Modeling Host Immune Responses in Infectious Diseases

Meghna Verma

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Translational Biology, Medicine and Health

Josep Bassaganya-Riera (Chair)
Raquel Hontecillas
Vida Abedi
Stefan Hoops

01 November 2019
Blacksburg, Virginia, USA

Keywords: immune-system, infectious diseases, computational models, agent-based models, ordinary differential equations, sensitivity analysis, calibration, Helicobacter pylori.

Copyright 2019, Meghna Verma
Modeling Host Immune Responses in Infectious Diseases

Meghna Verma

ABSTRACT

Infectious diseases caused by bacteria, fungi, viruses and parasites have affected humans historically. Infectious diseases remain a major cause of premature death and a public health concern globally with increased mortality and significant economic burden. Unvaccinated individuals, people with suppressed and compromised immune systems are at higher risk of suffering from infectious diseases. In spite of significant advancements in infectious diseases research, the control or treatment process faces challenges. The mucosal immune system plays a crucial role in safeguarding the body from harmful pathogens, while being constantly exposed to the environment. To develop treatment options for infectious diseases, it is vital to understand the immune responses that occur during infection. The two infectious diseases presented here are: i) Helicobacter pylori infection and ii) human immunodeficiency (HIV) and human papillomavirus (HPV) co-infection. H pylori, is a bacterium that colonizes the stomach and causes gastric cancer in 1-2% but is beneficial for protection against allergies and gastroesophageal diseases. An estimated 85% of H pylori colonized individuals show no detrimental effects. HIV is a virus that causes AIDS, one of the deadliest and most persistent epidemics. HIV-infected patients are at an increased risk of co-infection with HPV, and report an increased incidence of oral cancer. The goal of this thesis is to elucidate the host immune responses in infectious diseases via the use of computational and mathematical models. First, the thesis reviews the need for computational and mathematical models to study the immune responses in the course of infectious diseases. Second, it presents a novel sensitivity analysis method that identifies important parameters in a hybrid (agent-based/equation-based) model of H pylori infection. Third, it introduces a novel model representing the HIV/HPV co-infection and compares the simulation results with a clinical study. Fourth, it discusses the need of advanced modeling technologies to achieve a personalized systems wide approach and the challenges that can be encountered in the process. Taken together, the work in this dissertation presents modeling approaches that could lead to the identification of host immune factors in infectious diseases in a predictive and more resource-efficient manner.
Modeling Host Immune Responses in Infectious Diseases

Meghna Verma

GENERAL AUDIENCE ABSTRACT

Infectious diseases caused by bacteria, fungi, viruses and parasites have affected humans historically. Infectious diseases remain a major cause of premature death and a public health concern globally with increased mortality and significant economic burden. These infections can occur either via air, travel to at-risk places, direct person-to-person contact with an infected individual or through water or fecal route. Unvaccinated individuals, individuals with suppressed and compromised immune system such as that in HIV carriers are at higher risk of getting infectious diseases. In spite of significant advancements in infectious diseases research, the control and treatment of these diseases faces numerous challenges. The mucosal immune system plays a crucial role in safeguarding the body from harmful pathogens, while being exposed to the environment, mainly food antigens. To develop treatment options for infectious diseases, it is vital to understand the immune responses that occur during infection. In this work, we focus on gut immune system that acts like an ecosystem comprising of trillions of interacting cells and molecules, including members of the microbiome. The goal of this dissertation is to develop computational models that can simulate host immune responses in two infectious diseases- i) Helicobacter pylori infection and ii) human immunodeficiency virus (HIV)-human papilloma virus (HPV) co-infection. Firstly, it reviews the various mathematical techniques and systems biology based methods. Second, it introduces a “hybrid” model that combines different mathematical and statistical approaches to study H. pylori infection. Third, it highlights the development of a novel HIV/HPV co-infection model and compares the results from a clinical trial study. Fourth, it discusses the challenges that can be encountered in adapting machine learning based computational technologies. Taken together, the work in this dissertation presents modeling approaches that could lead to the identification of host immune factors in infectious diseases in a predictive and more resourceful way.
Dedication

This dissertation is dedicated to my dad (Mani Kant Verma), mom (Suman Lata Verma) and my sister (Swapnil Verma).

Throughout my life, my parent’s sacrifices have been a constant inspiration for me to work harder and my sister’s life challenges gave a purpose to my research.
Acknowledgements

I want to express my sincere gratitude to my advisor, Dr. Josep Bassaganya-Riera and Dr. Raquel Hontecillas for their support. They gave me the opportunity to pursue my doctoral work at the Nutritional Immunology and Molecular Medicine Laboratory (NIMML), Virginia Tech. The breadth of experience and opportunity to work on various infectious diseases problems that I pursued will be extremely valuable throughout my career. Dr. Bassaganya-Riera trusted in my abilities and gave me this exciting opportunity to work on various projects at NIMML. Thank you for always supporting me whether it be with research, visa or health related issues. Dr. Hontecillas has always helped me explain the experimental aspects of the modeling work and provided me the knowledge of immunology and infectious diseases over the years. Thank you for helping me learn how to formulate meaningful hypothesis and for trusting me and giving me an opportunity to be able to learn wet lab techniques. I am grateful for all the support that Josep and Raquel provided me over the years.

The guidance and support from my dissertation committee members: Dr. Vida Abedi, Dr. Stefan Hoops is also very appreciated. Dr. Abedi helped me transition to NIMML and has been a mentor that always pushed me to go a step further. Dr. Hoops has always been a constant support over the years. He always had an open-door policy which helped me develop as a researcher, I thank him for all the technical guidance, his patience and numerous work hours he spent debugging and discussing the collaborative projects with me. Dr. Luis S. Mayorga and I worked great together as a team and he is mentor who has always helped me believe on my capabilities. I am so thankful that Luis visited as a Fulbright scholar during my time in grad school and I got to know him and learn from him as a researcher. My ex-officio committee members Dr. Audra Van-Wart, Dr. Steven Poelzing and Dr. Michael Friedlander who not only gave me an opportunity to be in the first cohort of an exciting graduate program - Translational Biology, Medicine and Health but also provided me support in terms of bridge funding. I also wanted to thank Jay Read, Rebecca Guynn, Erin Cassidy and Vicky Stec with all help in the administrative matters. Thank you for always making sure that the financial matters were well taken care of.

It has been helpful to have friendly lab mates. Chase Heltzel and Nuria Tubau-Juni helped me in their position as lab managers to make sure that I was compliant with all the lab trainings. Dr. Andrew Leber helped navigate the computational projects at NIMML, I really appreciated his help in the agent-modeling project. Dr. Pinyi Lu helped in learning about machine learning algorithms and its application in Crohn's diseases patients. Nuria helped me learn mouse handling techniques, colony management and wet lab skills. I have mentored 9 undergraduate students, Gavin Vess, Ariana Herst, Misha Michael, Wyatt Lansford, Shivani Garg, Andrew Shedlock, Conor Dewey, David Dachille, Sara Elhannea and Alison Ritz. I learned from each one of you and mentoring all of my students was an absolute pleasure and one of the favorite parts of my graduate time here. The company of all the other undergraduate students in the lab has been extremely helpful.
Colleagues in the Fralin Life Sciences Institute, from the Kale lab and Hauf lab have provided a friendly and stimulating academic environment.

I would also like to thank my fantastic collaborators – Dr. Stanca M. Ciupe (Virgina Tech, Department of Mathematics), Dr. Luis S. Mayorga (IHEM CONICET Mendoza, Argentina), Dr. Xi Chen (Virginia Tech, Industrial Systems Engineering), Dr. Shiv Kale (Fralin Life Sciences Institute), Dr. Arindham Fadikar (Department of Statistics, Virginia Tech) and Dr. Sook Ha (Virginia Military Institute, Computer and Information Science). I learned so much from each one of you, having meaningful discussions, working on model simulations, running analyses, debugging the codes, and writing manuscripts helped me gain valuable experience in collaborating in an interdisciplinary environment.

