactivation of Notch receptors and ligands facilitates developmental patterning. *PLoS Computational Biology* 7.
- [21] Sprinzak D, Lakhanpal A, Lebon L, Santat La, Fontes ME, et al. (2010) Cis-interactions between Notch and Delta generate mutually exclusive signalling states. *Nature* 465: 86–90.
- [22] Collier JR, Monk Na, Maini PK, Lewis JH (1996) Pattern formation by lateral inhibition with feedback: a mathematical model of delta-notch intercellular signalling. *Journal of theoretical biology* 183: 429–46.
- [23] Koizumi Y, Iwasa Y, Hirashima T (2012) Mathematical study of the role of delta/notch lateral inhibition during primary branching of drosophila trachea development. *Bioophysical Journal* 103: 2549–2559.
- [24] Wang R, Liu K, Chen L, Aihara K (2011) Neural fate decisions mediated by trans-activation and cis-inhibition in Notch signaling. *Bioinformatics* (Oxford, England) 27: 3158–65.
- [25] Wang Z, Li Y, Kong D, Banerjee S, Ahmad A, et al. (2009) Acquisition of epithelial-mesenchymal transition phenotype of gemcitabine-resistant pancreatic cancer cells is linked with activation of the notch signaling pathway. *Cancer Research* 69: 2400–2407.
- [26] Campese AF, Grazioli P, de Cesaris P, Riccioli A, Bellavia D, et al. (2014) Mouse sertoli cells sustain de novo generation of regulatory t cells by triggering the notch pathway through soluble jagged1. *Biology of reproduction* .

- [27] Lu J, Ye X, Fan F, Xia L, Bhattacharya R, et al. (2013) Endothelial cells promote the colorectal cancer stem cell phenotype through a soluble form of Jagged-1. *Cancer cell* 23: 171–85.
- [28] Urs S, Roudabush A, O'Neill CF, Pinz I, Prudovsky I, et al. (2008) Soluble forms of the Notch ligands Delta1 and Jagged1 promote *in vivo* tumorigenicity in NIH3T3 fibroblasts with distinct phenotypes. *The American Journal of Pathology* 173: 865–878.
- [29] Huang Y, Lin L, Shanker a, Malhotra a, Yang L, et al. (2011) Resuscitating Cancer Immunosurveillance: Selective Stimulation of DLL1-Notch Signaling in T cells Rescues T-cell Function and Inhibits Tumor Growth. *Cancer Research* 71: 6122–6131.
- [30] Li JL, Sainson RCa, Shi W, Leek R, Harrington LS, et al. (2007) Delta-like 4 Notch ligand regulates tumor angiogenesis, improves tumor vascular function, and promotes tumor growth *in vivo*. *Cancer research* 67: 11244–11253.
- [31] Li D, Masiero M, Banham A, Harris A (2014) The Notch ligand Jagged1 as a target for anti-tumor therapy. *Frontiers in oncology* 4: 1–13.
- [32] Gupta GP, Massagué J (2006) Cancer metastasis: building a framework. *Cell* 127: 679–695.
- [33] Yamamoto M, Taguchi Y, Ito-Kureha T, Semba K, Yamaguchi N, et al. (2013) NF- $\kappa$ B non-cell-autonomously regulates cancer stem cell populations in the basal-like breast cancer subtype. *Nature communications* 4: 2299.
- [34] Phng LK, Gerhardt H (2009) Angiogenesis: a team effort coordinated by notch. *Developmental cell* 16: 196–208.
- [35] Kamdje aHN, Bassi G, Pacelli L, Malpeli G, Amati E, et al. (2012) Role of stromal cell-mediated Notch signaling in CLL resistance to chemotherapy. *Blood Cancer Journal* 2: e73.
- [36] Benedito R, Roca C, Sørensen I, Adams S, Gossler A, et al. (2009) The notch ligands Dll4 and Jagged1 have opposing effects on angiogenesis. *Cell* 137: 1124–35.