Finally, I would like to thank my family and friends. My parents and my sister, who always been a constant support for me. Thank you for always being there for me. From talking at 3am in night to celebrating all our birthdays via Skype calls, I know how much you have sacrificed just so that I can achieve my goals. It would not have been possible without your support. My close relatives - Dr. Gyan (John) Prakash (my uncle), Dr. Savita Prakash (aunt) and my cousins Dr. Shivaani Prakash and Gaurav Prakash who have always believed in me and their love and support has seen me through. My uncle and aunt have been my parents away from home, thank you for all your support, encouragement and love. Thank you for always believing in me and making sure that I never felt alone. Dr. Harshawardhan Deshpande (his cat Gracie, thank you for helping me discover the whole new ‘animal lover’ side of me), Dr. Christina K. Lee and Dr. Yeonwoo Lebovitz. Thank you for always listening to me, and for all the laughs and good times. You guys are the best! Vishal Thakur, Amruta Nandgaonkar, Dhanashree Pimple, Dhanesh Neelamana, Udayashree Datla and Saket Choudhary my best friends and confidants have continued to support me through the ups and downs in graduate school. To all my friends in India who visited my parents while I was away to make sure they are okay and helped them not feel my absence, thank you Vishal Thakur, Amrapali Dongre and Sheetal Salgarkar-Daiv, you guys are awesome! Dr. Nrusingh Mohapatra, Shilpa Soni, and their adorable daughter Neeva, have been my family in Blacksburg. Thank you for everything you have done for me, Shilpa and Nrusinghji. Their home has been my home away from home for the last 5 years. Dr. Pabitra Dash and Reena ji, thank you for making sure I wasn’t alone when my friends graduated and moved from Virginia Tech. All of my friends in Blacksburg, Roanoke, Mumbai and in my VT Badminton group made sure I wasn’t alone during my time here. With all the support that I have gotten over the years, I don’t think it would have been possible without the help and support that each one of you has provided. Thank you, I will forever be grateful to each one of you.
Attributions

The work in this dissertation resulted from efforts from following team members who contributed towards the projects outlined below.

Chapter 1
Josep Bassaganya-Riera (Virginia Tech, Nutritional Immunology and Molecular Medicine Laboratory) and Raquel Hontecillas (Virginia Tech, Nutritional Immunology and Molecular Medicine Laboratory) contributed to reviewing the chapter.

Chapter 2
Josep Bassaganya-Riera and Raquel Hontecillas and Vida Abedi (Virginia Tech, Nutritional Immunology and Molecular Medicine Laboratory) contributed to the design of the paper. Adria Carbo (Virginia Tech, Nutritional Immunology and Molecular Medicine Laboratory), Cassandra Philipson (Virginia Tech, Nutritional Immunology and Molecular Medicine Laboratory), Andrew Leber (Virginia Tech, Nutritional Immunology and Molecular Medicine Laboratory), and Nuria Tubau-Juni (Virginia Tech, Nutritional Immunology and Molecular Medicine Laboratory) contributed to the review and helped in making of the figures.

Chapter 3
Josep Bassaganya-Riera, Raquel Hontecillas, Andrew Leber and Stefan Hoops helped formulate the model. Stefan Hoops, Andrew Leber contributed towards implementing the code architecture. Xi Chen (Virginia Tech, Industrial Systems Engineering) contributed towards writing codes for global sensitivity analysis and generated the design matrices. Nuria Tubau-Juni contributed towards generating experimental data. Josep Bassaganya-Riera, Vida Abedi and Raquel Hontecillas supervised the project and edited the manuscript. Luis S. Mayorga (Universidad Nacional de Cuyo, CONICET) helped in the visualization of the simulation and provided critical feedback.

Chapter 4
Josep Bassaganya-Riera, Raquel Hontecillas and Stanca M. Ciupe helped formulate the model and supervised the project. Samantha Erwin (Virginia Tech, Department of Mathematics) and Stanca M. Ciupe (Virginia Tech, Department of Mathematics) contributed in the model analysis and creating the MATLAB version of the figures published. Andrew Leber contributed towards feedback on the results and discussion. Josep Bassaganya-Riera, Stanca M. Ciupe and Raquel Hontecillas reviewed and edited the work.

Chapter 5
Josep Bassaganya-Riera, Raquel Hontecillas and Vida Abedi helped design the architecture of review. Josep Bassaganya-Riera, Raquel Hontecillas, Vida Abedi and Nuria Tubau-Juni helped with the edits.

Chapter 6
Josep Bassaganya-Riera and Raquel Hontecillas contributed to reviewing the chapter.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>i</td>
</tr>
<tr>
<td>Academic</td>
<td>ii</td>
</tr>
<tr>
<td>General Audience</td>
<td></td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>v</td>
</tr>
<tr>
<td>Attributions</td>
<td>vi</td>
</tr>
<tr>
<td>Chapter 1: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 2: Modeling-Enabled Systems Nutritional Immunology</td>
<td>3</td>
</tr>
<tr>
<td>Abstract</td>
<td>3</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>3</td>
</tr>
<tr>
<td>2. Current approaches and the urgent need for paradigm shift</td>
<td>12</td>
</tr>
<tr>
<td>3. Computational Tools</td>
<td>18</td>
</tr>
<tr>
<td>4. Challenges in computational capabilities</td>
<td>18</td>
</tr>
<tr>
<td>5. Developing information processing representations of Systems Nutritional Immunology</td>
<td>19</td>
</tr>
<tr>
<td>Chapter 3: High-Resolution Computational Modeling of Immune Responses in the Gut</td>
<td>22</td>
</tr>
<tr>
<td>Abstract</td>
<td>22</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>24</td>
</tr>
<tr>
<td>2. Methods</td>
<td>24</td>
</tr>
<tr>
<td>3. Results</td>
<td>39</td>
</tr>
<tr>
<td>4. Discussion</td>
<td>47</td>
</tr>
<tr>
<td>Chapter 4: Modeling the Mechanisms by Which HIV-Associated Immunosuppression Influences HPV Persistence at the Oral Mucosa</td>
<td>69</td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td>----</td>
</tr>
<tr>
<td>Abstract</td>
<td>69</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>69</td>
</tr>
<tr>
<td>2. Methods</td>
<td>71</td>
</tr>
<tr>
<td>3. Results</td>
<td>74</td>
</tr>
<tr>
<td>4. Discussion</td>
<td>84</td>
</tr>
<tr>
<td>5. Supporting Files</td>
<td>86</td>
</tr>
</tbody>
</table>

**Chapter 5: Challenges in Personalized Nutrition and Health**

<table>
<thead>
<tr>
<th>Abstract</th>
<th>91</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>91</td>
</tr>
<tr>
<td>2. Challenges</td>
<td>94</td>
</tr>
<tr>
<td>3. Concluding Remarks</td>
<td>104</td>
</tr>
</tbody>
</table>

**Chapter 6: Conclusion**

| 106 |
Chapter 1. Introduction

Infectious diseases remain a major cause of premature death and a public health concern globally with increased mortality and significant economic burden. The infectious diseases including lower respiratory illnesses, HIV/AIDS, Tuberculosis and diarrheal diseases were some of the major causes of deaths worldwide in 2016 as reported by the World Health Organization. Infectious diseases have affected humans historically. The pandemic of 1918 Spanish flu killed up to 50 million worldwide and more than half a million people in the United States. These diseases are caused by bacteria, fungi, viruses and parasites either via air, exposure to infected individual or fecal route. Although there have been significant advancements in infectious diseases research, the control or treatment process faces challenges. Further, an increasing resistance of pathogens to the current treatments add to the difficulty of combating these diseases. The goal of the studies presented in this dissertation is aimed at elucidating the host immune responses in infectious diseases, via the use of mathematical models that could lead to the identification of host factors in a predictive and inexpensive way.