- [37] Benedito R, Hellström M (2013) Notch as a hub for signaling in angiogenesis. *Experimental Cell Research* 319: 1281–1288.
- [38] de Back W, Zhou JX, Brusch L (2013) On the role of lateral stabilization during early patterning in the pancreas. *Journal of The Royal Society Interface* 10: 20120766.
- [39] Formosa-Jordan P, Ibañes M, Ares S, Frade JM (2012) Regulation of neuronal differentiation at the neurogenic wavefront. *Development* 139: 2321–2329.
- [40] Plahte E (2001) Pattern formation in discrete cell lattices. *Journal of mathematical biology* 44: 411–445.
- [41] Owen MR, Sherratt Ja (1998) Mathematical modelling of juxtacrine cell signalling. *Mathematical Biosciences* 153: 125–150.
- [42] Miller AC, Lyons EL, Herman TG (2009) cis-Inhibition of Notch by Endogenous Delta Biases the Outcome of Lateral Inhibition. *Current Biology* 19: 1378–1383.
- [43] Cordle J, Johnson S, Tay JZY, Roversi P, Wilkin MB, et al. (2008) A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition. *Nature structural & molecular biology* 15: 849–57.
- [44] del Álamo D, Rouault H, Schweiguth F (2011) Mechanism and significance of cis-inhibition in notch signalling. *Current Biology* 21: R40–R47.
- [45] Formosa-Jordan P, Ibañes M (2014) Competition in Notch Signaling with Cis Enriches Cell Fate Decisions. *PLoS ONE* 9: e95744.
- [46] Narui Y, Salaita K (2013) Membrane Tethered Delta Activates Notch and Reveals a Role for Spatio-Mechanical Regulation of the Signaling Pathway. *Biophysical Journal* 105: 2655–65.
- [47] Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E (2013) MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proceedings of the National Academy of Sciences of the United States of America* .

- [48] Wang MM (2011) Notch signaling and Notch signaling modifiers. *The international journal of biochemistry & cell biology* 43: 1550–62.
- [49] Clewley R (2012) Hybrid Models and Biological Model Reduction with PyDSTool. *PLoS Computational Biology* 8.
- [50] Musse Aa, Meloty-Kapella L, Weinmaster G (2012) Notch ligand endocytosis: Mechanistic basis of signaling activity. *Seminars in Cell and Developmental Biology* 23: 429–436.
- [51] Hicks C, Ladi E, Lindsell C, Hsieh JJD, Hayward SD, et al. (2002) A secreted Delta1-Fc fusion protein functions both as an activator and inhibitor of Notch1 signaling. *Journal of Neuroscience Research* 68: 655–667.
- [52] Brabertz S, Bajdak K, Meidhof S, Burk U, Niedermann G, et al. (2011) The ZEB1/miR-200 feedback loop controls Notch signalling in cancer cells. *The EMBO journal* 30: 770–82.
- [53] Gill JG, Langer EM, Lindsley RC, Cai M, Murphy TL, et al. (2011) Snail and the microRNA-200 family act in opposition to regulate epithelial-to-mesenchymal transition and germ layer fate restriction in differentiating ESCs. *Stem Cells* 29: 764–776.
- [54] Amsen D, Blander JM, Lee GR, Tanigaki K, Honjo T, et al. (2004) Instruction of distinct CD4 T helper cell fates by different notch ligands on antigen-presenting cells. *Cell* 117: 515–26.
- [55] Shimojo H, Ohtsuka T, Kageyama R (2011) Dynamic expression of notch signaling genes in neural stem/progenitor cells. *Frontiers in neuroscience* 5: 78.
- [56] Beatus P, Lendahl U (1998) Notch and neurogenesis. *Journal of Neuroscience Research* 54: 125–136.
- [57] Savill NJ, Sherratt Ja (2003) Control of epidermal stem cell clusters by Notch-mediated lateral induction. *Developmental Biology* 258: 141–153.
- [58] Kopan R, Ilagan MXG (2009) The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell* 137: 216–33.

- [59] Wilkinson MA, Fitzgerald K (1994) Reciprocal Changes in Expression of the Receptor lin-72 and Its Ligand lag-2 Prior to Commitment in a *C. elegans* Cell Fate Decision. *Cell* 79: 1187–98.
- [60] Jakobsson L, Franco Ca, Bentley K, Collins RT, Ponsioen B, et al. (2010) Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting. *Nature cell biology* 12: 943–953.
- [61] Formosa-Jordan P, Ibañes M, Ares S, Frade JM (2013) Lateral inhibition and neurogenesis: novel aspects in motion. *The International Journal of Developmental Biology* 57: 341–50.
- [62] Barad O, Hornstein E, Barkai N (2011) Robust selection of sensory organ precursors by the Notch-Delta pathway. *Current opinion in cell biology* 23: 663–7.
- [63] Kiernan AE (2013) Notch signaling during cell fate determination in the inner ear. *Seminars in cell & developmental biology* 24: 470–9.
- [64] Owen MR, Sherratt JA, Wearing HJ (2000) Lateral induction by juxtacrine signaling is a new mechanism for pattern formation. *Developmental biology* 217: 54–61.
- [65] Back WD, Zhou JX, Brusch L (2013) On the role of lateral stabilization during early patterning in the On the role of lateral stabilization during early patterning in the pancreas. *Journal of Royal Society Interface* 10: 20120766.
- [66] Lewis J (1998) Notch signalling and the control of cell fate choices in vertebrates. *Seminars in cell & developmental biology* 9: 583–9.
- [67] Tien AC, Rajan A, Bellen HJ (2009) A Notch updated. *The Journal of cell biology* 184: 621–9.
- [68] Nikolaou N, Watanabe-Asaka T, Gerety S, Distel M, Köster RW, et al. (2009) Lunatic fringe promotes the lateral inhibition of neurogenesis. *Development* 136: 2523–33.
- [69] Xu K, Usary J, Kousis PC, Prat A, Wang Dy, et al. (2012) Lunatic Fringe Deficiency Cooperates with the Met / Caveolin Gene Amplicon to Induce Basal-like Breast Cancer. *Cancer Cell* 21: 626–641.

- [70] Yi F, Amarasinghe B, Dang T (2013) Manic fringe inhibits tumor growth by suppressing Notch3 degradation in lung cancer. *American journal of cancer* . . . 3: 490–499.
- [71] Zhang S, Chung WC, Wu G, Egan SE, Xu K (2014) Tumor-Suppressive Activity of Lunatic Fringe in Prostate through Differential Modulation of Notch Receptor Activation. *Neoplasia (New York, NY)* 16: 158–67.
- [72] Chen JS, Gumbayan AM, Zeller RW, Mahaffy JM (2014) An expanded Notch-Delta model exhibiting long-range patterning and incorporating MicroRNA regulation. *PLoS computational biology* 10: e1003655.
- [73] Bao B, Wang Z, Ali S, Kong D, Li Y, et al. (2011) Notch-1 induces epithelial-mesenchymal transition consistent with cancer stem cell phenotype in pancreatic cancer cells. *Cancer letters* 307: 26–36.
- [74] Wang Z, Li Y, Banerjee S, Kong D, Ahmad A, et al. (2010) Down-regulation of Notch-1 and Jagged-1 inhibits prostate cancer cell growth, migration and invasion, and induces apoptosis via inactivation of Akt, mTOR, and NF- $\kappa$ B signaling pathways. *Journal of Cellular Biochemistry* 109: 726–736.
- [75] Yoon CH, Choi YE, Koh SJ, Choi JI, Park YB, et al. (2013) High glucose-induced jagged 1 in endothelial cells disturbs notch signaling for angiogenesis: A novel mechanism of diabetic vasculopathy. *Journal of molecular and cellular cardiology* 69C: 52–66.
- [76] Jolly MK, Huang B, Lu M, Mani SA, Levine H, et al. (2014) Towards elucidating the connection between epithelial to mesenchymal transitions and stemness. *Journal of Royal Society Interface* 11: 20140962.
- [77] Lu M, Jolly MK, Onuchic J, Ben-Jacob E (2014) Toward Decoding the Principles of Cancer Metastasis Circuits. *Cancer Research* 74: 4574–4587.
- [78] Nieto MA (2013) Epithelial Plasticity: A Common Theme in Embryonic and Cancer Cells. *Science* 342: 1234850.