The immune system is comprised of a complex network of interactions between the molecules, pathways, host cells and pathogens. Experimental studies help collect huge amounts of cellular and biochemical data that can interrogate the immune system at different scales ranging from genes, cells, tissues, organs and tissue scale, however, this approach is reductionist. Reductionist approaches are limited in their ability to study the complex interactions. Employing system level approaches can aid in the study of complex interactions across - i) dimensions ranging from intracellular, intercellular, cellular and tissue scales and ii) time scales ranging from nanoseconds, seconds, minutes to days. The knowledge generated at the different scales need to be integrated to recognize the emergent behavior of the system as a whole. Mathematical and computational modeling is useful as it adopts a predictive and inexpensive approach to describe, simulate and analyze complex behaviors in multi-dimensional immune system network. The findings from the computational models can complement preclinical studies and the generation of novel hypotheses to guide the design of new experiments. The models can help track biological events that are in line with the hypothesis and also provide insights on whether certain hypothesis testing is feasible. Chapter 2 of this dissertation highlights the importance of a required paradigm shift from reductionist to systems immunology based methods and provides an overview of the computational methods that can be employed to capture the features of a massively interacting complex immune system.

The two research studies presented in Chapters 3 and 4 of this dissertation highlight the different mathematical modeling techniques employed to study influential aspects of the host immune responses (specifically in the mucosal immune system) in Helicobacter pylori infection, and human papillomavirus (HPV) and human immunodeficiency virus (HIV) co-infections. The techniques used comprise of –i) “hybrid” model that combines the features of an agent-based model (ABM) with equation-based methods such as ordinary and partial differential equations and ii) equation-based methods. ABM is a rule-based, bottom-up approach which focuses on spatial and temporal aspects of the components modeled, also known as ‘agents’. There is an increased compartmentalization and spatial discretization that enhances the ability to model the architecture of the system being modeled. The major players in the ABM, ‘agents’, move, follow rules and interact with other agents of the system thereby contributing towards the emergent property of the system. The ABM, generate a simulation of interacting agents in a virtual world created by the
user to simulate an *in silico* model. Equation-based methods such as ordinary differential equations (ODE), on the other hand, are deterministic and based on the average response of cells over time. These models do not incorporate the spatial features required to understand the immune cell dynamics.

Sensitivity analysis (SA) is a crucial step in the model analysis that aids in studying the influence of varying the input parameters on model outputs. Further, it helps evaluate the robustness of the model outputs when the input parameter values are changed. The range of parameter value selection is based on a range of assumptions. SA helps quantify the uncertainty caused due to the parameter set on the model outcomes. The output variability can be attributed to the input parameters and various “what-if” conditions can be tested following SA. The different sensitivity analysis techniques focused here include local sensitivity analysis and global sensitivity analysis - regression based and Sobol’s methods. The local methods include taking the partial derivative of the output with respect to an input. It involves studying the effect of varying each model parameter one at a time, on the model output. Global sensitivity analysis is focused on studying the interaction effect of a wide range of parameter values. The proportion of the model variance is attributed to changes in the input parameters. Regression based methods include fitting regression function of input parameters on the model output. Sobol’ based methods decompose the model variance and quantitatively measure the effect of each input using Sobol’ first order and interaction effects using total order sensitivity indexes.

Lastly, even though computational approaches are predictive and are capable of elucidating the complex mechanisms of the immune system, these are not free from challenges. Chapter 5, of the dissertation highlights the challenges that can be encountered when advanced computational approaches are employed to make recommendations towards personalized health.
Chapter 2. Modeling-Enabled Systems Nutritional Immunology

This work was published in *Frontiers in Nutrition*:


**Abstract**

This review highlights the fundamental role of nutrition in the maintenance of health, the immune response, and disease prevention. Emerging global mechanistic insights in the field of nutritional immunology cannot be gained through reductionist methods alone or by analyzing a single nutrient at a time. We propose to investigate nutritional immunology as a massively interacting system of interconnected multistage and multiscale networks that encompass hidden mechanisms by which nutrition, microbiome, metabolism, genetic predisposition, and the immune system interact to delineate health and disease. The review sets an unconventional path to apply complex science methodologies to nutritional immunology research, discovery, and development through “use cases” centered around the impact of nutrition on the gut microbiome and immune responses. Our systems nutritional immunology analyses, which include modeling and informatics methodologies in combination with pre-clinical and clinical studies, have the potential to discover emerging systems-wide properties at the interface of the immune system, nutrition, microbiome, and metabolism.

**Introduction**

The knowledge that food affects health was first mentioned in the writings of ancient Egyptians and Indians [1-3]. Around 2,500 years ago, Hippocrates, stated “Let food be your medicine and medicine be your food” [4]. Modern nutritional immunology dates back to the 18th century, when the explanation of lymphoid tissue atrophy in malnourished population in England [1, 3] suggested an association between nutritional status and immune function. Epidemiological and clinical data also suggests that nutritional deficiencies of essential dietary components such as vitamins and micronutrients alter immune competence and increase the risk of infection. The deficiency of adequate macronutrients and selected micronutrients such as zinc, selenium, iron, copper, vitamins A, B-6, C, E lead to immune deficiency-related infections in children [5, 6]. Micronutrient deficiencies affect innate immune responses as well as adaptive cellular immune responses [7]. The immune response is dependent on the nutritional components of food intake, which modulates the induction of regulatory versus effector response at the gut mucosal level [3]. However, recent studies [3] suggest that the current immune deficiency cases are also the result of increased stress, increased caloric intake, obesity, autoimmunity, allergic disorders and an aging population, which do not necessarily relate to under-nutrition. Thus, unbalanced nutrition, unhealthy lifestyle choices,
limited physical activity and the effect of the environment in general, compromise the host immune response, thereby increasing susceptibility to a wide range of diseases. The field of nutritional immunology primarily focuses on the role of diet and its nutritional contents in disease prevention. However, advancement in the field of nutritional immunology has not been investigated through the point of view of a massively interacting system of interconnected networks that includes four key players - nutrition, microbiome, metabolism and the immune system. Recent evidence [8] also suggests the involvement of diet and the role of composition of microbiota in reduced risk of Parkinson’s disease (PD). There are findings that support the role of altered gut microbiome involved in influencing the activity of enteric neurons in PD patients [8]. Although it is still unclear, the neuroendocrine system can be considered as an important part of the massively interacting multistage networks that define health and wellness. An understanding of the interaction between networks can help design better strategies for primary prevention for diseases such as PD, which show the involvement of gut brain axis in the disease pathogenesis [8]. The investigation from the above mentioned point of view requires modeling tools, informatics techniques and major computational resources in order to gain a better understanding of the mechanisms by which the four key players interact, to delineate health and disease. The vast aspects of this interconnected network operate on the basis of complex regulatory networks that can be analyzed in a well-defined manner using mathematical and computational modeling. The recent modeling frameworks applied include the use of 1) ordinary differential equations that are used for cancer immunology, natural killer cell responses, B cell responses, (naïve and memory), T regulatory cell dynamics and T cell responses; 2) partial differential equations are used for modeling age structured and spatio-temporal models; 3) stochastic differential equations account for noise and sporadic events, 4) agent-based models account for probabilistic uncertainty in biological interaction [9], and 5) advanced machine learning algorithms that correlate cellular and molecular events to changes in health and disease outcomes. In the following sections we dissect the essence of interactions between the four key players following with the review of technological advances in the field of nutritional immunology research and development.

1. The interplay between diet, microbiome, metabolism and immune response

The proper nutritional supply during the period of gestation, neonatal maturation and weaning contributes towards the development of balanced immune responses. With an increasing shift in our focus towards using dietary interventions to regulate the host defense, it is important to understand the effect of overall nutrition derived from these interventions. The nutritional quality of the wholesome diet modulates the interactions between the immune system, microbiome and metabolism. It is estimated that demand of feeding a population will increase up to 9 billion people needing food by 2,050, which necessitates the need for devising methods that not only meet the demand but also ensure continuous wholesome food supply [10]. Therefore, understanding the relationship between immune system, microbiome, metabolism regulated by nutrition (as shown in Figure 1) will assist in targeting one component at a time, while recognizing their systems-wide effects. This would lead to identification of emerging properties of this complex system and utilization of the newly derived information and knowledge for improved health outcomes.
1.1 Microbiome, and its interaction with nutrition, immune response and metabolism

Microbes are important components of the human ecosystem, and they account for approximately 100 trillion including both the ones residing outside as well as inside the human body [11, 12]. The gut microbiome is a key player in regulating the defense responses and metabolism, thereby contributing towards shaping the immune responses (regulatory or effector) and aiding in the maturation of the immune system. The various physiological factors responsible for differences in genetic elements of the microbiome within a host includes diet, geographical location, and environmental interaction [13]. The interactions between the gut microbiome, immune system, metabolism and nutrition are crucial determinants of health outcomes. However, their systems-wide mechanisms of interaction remain largely unknown. The advent of computational modeling and informatics provides the technology, to integrate and comprehensively analyze the multiscale interactions within such networks. Thus, a systems wide approach can provide significant insights into nutritional regulation of this holistic network, without unnecessarily resorting to reductionism.

1.1.1 Interplay between microbiome and nutrition

Diet and nutritional status are the key players in defining the composition and function of the gut microbiome as well as the host immune response. The nutritional value of food is influenced by microbial content inside a person’s gut. A study by Turnbaugh, P.J., et al in 2008 [14]
demonstrated that transfer of microbiota from mice with diet-induced obesity to lean germ-free mice, showed a greater fat deposition in the lean mice versus the lean ones with transplants from the lean donors. Another study by Turnbaugh, P.J., et al [15] explored the use of humanized gnotobiotic mice wherein adult human fecal microbial communities were transplanted into germ-free mice to show the effect of Western diet on the varying bacterial colonization in adult mice. The switch from the regular to Western diet showed the colonization of Bacilli classes of Firmicutes along the length of the gut, lead to increased adiposity. The effect was shown to be reversible, based on the combinations of recipient-donor diets. These studies [14-17] show that dietary intake influences the composition and activity of the gut microbiome in humans. Specific strains of bacteria have been implicated in the regulation of the intestinal homeostasis, that deliver regulatory signals to the epithelium and the mucosal immune system [18]. Even a short-term consumption of animal-based diet versus plant products has a differential effect on the bacterial colonization inside the gut [19]. A recent study by Daniel et al. [20] showed that a high fat diet induced changes in the chemical composition of cecum thereby causing changes in the bacterial physiology and metabolism. Furthermore, the distance between diet dependent clusters of the microbial composition was higher than microbiota driven clusters, thereby illustrating how diet can alter the microbiota profiles to a higher extent than bacterial composition. The effect of diet on the composition of every individual’s microbiome is shown to be individual-specific at the operational taxonomic units and stable over a period of time in a healthy adult [21]. However, as the individual ages there is an extreme variability observed in composition of the core microbiota. Furthermore, there are other environmental factors, such as body weight, physical activity, and exposure to toxins, which also play an important role on the composition of microbiota. A comprehensive understanding of nutritional quality of the dietary interventions [22-25] that modulate the components of the gut microbiota and mucosal immune responses can prove useful for maintenance of health. A systems level framework that integrates various in vitro and in vivo models, including human data can facilitate the systems-wide mechanistic insights [26].

1.1.2 Role of the microbiome in shaping a healthy immune system

Microbiome plays a crucial role in shaping the functions of the immune system thereby providing a protective mechanism to fight against infection. The commensal bacteria help in maintaining the balance with the foreign (often pathogenic) bacteria, by modulating the components of host innate immune system. A dysregulation of homeostasis between host and gut microbes leads to dysbiosis, which can give rise to pathogenic states such as inflammatory bowel disease (IBD) [21, 27] as shown in the network model of IBD in figure 2. A change in the composition of gut microbes has been associated with development of asthma in animal models. A recent study by Arrieta et al., 2015 [28] demonstrated that infants who exhibit transient gut microbial dysbiosis during the early days of life are at high risk of asthma. The inoculation of germ free mice with the bacterial genera Lachnospira, Veillonella, Faecalibacterium and Rothia (missing in children at high risk asthma), ameliorated the airway inflammation in germfree adult offspring [28]. The study elucidates the role of gut microbiome in protecting the body against asthma. A recent study by Fonseca et al., [29] showed that during the post resolution of infection stage from Y. pseudotuberculosis, the signals derived from the gut microbiota aided in the maintenance of inflammatory mesentery remodeling and restoration of mucosal immunity. The intestinal immune system plays an important role in maintaining the balance of commensal and foreign microorganisms inside the gut along with keeping the diversity of the commensal microorganisms. However, due to high bacterial densities inside the gut, the task is challenging as compared to other organs and tissues.
The immune system has adopted certain ways such as immunological tolerance by diverting various resources to segregate the microbiome on the luminal side of the epithelial barrier [30]. The production of mucus by the goblet cells residing in the intestinal epithelium creates a protective layer that separates the commensal and pathogenic bacteria. This protective zone aids in the maintenance of the symbiotic relationship with the lumen microbiota. The compartmentalization of pathogenic bacteria includes the trapping of bacteria inside the mucus layer, complement associated bacterial killing and promotion of phagocytosis of bacteria that invade the epithelial barrier [31]. Another mechanism that promotes the segregation of the bacterial colonies is the secretion of antimicrobial proteins. Activation of the intestinal epithelial cells triggers the expression of antimicrobial proteins that provides a protective mechanism against the invasion of pathogenic bacteria into the host tissues [31]. The production of IgA also helps in the maintenance of the symbiotic relationship, but the mechanisms of protection by IgA remain unclear.

Overall the changes in the composition of the gut microbiome can modulate the induction of regulatory versus effector immune responses. Probiotics have been shown to beneficially modulate the intestinal ecosystem. Another group of non-digestible food ingredients is the prebiotics that favors the growth of health promoting bacteria, proving beneficial to the host [18, 22, 25, 32]. A large group of prebiotics comprises the carbohydrates that are indigestible by human digestive enzymes such as resistant starches. The bacterial community inside the intestinal mucosal surface ferment the undigested fibers to generate lipid molecules such as oleic acid, conjugated linoleic acids (CLAs), and short chain fatty acids (SCFAs) such as acetate, propionate, and butyrate that influences the colonic mucosal growth and intestinal permeability that enhances the gastrointestinal health [32, 33]. A study by Bassaganya-Riera et al. [33] compared the anti-inflammatory efficacy and studied how different dietary soluble fibers and resistant starch influence regulatory T cells (T regs), colonic peroxisome proliferator-activated receptor gamma (PPAR γ), and interferon gamma (IFN-γ) to suppress gut inflammation. Thus, both probiotics and prebiotics can influence the composition of the intestinal microflora and alter the metabolic composition of the microbiome [22, 33, 34] In fact, in cases of dysbiosis, the possibility of manipulating the gut bacterial composition by using probiotic bacteria has already been explored as a promising therapeutic intervention against IBD [22]. The study [22] investigated the molecular mechanism underlying the anti-inflammatory effect of probiotic bacteria using a mouse model of colitis. The results from the study [22] showed that probiotic bacteria modulated microbial diversity of the gut and favored the production of CLA that targeted myeloid cells PPAR γ to suppress colitis. The network topology model of IBD shown in figure 2 refers to inflammatory bowel disease condition caused due to dysbiosis and highlights the complexity of the multi-network, multiscale mucosal immune responses that influences initiation, progression and outcome of the disease.
Figure 2: Network topology of model illustrating mucosal responses to inflammatory bowel disease with novel therapeutic targets in view. Systems biology markup language (SBML) compliant network of interactions between commensal and foreign bacteria on the cellular immune components is created using CellDesigner (http://www.cellDesigner.org/). The bigger panel in the figure represents the different compartments of the gut that includes the lumen, epithelium, lamina propria and mesenteric lymph node. The red and the green objects represent foreign bacteria and commensal bacteria respectively found inside the lumen of the gut. The stacked column bar graph depicts the relative abundances and distribution of the various microbial communities present inside the gut. The imbalance between the red (foreign bacteria) and green (commensal bacteria) objects represents the dysbiosis inside the lumen in inflammatory bowel disease (IBD). The dysbiosis in the lumen causes the activation of inflammatory cytokines (shown by green arrows) in the lamina propria. The three in-view of the molecules represents the modeling enabled discovery of lanthionine synthetase cyclase-like 2 (LANCL2), nod like receptor-X1 (NLRX1), and peroxisome proliferator-activated receptor gamma (PPAR γ) that are the targets for therapeutic intervention for treatment of IBD. The rectangular in-view represents the complex intracellular signaling pathways and transcriptional factors controlling T cell network [35].