- [79] Arnoux V, Nassour M, Helgoualc AL, Hipskind RA, Savagner P, et al. (2008) Erk5 Controls Slug Expression and Keratinocyte Activation during Wound Healing 19: 4738–4749.
- [80] Micalizzi DS, Farabaugh SM, Ford HL (2010) Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. Journal of mammary gland biology and neoplasia 15: 117–34.
- [81] Revenu C, Gilmour D (2009) EMT 2.0: shaping epithelia through collective migration. Current opinion in genetics & development 19: 338–42.
- [82] Chigurupati S, Arumugam TV, Son TG, Lathia JD, Jameel S, et al. (2007) Involvement of notch signaling in wound healing. PloS one 2: e1167.
- [83] Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, et al. (2013) Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science (New York, NY) 339: 580–4.
- [84] Aceto N, Bardia A, Miyamoto DT, Donaldson MC, Wittner BS, et al. (2014) Circulating Tumor Cell Clusters Are Oligoclonal Precursors of Breast Cancer Metastasis. Cell 158: 1110–1122.
- [85] Bednarz-Knoll N, Alix-Panabières C, Pantel K (2012) Plasticity of disseminating cancer cells in patients with epithelial malignancies. Cancer metastasis reviews 31: 673–87.
- [86] Vallejo DM, Caparros E, Dominguez M (2011) Targeting Notch signalling by the conserved miR-8/200 microRNA family in development and cancer cells. The EMBO journal 30: 756–69.
- [87] Bu P, Chen KY, Chen JH, Wang L, Walters J, et al. (2013) A microRNA miR-34a-regulated bimodal switch targets notch in colon cancer stem cells. Cell stem cell 12: 602–15.
- [88] Pang RTK, Leung CON, Ye TM, Liu W, Chiu PCN, et al. (2010) MicroRNA-34a suppresses invasion through downregulation of Notch1 and Jagged1 in cervical carcinoma and choriocarcinoma cells. Carcinogenesis 31: 1037–44.

- [89] de Antonellis P, Medaglia C, Cusanelli E, Andolfo I, Liguori L, et al. (2011) MiR-34a targeting of Notch ligand delta-like 1 impairs CD15+/CD133+ tumor-propagating cells and supports neural differentiation in medulloblastoma. *PloS one* 6: e24584.
- [90] Sahlgren C, Gustafsson MV, Jin S, Poellinger L, Lendahl U (2008) Notch signaling mediates hypoxia-induced tumor cell migration and invasion. *Proceedings of the National Academy of Sciences of the United States of America* 105: 6392–7.
- [91] Niessen K, Fu Y, Chang L, Hoodless Pa, McFadden D, et al. (2008) Slug is a direct Notch target required for initiation of cardiac cushion cellularization. *The Journal of cell biology* 182: 315–25.
- [92] Savagner P, Kusewitt DF, Carver EA, Magnino F, Choi C, et al. (2005) Developmental transcription factor slug is required for effective re-epithelialization by adult keratinocytes. *Journal of cellular physiology* 202: 858–866.
- [93] Leroy P, Mostov KE (2007) Slug is required for cell survival during partial epithelial-mesenchymal transition of hgf-induced tubulogenesis. *Molecular biology of the cell* 18: 1943–1952.
- [94] Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, et al. (2008) The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 133: 704–715.
- [95] Morel AP, Lièvre M, Thomas C, Hinkal G, Ansieau S, et al. (2008) Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PloS one* 3: e2888.
- [96] Ombrato L, Malanchi I (2014) The emt universe: space between cancer cell dissemination and metastasis initiation. *Critical Reviews in Oncogenesis* 19.
- [97] Pattabiraman DR, Weinberg RA (2014) Tackling the cancer stem cells-what challenges do they pose? *Nature Reviews Drug Discovery* 13: 497–512.
- [98] Conigliaro a, Amicone L, Costa V, De Santis Puzzonia M, Mancone C, et al. (2013) Evidence for a common progenitor of epithelial and mesenchymal components of the liver. *Cell death and differentiation* 20: 1116–23.