1.2 Metabolism and its effect on immune system and microbiome
Multiple bacterial genomes modulate the metabolic reactions inside the body exemplified by the production of short chain fatty acids (SCFAs), an essential component of host health. Humans lack enzymes required for digestion of dietary fibers [36]. The microbial community inside the gut ferments these undigested carbohydrates for energy storage. As mentioned in the previous section, the fermentation results in a wide variety of lipid molecules including oleic acid and SCFAs [36] such as butyrate, propionate, acetate that provide the colon with energy required during metabolic demands as well as regulatory signals that help in the maintenance of homeostasis. Along with being a local nutrient source for colonocytes, SCFAs regulate energy homeostasis by stimulating lectin production in adipocytes as well as glucagon like peptide secretion by the intestinal endocrine cells. The SCFAs also regulate neutrophil function and migration, inhibit inflammatory cytokine-induced expression of vascular cell adhesion molecule-1, and increase the expression of tight junction proteins in the colon epithelia. Overall, they affect a wide range of host processes including energy utilization, host-microbe signaling, epithelial cell integrity and gut mobility [37]. Oleic acid is a commonly found dietary component and is also a microbial metabolism product. Increased concentrations of oleic acid are found within Parabacterioides [38], and oral treatment with commensal Parabacteroides distasonis has been shown to significantly reduce the severity of intestinal inflammation in murine models of acute and chronic colitis [39]. Thus, it is important to understand whether the diet-derived products of microbial metabolism are released under similar conditions in presence of varying food substrates that may include proteins, carbohydrates and fat. The host metabolome is a rich resource for studying metabolic function of the gut microbiome. Multi-omic data integration through modeling can facilitate a comprehensive mechanistic understanding of how dietary and microbial components in the gut modulate immune responses. These technologies are at the very core of advancing nutritional-based precision medicine interventions and moving from understanding single nutrients to understanding the impact of nutrition at the systems level.

### 1.3 Nutrition – a key player in the immune system-diet interaction network

The nutritional status of an individual is a key determinant of the susceptibility of the immune system to infection and disease [10, 40]. During infection, the host requirements for energy substrates and nutrients rapidly increase in the presence of invading microorganisms or in any immune mediated disease that involves proliferation of immune cell subsets. However, it is widely known that infectious agents reduce the motivation for voluntary food intake due to the stimulation of leucocytes to produce inflammatory cytokines. The immune cells use these cytokines to convey information to other physiological systems including the brain that modulates the food intake [41, 42]. The increased metabolic demands are utilized to raise the body temperature (for e.g. in fever) [43] required for the proliferation of the immune cells in the course of elimination of an infectious pathogen. The growth, survival and differentiation of the activated immune cells depends on glucose metabolism as a source of energy, which has a huge impact on our health [44]. The identification of the metabolic processes during the inflammatory processes would provide new therapeutic opportunities. The study of the T cell metabolism has provided ample resources regarding the pathways important for the T cell plasticity and effector functions [44]. The metabolic demand of every immune cell depends on the particular function it performs, this is evident in the subsets of CD4+ T cells where effector and Th17 cells rely on aerobic glycolysis while memory T cells and regulatory T cells rely on fatty acid oxidation to produce energy [44]. The other activated immune cells such as dendritic cells, neutrophils, and pro-inflammatory macrophages rely on aerobic glycolysis for energy. During activation T cells increase their glucose
uptake through Glut1, that facilitates increased oxidative phosphorylation and glycolysis to sustain proliferation of these cells [44]. The two main biochemical pathways that lead to generation of adenosine triphosphate (ATP) and the metabolic precursors for biosynthesis of immune cells includes glycolysis and tricarboxylic acid (TCA) cycle [45]. In proliferating cells, glucose is broken down to pyruvate by glycolysis, which is further oxidized by the TCA cycle in mitochondria [45]. The study by Michalak et al. [46] determined that pro-inflammatory cells displayed a stronger bias towards glycolysis whereas the induced regulatory cells displayed mixed metabolism including glycolysis and lipid oxidation. Since, nutrients affect metabolic changes, which in turn affect the differentiation state of the immune cells, dietary interventions could be used to cause metabolic changes in a response to infection.

Malnutrition is an important example of immunosuppression caused due to macronutrient and micronutrient deficiencies in our immune system [7]. It predisposes individuals to infection by impairing the integrity of epithelial cell barrier and suppressing the immune responses [47]. Enteroaggregative *E. coli* (EAEC) infections causes diarrhea like symptoms in immune-compromised individuals and particularly in severe cases in children with malnutrition. A study by Philipson et al., 2013 [48] demonstrated that malnourished mice exhibited an impaired ability to induce proinflammatory cytokine during the EAEC infection. The observed immunodeficiency of the mice demonstrated that the malnourished mice were unable to mount protective innate or adaptive immune responses against EAEC infection [48]. Another study by Philipson et al., 2014 [49] showed that tryptophan is a crucial element for antibacterial protection against infection. Mice fed with tryptophan free diet had reduced antimicrobial peptide production against the high EAEC pathogen levels. A study by Bolick et al. [50] demonstrated that zinc deficiency impaired the immune responses in response to EAEC by increasing the virulence factor associated with it. The zinc-deficient mice challenged with EAEC had greater weight loss, mucus production and diarrhea compared to the control group. The nutritional supplements and interventions, such as vitamins and mineral supplements, polyunsaturated fatty acids have been studied extensively over the past decade [51]. Additional breakthrough studies include the association between vitamin E and T cells, vitamin A and mucosal immunity, role of zinc in T and B cell development and the effect of polyunsaturated fatty acids composition of the diet on inflammation and immunity. A study by Meydani et al., 2008 [51] demonstrated the reversing effect of vitamin E on age-associated defect in T cells. Vitamin E enhances the T cells via a direct effect on T cells and an indirect effect by reduced production of PGE2 in macrophages [51]. Furthermore, several studies have shown that retinoic acid (RA), a major oxidative metabolite of vitamin A, plays a key role in the differentiation of T cell subsets, migration of T cells into tissues and their regulatory function [52] that provides further evidence for the role of vitamin A in mucosal immunity. Adequate vitamin A status in animal models, whether derived from ingestion of preformed retinol or β-carotene, is important for maintenance of the proper balance of well-regulated T cell functions and prevention of excessive or prolonged inflammatory reactions. In addition, zinc deficiency [53] has been shown to be partially responsible for increased apoptosis of pre-T cells; and crucial for the balance between the different T cell subsets. Accordingly, zinc supplementation restores the Th1/Th2 balance; however, high dose of Zn<sup>2+</sup> reduces the development of Th17 cells. Furthermore, zinc deficiency is known to cause the reprogramming of immune system that accelerates apoptosis among premature and immature B cells, and causes decreased antibody production due to the chronic production of glucocorticoids [54]. Another, important component includes the proportion of different types of polyunsaturated fatty acids (PUFAs) present in the diet and its effect on immune cell functions. The dietary n-3 PUFAs present in fish oil modulate immune responses and
the expression of transcription factors involved in controlling inflammation [55-57]. Dietary n-3 PUFAs also aid in the suppression of pro-inflammatory cytokines produced by the macrophages and reduce the symptoms of animal models of autoimmune disease [24].

These studies show the effect of various dietary components on the immune system. However, a global mechanistic understanding of the interplay between infection, microbiome, metabolism, and nutrition is currently lacking.