- [99] Sikandar SS, Pate KT, Anderson S, Dizon D, Edwards RA, et al. (2010) Notch signaling is required for formation and self-renewal of tumor-initiating cells and for repression of secretory cell differentiation in colon cancer. *Cancer research* 70: 1469–1478.
- [100] Zhu TS, Costello Ma, Talsma CE, Flack CG, Crowley JG, et al. (2011) Endothelial cells create a stem cell niche in glioblastoma by providing NOTCH ligands that nurture self-renewal of cancer stem-like cells. *Cancer research* 71: 6061–72.
- [101] Nefedova Y, Sullivan DM, Bolick SC, Dalton WS, Gabrilovich DI (2008) Inhibition of Notch signaling induces apoptosis of myeloma cells and enhances sensitivity to chemotherapy. *Blood* 111: 2220–9.
- [102] Iliopoulos D, Hirsch Ha, Wang G, Struhl K (2011) Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. *Proceedings of the National Academy of Sciences of the United States of America* 108: 1397–402.
- [103] Lee SH, Hong HS, Liu ZX, Kim RH, Kang MK, et al. (2012) TNF-alpha enhances cancer stem cell-like phenotype via Notch-Hes1 activation in oral squamous cell carcinoma cells. *Biochemical and biophysical research communications* 424: 58–64.
- [104] Chang Q, Bournazou E, Sansone P, Berishaj M, Gao SP, et al. (2013) The IL-6 / JAK / Stat3 Feed-Forward Loop Drives Tumorigenesis. *Neoplasia* 15: 848–862, IN40–IN45.
- [105] Liu X, Johnson S, Liu S, Kanodia D, Yue W, et al. (2013) Nonlinear growth kinetics of breast cancer stem cells: implications for cancer stem cell targeted therapy. *Scientific reports* 3: 2473.
- [106] Fornari C, Beccuti M, Lanzardo S, Conti L, Balbo G, et al. (2014) A mathematical-biological joint effort to investigate the tumor-initiating ability of Cancer Stem Cells. *PloS one* 9: e106193.
- [107] Zhou D, Wang Y, Wu B (2014) A multi-phenotypic cancer model with cell plasticity. *Journal of theoretical biology* 357: 35–45.

- [108] Turner C, Kohandel M (2010) Investigating the link between epithelial-mesenchymal transition and the cancer stem cell phenotype: A mathematical approach. *Journal of theoretical biology* 265: 329–35.
- [109] Ben-Jacob E, Coffey DS, Levine H (2012) Bacterial survival strategies suggest rethinking cancer cooperativity. *Trends in microbiology* 20: 403–10.
- [110] Arwert EN, Hoste E, Watt FM (2012) Epithelial stem cells, wound healing and cancer. *Nature Reviews Cancer* 12: 170–180.
- [111] Riehl A, Németh J, Angel P, Hess J (2009) The receptor RAGE: Bridging inflammation and cancer. *Cell communication and signaling : CCS* 7: 12.
- [112] Korkaya H, Liu S, Wicha MS (2011) Regulation of Cancer Stem Cells by Cytokine Networks : Attacking Cancer ' s Inflammatory Roots : 6125–6129.
- [113] Johnston DA, Dong B, Hughes CCW (2009) TNF induction of jagged-1 in endothelial cells is NF  $\kappa$  B-dependent. *Gene* 435: 36–44.
- [114] Sansone P, Storci G, Tavolari S, Guarnieri T, Giovannini C, et al. (2007) IL-6 triggers malignant features in mammospheres from human ductal breast carcinoma and normal mammary gland 117.
- [115] Foldi J, Chung AY, Xu H, Zhu J, Outtz HH, et al. (2010) Autoamplification of notch signaling in macrophages by tlr-induced and rbp-j-dependent induction of jagged1. *The Journal of Immunology* 185: 5023–5031.
- [116] Xing F, Kobayashi A (2013) Reactive astrocytes promote the metastatic growth of breast cancer stem-like cells by activating Notch signalling in brain. *EMBO molecular medicine* : 384–396.
- [117] Morga E, Mouad-Amazzal L, Felten P, Heurtaux T, Moro M, et al. (2009) Jagged1 regulates the activation of astrocytes via modulation of nf $\kappa$ b and jak/stat/socs pathways. *Glia* 57: 1741–1753.