The direct alteration of mucosal communities by the nutritional interventions has led to the evolution of nutritional immunology, leading to advancement in the field of medicine. One such ‘use case’ for the effective use of nutritional based intervention is the use of conjugated linoleic acid (CLA) in the treatment of immune mediated inflammatory disorder of the gastrointestinal tract such as Crohn’s disease. CLA is a mixture of positional and geometric isomers of octadecadienoic acid. The use of CLA has been explored due to numerous anti-inflammatory and anti-oxidant properties that have been characterized in animal models [58-60]. Dietary CLA supplementation has been shown to suppress colonic inflammation in pigs with bacterial-induced colitis [60] by the up regulation of the colonic peroxisome proliferative activated receptors (PPARs) expression. CLA decreased the disease severity of experimental IBD in pigs by activating colonic PPAR Y [60]. Another mechanistic theory proposed to explain the benefits of dietary CLA includes inducible eicosanoid suppression in the endoplasmic reticulum. CLA has also been shown to ameliorate inflammation driven colorectal cancer in mice [61] and has enhanced cellular immunity by modulation of the effector function of CD8+ T cells and antiviral responses in pig models [62, 63]. It is a unique compound known to exhibit anti-inflammatory effects along with stimulating cellular and adaptive immune responses to bacterial and viral infections.

The immunomodulatory efficacy of CLA was tested in patients with mild to moderate Crohn’s disease (CD) in an open-label study for 12 weeks [64]. Oral CLA administration was well tolerated in these patients, and CLA suppressed the ability of the peripheral blood T cells to produce pro-inflammatory cytokines such as interferon gamma (IFN-γ), tumor necrosis factor – alpha (TNF-α), and IL-17. The study demonstrated decreased CD activity index and increased quality of life of patients with CD [64]. It also provided insights on possible mechanisms of immune modulation by CLA, a nutritional intervention targeting the human system [64]. The patient level data obtained from the clinical study was used as a training dataset to develop a larger synthetic population for in silico experimentation of the Phase III placebo-controlled, randomized clinical trial [65]. The study [65] demonstrated that post-treatment highlighted a positive correlation between the initial disease activity score and the drop in Crohn’s disease activity index (CDAI) score. It highlighted the need for precision medicine strategies for IBD treatment, wherein treatments specific to an individual would yield better outcome as opposed to the one size fit all strategy.

Another ‘use case’ for nutritional immunology research is abscisic acid (ABA), a plant phytohormone, which when used as a dietary component elicits immunomodulatory properties. A benefit of dietary ABA-supplementation in mice includes anti-diabetic effects, anti-atherosclerotic and an anti-hypersensitive effect that has been shown in various studies [66-68]. The study by Guri et al., 2007 [66] showed that ABA improved insulin sensitivity and reduced adipose tissue inflammation when supplemented into diets of obese mice. Another study by Guri et al., 2010 [67] showed that mice treated with 100mg/kg of racemic ABA mixture significantly reduced recruitment of CD4+ T cells in the aortic root [69]. ABA has also been identified as a ligand of
lanthionine synthetase C-like 2 (LANCL2), a novel therapeutic target. A study by Hontecillas et al., 2013 [70] investigated the immune modulatory mechanisms underlying the anti-inflammatory efficacy of ABA against influenza-associated pulmonary inflammation. When ABA was given preventively or therapeutically, it ameliorated the influenza virus-induced pathology by the activation of peroxisome proliferator-activated receptor γ (PPAR γ) in pulmonary immune cells, along with suppression in the initial proinflammatory responses and promoted resolution of the infection. A recent study by Magnone et al., 2015 [71] showed that the mechanism by which low dose of ABA (found in fruit extracts or exogenous) lowers the blood glucose level does not involve insulin release at all. They showed that ABA had a lowering effect on glycemia without having an effect on insulin concentration in the blood. The study focused on finding the bioavailability of dietary ABA mainly the one found in fruits (apricots primarily used in the study) and the effect of these fruits in general on glucose tolerance. The rats and human fed with fruits extract (with ABA), when compared to the control group had lower glycemia and insulinemia. When a dose ABA was administered orally without fruit, an equivalent dose of ~1ug/kg [71] successfully lowered glycemia and insulinemia during the oral glucose tolerance test. The mean glycemia with the fruit extract was significantly lower than the exogenous ABA. The lowering effect of ABA on glycemia lasted for at least 6h after intake [71], showing that it contributed towards disposal of glucose in the blood. The results also showed that apricot extracts increased ABAP (ABA plasma levels) higher than glucose did, which led them to the conclusion that high bioavailability of oral ABA can be obtained from the fruit extracts. The mechanisms by which this plant hormone and secondary by-product of soil fungal metabolism regulate glucose metabolism and immune responses in humans remain largely unknown.

The research on the role of single nutrients in immune functions is extensive however this is not the case for multiple nutrients and the existing combinatorial effect of interactions between the various nutrients remains largely unknown. The interactions between multiple nutrients can negatively affect the immune system, for example excess of calcium interferes with leukocyte function by displacing magnesium ions, causing reduction in cell adhesion processes [72]. The nutrient deficiencies can either singly or combinatorial affect the host immune system in multiple ways. The regulation of the immune system by the nutrients can either be beneficial or detrimental. For example, the nutrients involved in antimicrobial and antitumoral function of macrophages can be modified by nutrients that promote synthesis of reactive oxygen or nitrogen intermediates [72]. A recent study by Lacroix S et al., 2015 [73] showed how systems biology methods can be applied to better understand the potential role of nutritional interventions such as caloric restriction and polyphenol supplementation to promote health aging processes and reduce metabolic risk factors. Thus, although a comprehensive level understanding of the complex mechanisms underlying the combinatorial effect of nutrients is challenging, a systems-wide approach integrated with computational modeling and informatics can aid in elucidating this complex process.

2. The current approaches and the urgent need for paradigm shift

2.1. Understanding reductionist approaches towards nutritional immunology

Traditional reductionist nutritional immunology approaches have prevailed in the field and focused on studying the interplay between nutritional deficiency or supplementation and their effects on specific parts of the system while disregarding global effects. Until recently, researchers have only been able to extrapolate data that involved a subset of nutrients and their gene interactions, along
with the key pathways of the immune system. A comprehensive systems-wide understanding of any biological system requires the harnessing of data that includes genes, proteins, RNAs, their interactions, changes in concentration, and regulation under certain conditions [74]. Traditional approaches are based on reductionist methods alone, which do not take into account that systems are a part of greater networks of interacting entities, including genes and nutrients. However, with the advent of fields like Nutrigenomics [75] and Nutrigenetics the field is slowly advancing towards using the tools initially developed for genetics research. However, when analyzing massively interacting systems such as the relationship between nutrients, microbiome, metabolism and immune response, there is a need for computational modeling techniques [76]. Nutrigenomics and Nutrigenetics refer to the interface between nutritional environment and their interaction with cellular and genetics approaches. The development of novel sequencing tools in these fields of nutritional science focuses in determining the overall effect of nutrition on the human genome and the modulation of several molecular mechanisms that affect different physiological functions inside a human body. The advancements in genomics have resulted in incremental knowledge discovery which takes into consideration: how an individual’s genome expresses itself at different omic levels (proteomics, metabolomics, lipidomics) in response to nutrition. An effort towards post-genomics data, and multi-omic data integration by using modeling provides a deeper insight of the interaction between our genes, microbiome, and diet. Metabolomics is one of such -omics technology that involves the study of small molecules or metabolites present in the biological samples in order to study the effects on metabolic process under varying biological conditions [77]. The study of metabolites yields information about the biological processes since metabolites are implicated in number of human diseases [77]. The application of metabolomics in nutritional immunology would include detailed study of alterations caused in metabolic pathway following nutritional interventions. This will allow enhanced understanding of the effect of nutrition on metabolic pathways. The study by Hendriks et al. [78] is an example where an integrated metabolomics approach was used to study effects of dietary products that showed anti-inflammatory properties, in a population of overweight men. The profiles of gene expression, proteins and metabolites were integrated with the measures of inflammation markers and the results obtained after integrated omics approach, demonstrated that the dietary products modulated inflammation and oxidation with alteration in the metabolism status of the healthy overweight men. A more transformative approach that would include information-processing representations of nutritional immunology is required to tackle the challenges in this field. This would involve using interdisciplinary approaches from computer science, systems modeling, bioinformatics and data science for big data analysis, which would allow researchers to reverse-engineer the system. Thus, the application of systems biology methods in nutritional immunology research has the potential to accelerate the discovery of novel biomarkers and systems level mechanistic understanding of how nutrition modulates our immune system and health outcomes. One key step in this iterative process is the validation of modeling-derived predictions that require targeted pre-clinical, mechanistic or clinical studies. This step represents the confluence between systems-level analyses and the need for reductionist validation studies.

2.2. Systems nutritional immunology: A systems-level approach to nutrition-microbiota-immune system interactions
Figure 3: Integrated information biology methods applied to Nutritional Immunology

The Modeling Immunity to Enteric Pathogens project (MIEP) and the Nutritional Immunology and Molecular Medicine Laboratory are examples of successful implementation of modeling approaches for the study of complex mucosal immune responses in the context of infectious diseases. Under the MIEP project, a first step toward building information processing representations of the mucosal immune system was undertaken. However, similar initiatives are lacking in the field of nutritional immunology or for chronic and autoimmune diseases. Computational modeling in combination with big data analytics, portal science and informatics, enabled by high-performance computing [79-81] are essential components in the study of massively interacting systems such as host immune response-gut microbiota-nutritional interactions. As proposed in Goals in Nutrition science 2015-2020 [76] a mechanistic understanding of the host-nutrient microbiota interactions enabled through computational modeling based on integrated information biology methods have an enormous potential to predict the outcomes of the nutrient-microbiota-immune system interactions.
The main challenges in systems biology frameworks are the complexity of systems and the output in terms of vast amount of data with scattered pieces of knowledge that needs to be connected together and be made sense of. The need for development of computational tools becomes imperative for the integration of the data [82]. The advent of user-friendly tools for informatics, modeling and advanced big data analytics enable the prediction of emerging global behaviors of biological systems and the characterization of novel molecular and cellular mechanisms [82].

Over the past decade, a significant increase in computational power and availability of larger experimental datasets, has allowed models to be more comprehensive and in some cases multi-scaled [80, 83-86]. In addition, development of software workflow such as Epidemic Simulation Systems (EPISIM) [87] has facilitated the semantic model integration for biologically skilled scientists especially with the growing number of available models (over 163 nutrition themed systems biology markup language models (SBML) that are already available in the Biomodels database [88]. The methodologies to extract actionable knowledge from such rich data and metadata, has been facilitated by the development of standards (such as systems biology markup language models (SBML) [89] and more recently Markup language for Allergens (AllerML) [90]), ontologies [91] (for example: Medical Subject Headings (MeSH) [92], Unified Medical Language System (UMLS) [93] and Gene Ontology (GO)) and curated specialized databases such as Therapeutic Target Database (TTD) [94], hepatotoxicity database [95], drug combination database [96], Food and Drug administration (FDA) toxicity databases, real-time data entry [97] and White adipose tissue reference network (WATRefNet) [98]. New bioinformation technologies combine relevant models and data to address important questions whereby the answers reside in the interface between networks.

2.2.1 Computational tools

Computational tools and big data help revolutionize biological research in a way that is shifting the paradigm from top-down or bottom-up approaches to a middle-out approach. The latter is based on conceptualizing models and tools at the level that provides the richest data and connection that to higher or lower levels for comprehensive integrated systems. Building massively interacting multiscale models (theoretical and data-driven) anchored around unanswered immunological questions holds a promise for the advancement of the field of nutritional immunology into unprecedented scientific discoveries. A recent work by Ramsundar et al. [99] investigated the aspects of multitask learning with an application to virtual screening. The multitask networks trained on 40 million experimental measurements for more than 200 targets showed significant improvements in comparison to the basic machine learning methods [99]. Their findings show that the amount of data and tasks both had an impact on the outputs. The efficacy of multitask learning was directly correlated to the available relevant data, which emphasized the concept that bigger data are of critical importance for improved predictions [99]. Furthermore, metabolomics studies can be integrated, through rigorous methods, with biochemical, metabolism and immunological networks in order to form more comprehensive pictures regarding the complex systems level interaction. Immunometabolism data includes changes in metabolite composition and immunological parameters that can aid in studying the interaction between nutrients, immune system, microbiome that changes during the progression of a disease. Such data and theory can be used to build computational models with an inclusion of the interaction effects of nutrients, such that the model can be calibrated with large time series multi-omics datasets. The outputs obtained from the models can be integrated with the experimental studies along with inclusion of molecular
modeling techniques such as molecular docking approaches. The docking studies can determine how nutrients modulate the various metabolic and immunological networks which can be experimentally validated with surface plasmon resonance (SPR) spectroscopy. For example, a better understanding of the changes in metabolites caused during the varying nutritional demands of the immune cells which includes - the reliance of effector and Th17 cells on glucose and memory T cells and regulatory T cells on fatty acid oxidation can elucidate 1) how a set of nutrients modulate biochemical pathways and immune responses of specific cell types and 2) mechanisms underlying the nutritional prevention or amelioration of disease. Systems nutritional immunology can be built with the concept that analyses of multitask networks across nutrition, metabolism, microbiome and immune system and are required to elucidate emerging mechanistic behaviors that inform health-promoting interventions. In addition to deterministic models, machine learning methods can also be highly effective in bridging the gap between big data and knowledge. For instance, it has been shown that supervised learning methods such as Artificial Neural Network (ANN) or Random Forest (RF) can be an alternative solution to the ordinary differential equation (ODE)-based modeling [100-103], and can more efficiently be used to model complex systems, such as the CD4+ T Cell differentiation [100]. Unsupervised method can also be valuable in knowledge discovery as they allow deeper analysis of large datasets and can be instrumental in developing mechanistic models as well candidate gene prioritization, and overall understanding of the complex intertwined systems [104, 105].

2.2.2 Systems biology tools: contribution of agent based modeling and multiscale modeling

With the continuous generation of massive amount of data, there is an urgent need to integrate big data, theory, procedural knowledge and mechanistic information to synthesize and simulate recognizable behaviors of massively interacting systems. Mathematical modeling and simulation are the techniques that can be utilized for dynamic knowledge testing. Models have the power to discover new findings through effective computational technologies. The knowledge acquired from the computer simulations can form a formal basis of testing the finding in the lab, and validate the known findings [106]. The two major categories of modeling technologies include equation based and agent based modeling, with a limited number of equations, mathematical models provide an efficient solution; however, it is challenging to incorporate the biological processes in the mathematical equation. Agent-based modeling (ABM) on the other hand uses agents to represent the units in the biological processes [106] wherein these agents follow certain rules and have unique properties that represent different states of the biological entities such as their location, genotype, and movement. The enhanced capability of ABMs can simulate extremely complex biological behaviors for which the requirement of high performance computing is a must [108]. Traditional ordinary differential equation (ODE) methods and ABMs provide useful information. However, analysis of the complex nutrient-immune system–microbiome interactions and study of how these interactions change over time, requires an understanding of all the key components at varying scales in space that include molecular to tissue level, population level scales and time from nanoseconds to year [109]. This necessitates computational modeling across spatiotemporal scales. The advancement in computer hardware, algorithms as well as computational power have contributed to the development of multi-scale models (MSM).

The Modeling Immunity of Enteric Pathogens (MIEP) has developed the Enteric immunity simulator multi-sale modeling (ENISI MSM) [80, 86], the first agent based simulator for enteric immune systems. ENISI MSM integrates five orders of spatiotemporal scales and is based on both
deterministic (ordinary differential equations and partial differential equations) and agent based models, integrated in a single unit. It is designed specifically for application in computational immunology along with strong visualization module for the representation of the tissue level scale in the MSM system. ENISI MSM allows the combination of different tools, techniques, modeling strategies thereby integrating diverse types of data across different scales along with sensitivity analysis in order to validate the model driven hypothesis with experimental data. With respect to components, the ENISI MSM model can stimulate signaling pathways, metabolic networks, cytokine diffusions, cell movement, and tissue modeling [107]. A multi-scale CD4+ T cell differentiation model when calibrated with experimental data and tested in the context of gut inflammatory was able to produce in silico experimentation that was used to study the complex host-pathogen interactions as well as host-nutrient-microbiota actions [110]. ENISI Visual [108], provides a user-friendly interface for users to change the number of immune cells and observe simulation speed. The MSM tools can be utilized in the field of systems nutritional immunology, wherein the effect of nutritional components on the immune cell parameters can be modeled and modified accordingly. ENISI MSM is a tool designed for modeling the mucosal immune responses that can simulate 10^7-10^{11} cells in high performance simulations [107, 108, 111-115]. The high-performance computing driven ENISI MSM enhanced the development of massively interacting models of the mucosal immune system and significantly increased the power of in silico experimentation with a scalability of 10^9-10^{12} [110]. While the tool was initially developed to address problems related to infectious and immune-mediated diseases, ENISI can be adapted to develop new information processing representations of host-microbiota-nutrient massive interactions.