- [118] Rokavec M, Öner MG, Li H, Jackstadt R, Jiang L, et al. (2014) Il-6r/stat3/mir-34a feedback loop promotes emt-mediated colorectal cancer invasion and metastasis. *The Journal of clinical investigation* 124: 0–0.
- [119] Smith AK, Conneely KN, Pace TW, Mister D, Felger JC, et al. (2014) Epigenetic changes associated with inflammation in breast cancer patients treated with chemotherapy. *Brain, behavior, and immunity* 38: 227–236.
- [120] Multhoff G, Radons J (2012) Radiation, inflammation, and immune responses in cancer. *Frontiers in oncology* 2: 1–18.
- [121] Steg AD, Katre AA, Goodman B, Han HD, Nick AM, et al. (2011) Targeting the notch ligand jagged1 in both tumor cells and stroma in ovarian cancer. *Clinical Cancer Research* 17: 5674–5685.
- [122] Reedijk M, Pinnaduwage D (2008) JAG1 expression is associated with a basal phenotype and recurrence in lymph node-negative breast cancer. *Breast cancer research* . . . : 439–448.
- [123] Li Y, Wang L, Pappan L, Galliher-beckley A, Shi J (2012) IL-1  $\beta$  promotes stemness and invasiveness of colon cancer cells through Zeb1 activation : 1–13.
- [124] Yang X, Lin X, Zhong X, Kaur S, Li N, et al. (2010) Double-negative feedback loop between reprogramming factor lin28 and microrna let-7 regulates aldehyde dehydrogenase 1-positive cancer stem cells. *Cancer research* 70: 9463–9472.
- [125] Lee J, Lee J, Farquhar KS, Yun J, Frankenberger CA, et al. (2014) Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. *Proceedings of the National Academy of Sciences* 111: E364–E373.
- [126] Huang B, Lu M, Jolly M, Tsarfaty I (2014) The three-way switch operation of Rac1/RhoA GTPase-based circuit controlling amoeboid-hybrid-mesenchymal transition. *Scientific reports* : 1–11.
- [127] Augsten M (2014) Cancer-associated fibroblasts as another polarized cell type of the tumor microenvironment. *Frontiers in oncology* 4: 1–8.

- [128] Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, et al. (2007) Universally sloppy parameter sensitivities in systems biology models. PLoS Computational Biology 3: 1871–1878.
- [129] Shi L, Campbell G, Jones WD, Campagne F, et al (2010) The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. Nature biotechnology 28: 827–38.
- [130] Haury AC, Gestraud P, Vert JP (2011) The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures. PloS one 6: e28210.
- [131] Hammer B, Villmann T (2002) Generalized relevance learning vector quantization. Neural Networks 15: 1059-1068.
- [132] Schneider P, Biehl M, Hammer B (2009) Adaptive relevance matrices in learning vector quantization. Neural Computation 21: 3532-3561.
- [133] Arlt W, Biehl M, Taylor AE, Hahner S, Libe R, et al. (2011) Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors. J Clinical Endocrinology and Metabolism 96: 3775-3784.
- [134] Kästner M, Hammer B, Biehl M, Villmann T (2012) Functional relevance learning in generalized learning vector quantization. Neurocomputing 90: 85 - 95.
- [135] Schneider P, Bunte K, Stiekema H, Hammer B, Villmann T, et al. (2010) Regularization in matrix relevance learning. Neural Networks, IEEE Transactions on 21: 831–840.
- [136] Weinberger KQ, Saul LK (2009) Distance metric learning for large margin. Journal of Machine Learning Research 10: 207-244.
- [137] Kulis B (2012) Metric learning: A survey. Foundations and Trends in Machine Learning 5.
- [138] Bellet A, Habrard A, Sebban M (2013) A survey on metric learning for feature vectors and structured data. arXiv preprint arXiv:13066709 .

- [139] Mahalanobis PC (1936) On the generalized distance in statistics. Proceedings of the National Institute of Sciences of India 2: 49?55.
- [140] Gorodkin J (2004) Comparing two k-category assignments by a k-category correlation coefficient. Comput Biol Chem 28: 367–374.
- [141] Jurman G, Riccadonna S, Furlanello C (2012) A comparison of mcc and cen error measures in multi-class prediction. PloS one 7: e41882.
- [142] Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249–264.
- [143] Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, et al. (2002) MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nature genetics 30: 41–7.
- [144] Pedregosa F, *et al* (2011) Scikit-learn: Machine learning in python. Journal of Machine Learning Research 12: 2825–2830.
- [145] Ametaj BN, Zebeli Q, Saleem F, Psychogios N, Lewis MJ, et al. (2010) Metabolomics reveals unhealthy alterations in rumen metabolism with increased proportion of cereal grain in the diet of dairy cows. Metabolomics 6: 583–594.
- [146] Ein-Dor L, Zuk O, Domany E (2006) Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. Proceedings of the National Academy of Sciences of the United States of America 103: 5923–8.
- [147] Ein-Dor L, Kela I, Getz G, Givol D, Domany E (2005) Outcome signature genes in breast cancer: is there a unique set? Bioinformatics (Oxford, England) 21: 171–8.
- [148] Cai YD, Zhou GP, Chou KC (2005) Predicting enzyme family classes by hybridizing gene product composition and pseudo-amino acid composition. Journal of theoretical biology 234: 145–149.
- [149] Huang WL, Chen HM, Hwang SF, Ho SY Accurate prediction of enzyme subfamily class using an adaptive fuzzy k-nearest neighbor method. Biosystems : 405–413.

- [150] Nasibov E, Kandemir-Cavas C (2009) Efficiency analysis of KNN and minimum distance-based classifiers in enzyme family prediction. Computational biology and chemistry 33: 461–464.
- [151] Cai YD, Chou KC (2005) Using functional domain composition to predict enzyme family classes. Journal of proteome research 4: 109–111.
- [152] Webb EC (1992) Enzyme nomenclature: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes. New York: Academic Press, 862 pp.
- [153] Dobson PD, Doig AJ (2005) Predicting enzyme class from protein structure without alignments. Journal of molecular biology 345: 187–199.
- [154] Borro LC, Oliveira SRM, Yamagishi MEB, Mancini AL, Jardine JG, et al. (2006) Predicting enzyme class from protein structure using Bayesian classification. Genetics and molecular research 5: 193–202.
- [155] Bray T, Doig AJ, Warwicker J (2009) Sequence and structural features of enzymes and their active sites by EC class. Journal of molecular biology 386: 1423–1436.
- [156] Laskowski RA, Watson JD, Thornton JM (2005) ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res 33: W89-W93.
- [157] (2008). Scop database. URL <http://www.ebi.ac.uk>.
- [158] Neshich G, Togawa RC, Mancini AL, Kuser PR, Yamagishi MEB, et al. (2003) STING Millennium: a web-based suite of programs for comprehensive and simultaneous analysis of protein structure and sequence. Nucleic Acids Research 31: 3386–3392.
- [159] Bojer T, Hammer B, Schunk D, von Toschanowitz KT (2001) Relevance determination in learning vector quantization. In: PROC. OF EUROPEAN SYMPOSIUM ON ARTIFICIAL NEURAL NETWORKS.
- [160] Blatt M, Wiseman S, Domany E (1996) Superparamagnetic clustering of data. Physical Review Letters 76: 3251–3254.

- [161] Galperin MY, Walker DR, Koonin EV (1998) Analogous enzymes: independent inventions in enzyme evolution. *Genome Research* 8: 779-790.
- [162] Gherardini PF, Wass MN, Helmer-Citterich M, Sternberg MJE (2007) Convergent evolution of enzyme active sites is not a rare phenomenon. *Journal of molecular biology* 372: 817–45.
- [163] Rost B (2002) Enzyme function less conserved than anticipated. *Journal of Molecular Biology* 318: 595–608.
- [164] Martin AC, Orengo CA, Hutchinson EG, Jones S, Karmirantzou M, et al. (1998) Protein folds and functions. *Structure* 6: 875-884.
- [165] Amitai G, Shemesh A, Sitbon E, Shkler M, Netanel D, et al. (2004) Network analysis of protein structures identifies functional residues. *Journal of molecular biology* 344: 1135-1146.
- [166] Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, et al. (2004) ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics* 20: 1322–1324.
- [167] Capra JA, Singh M (2007) Predicting functionally important residues from sequence conservation. *Bioinformatics* 23: 1875–1882.
- [168] Bray T, Chan P, Bougouffa S, Greaves R, Doig AJ, et al. (2009) SitesIdentify: a protein functional site prediction tool. *BMC Bioinformatics* 10: 379.
- [169] Brylinski M, Prymula K, Jurkowski W, Kochanczyk M, Stawowczyk E, et al. (2007) Prediction of functional sites based on the fuzzy oil drop model. *PLoS Computational Biology* 3: 909–923.
- [170] Onuchic JN, Wolynes PG (2004) Theory of protein folding. *Current Opinion in Structural Biology* 14: 70-75.
- [171] Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. *Annual Review of Physical Chemistry* 48.

- [172] Bryngelson JD, Wolynes PG (1987) Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences of the United States of America* 84: 7524–7528.
- [173] Goldstein RA, Luthey-Schulten ZA, Wolynes PG (1992) Optimal protein-folding codes from spin-glass theory. *Proceedings of the National Academy of Sciences of the United States of America* 89: 4918–4922.
- [174] Ferreiro DU, Hegler JA, Komives EA, Wolynes PG (2007) Localizing frustration in native proteins and protein assemblies. *Proceedings of the National Academy of Sciences of the United States of America* 104: 19819–19824.
- [175] Orengo CA, Flores TP, Jones DT, Taylor WR, Thornton JM (1993) Recurring structural motifs in proteins with different functions. *Current biology* 3: 131–139.
- [176] Lobley A, Swindells MB, Orengo CA, Jones DT (2007) Inferring function using patterns of native disorder in proteins. *PLoS Comput Biol* 3: e162.
- [177] Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* 18: 756–764.
- [178] Sadowski MI, Jones DT (2009) The sequence-structure relationship and protein function prediction. *Curr Opin Struct Biol* : 357–362.
- [179] Li C, Wong WH (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proceedings of the National Academy of Sciences of the United States of America* 98: 31–36.
- [180] Wu Z (2009) A review of statistical methods for preprocessing oligonucleotide microarrays. *Statistical methods in medical research* 18: 533–541.
- [181] Affymetrix (2005). Guide to probe logarithmic intensity error (plier) estimation. URL <http://www.affymetrix.com>.
- [182] Shi L, Tong W, Fang H, Scherf U, Han J, et al. (2005) Cross-platform comparability of microarray technology: intra-platform consistency and appropriate data analysis procedures are essential. *BMC bioinformatics* .

- [183] Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, et al. (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature biotechnology* 24: 1151–61.
- [184] Allison DB, Cui X, Page GP, Sabripour M (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nature reviews Genetics* 7: 55–65.
- [185] Jeanmougin M, de Reynies A, Marisa L, Paccard C, Nuel G, et al. (2010) Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies. *PloS one* 5.
- [186] Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* 98: 5116–21.
- [187] Cui X, Hwang J, Qiu J, Blades N, Churchill G (2005) Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics* 6: 59–75.
- [188] Wright GW, Simon RM (2003) A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics* 19: 2448–2455.
- [189] Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* 3.
- [190] Zeisel A, Amir A, Kostler WJ, Domany E (2010) Intensity dependent estimation of noise in microarrays improves detection of differentially expressed genes. *BMC bioinformatics* 11: 400.
- [191] Baldi P, Long AD (2001) A bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics* 17: 509–519.
- [192] Stevens JR, Bell JL, Aston KI, White KL (2010) A comparison of probe-level and probeset models for small-sample gene expression data. *BMC bioinformatics* 11: 281.

- [193] Lemieux S (2006) Probe-level linear model fitting and mixture modeling results in high accuracy detection of differential gene expression. *BMC bioinformatics* 7: 391.
- [194] Barrera L, Benner C, Tao Y (2004) Leveraging two-way probe-level block design for identifying differential gene expression with high-density oligonucleotide arrays. *BMC bioinformatics* 14: 1–14.
- [195] Astrand M, Mostad P, Rudemo M (2008) Empirical Bayes models for multiple probe type microarrays at the probe level. *BMC bioinformatics* 9: 156.
- [196] Chu JT (1955) On the distribution of the sample median. *The Annals of Mathematical Statistics* : 112–116.
- [197] Affymetrix. Latin square data for expression algorithm assessment. URL <http://www.affymetrix.com>.
- [198] Cope LM, Irizarry RA, Jaffee HA, Wu Z, Speed TP (2004) A benchmark for affymetrix genechip expression measures. *Bioinformatics* 20: 323–331.
- [199] Bolstad BM, Irizarry Ra, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics (Oxford, England)* 19: 185–93.
- [200] Shi L, Campbell G, Jones WD, Campagne F, Wen Z, et al. (2010) The microarray quality control (maqc)-ii study of common practices for the development and validation of microarray-based predictive models. *Nature biotechnology* 28: 827.
- [201] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, et al. (1999) Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science* 286: 531–537.
- [202] Gyorffy B, Molnar B, Lage H, Szallasi Z, Eklund AC (2009) Evaluation of microarray preprocessing algorithms based on concordance with RT-PCR in clinical samples. *PloS one* 4: e5645.
- [203] Therneau TM, Ballman KV (2008) What does PLIER really do? *Cancer informatics* 6: 423–31.