2.2.3 In silico techniques –a nutritional immunology revolution

Traditional tools such as in vivo and in vitro models have been consistently used in order to test hypothesis and perform quantitative studies. However, traditional reductionist experiments have led incremental knowledge generation due to the abundance of reductionist approaches. Over the past decade, new computational techniques such as in silico methods [116] have been applied to address the failures in trials for Alzheimer’s disease [117] and the clinical trial related to trauma-induced critical illness [118]. The in silico methods are based on quantitative relationship between the parameters, and include homology modeling, machine learning, data mining, network analysis tools and data analysis tools that require high computational power and capabilities [116]. For instance, in silico pharmacology is a rapidly growing new field that incorporates the newly developed techniques in order to integrate patient clinical data. It involves the development of computational models based on certain algorithms to make predictions, propose new hypotheses and advance towards new horizons in medicine and therapeutics. In silico clinical trials provide an opportunity to develop synthetic population and conduct large-scale clinical trial simulation thereby aiding in the design and testing of new nutritional components. The application of in silico methods has also been utilized in the complex process of drug discovery. The review by Ekins et al., 2007 [116] describes various in silico methods for pharmacology that are being utilized in the drug discovery process. The process of ‘virtual screening’ involves scoring and ranking the molecules in large chemical libraries according to their strength of affinity to a certain target [116]. Thus, the valuable information provided by the power of in silico methods can be extended beyond the fields of immunology & pharmacology and be applied to systems nutritional immunology in order to predict the outcome of dietary interventions on the human health.
The concept of *in silico* trials provides insights and guidance into the design of clinical trials of immunomodulatory therapies, especially the ones that have severe side effects. The process ranges from optimal patient selection to individualized dosage and duration of proposed nutritional / therapeutic intervention [119]. Machine learning algorithms or ABMs can be utilized to create synthetic patients from existing clinical trials [65].

**Case study**

One of the ‘use cases’ that explains the success of *in silico* experimentation is the identification of lanthionine synthetase component cyclase-like 2 protein (LANCL2) and its application as a treatment option for the *in silico* clinical trial. MIEP performed series of modeling studies that included computational-based drug design methods [120], biochemical and *in vivo* studies [121] to confirm LANCL2 as a novel and promising target for the discovery and development of orally active, broad based drugs against inflammatory, immune mediated and chronic metabolic disease. The *in vivo* studies comprised of using LANCL2 ligands as a treatment option on human peripheral blood mononuclear cells that showed a significant drop in the inflammatory and pro-inflammatory cytokine levels [121]. The results were validated in mouse models of IBD in which LANCL2 ligands were used as an alternative treatment for Crohn’s disease (CD) patients [121]. Based on the preliminary results and experimental data, MIEP used advanced machine learning algorithms to create a large synthetic population of CD patients and designed a Phase III clinical *in silico* clinical trial study [65]. The synthetic patients were randomly allocated to different treatments under the study and the effectiveness of these treatments were analyzed based on the changes in Crohn’s disease activity index (CDAI) scores [65]. The results from the study provided an insight that the efficacy of LANCL2 therapeutics can be extendable to all stages of Crohn’s disease.

Similar approaches can be designed and extended to gain new insights of the interactions between diet, genetic factors, microbiome populations, and response to treatment to precision medicine intervention. The effect of treating the set of synthetic population with biologics, therapeutics, pharmaceuticals, nutritional components or combinatorial interventions (i.e. nutritional adjuncts along with therapeutics) can be examined. The output can provide valuable data for accelerating drug development pipelines with improved capability to predict the likely response to any treatment [119] [65].

**3. Challenges in Computational capabilities**

Despite the fact that, *in silico* experimentation and modeling have the capability to yield basic insights and translational applications in critical illnesses, many challenges still remain in this rapidly evolving field [20]. The key challenges involved in modeling the biologically complex systems are that they encompass many levels of systems and scales. The inherently multi-scale, multi-system, multi-network nature of critical illness adds on to this complexity. The challenge also comprises the integration of nutrition, immunological, metabolic and physiologic processes required to decipher the multi-compartment and multi-dimensional landscape of biological systems [20]. Furthermore, the process of building mathematical models of complex biological processes and their computer simulation is an iterative one. The initial information processing representation of the *in vivo* counterparts would incorporate genomic, biochemical, microbial, immunologic and physiological data [122], that would eventually require the data calibration from experimental counterparts. The major challenges that would arise include explosion of massive
time-series biological data, with increasing demand for storage of the larger datasets and computational power. This would require the development of new tools and techniques for the management of knowledge from the biological datasets. Furthermore, given that analysis and visualization of the massive data is complicated, the importance of open science becomes increasingly important [123]. Data availability would enhance the progression of data analysis procedures. It is important to understand the power of big data and the analysis process, as it adds knowledge to the existing hypothesis [123]. However, the high dimensionality of big data puts forth computational and statistical challenges that include scalability, storage bottleneck, computational cost and algorithm instability [124]. In order to extract the knowledge and exploit parallelization of computation, skilled programmers and bioinformatician need to adopt to the new programming platforms, tools and practices such as in memory processing, graph databases, and advanced machine learning algorithms [125]. Along with modeling and informatics tools, a web portal that facilitates the collection and integration of experimental and computational data and metadata along with analysis processes, model building and quality assurance are needed to support precision medicine interventions. Therefore, it is challenging to not only overcome the hardware bottlenecks and software complexity, but also to keep up with the ever-changing technological advancements and the need for seamless integration.

The multiscale models represent different spatiotemporal scales with distinct spatiotemporal properties. This increases the need to improve the computational performance and synchronization across scales. In a hybrid approach, for example ENISI MSM [107] calls the sub-models in the different scales in each simulation cycle, and the final output can be the integration of the outputs form each scale. Since, the ODE solver Complex Pathway Simulator (COPASI) [126, 127] used in ENISI MSM is a large object, loading millions of objects with different scales significantly slows the simulations, due to high memory processing activities. The implementation of the system with the use of high computational power for model simulations can aid the analysis of increased number of realistic number of agents required for in silico studies. Also, there is a need for improvement in enhancement of the visualization components of the models that can aid the adaptability of the system amongst experimentalists. The solutions designed to address the challenges related to the tools used in computational immunology can be extended to develop bio-information systems, models, web portals, and tools adaptable to systems nutritional immunology research.

4. Developing information processing representations of Systems Nutritional Immunology

The Modeling Immunity of Enteric Pathogens project (www.modelingimmunity.org) has successfully developed user-friendly tools and models to characterize the mechanisms of immunoregulation underlying immune responses to enteric pathogens. The MIEP technology is high performance computing (HPC)-driven as illustrated by ENISI MSMv2, a tool that models mucosal immune response and scales up to $10^{11}$ agents in HPC simulations [110]. This is an important hallmark achieved towards building large-scale information processing representation of immune response at multiple levels. Overall, the HPC-driven ENISI MSM platform combines the study of molecular pathways controlling T cell differentiation and tissue level interactions between cells with an aim to characterize novel mechanisms of immunoregulation at the gut mucosa. MIEP is also working towards the integration of bioinformatics, computational modeling and experimental validation in order to study the mechanisms of tolerance during bacterial and
viral infection. These modeling-driven predictions have the potential to accelerate the scientific discovery process.

Notable MIEP-based achievements include: (1) development and enhancement of a suite of tools for ABM (ENISI) and ODE (COPASI) based modeling of immunological processes [81, 128-130]; COPASI supports models in the SBML standard and can import and export models in the SBML format. The software allows to perform simulations either with stochastic kinetics or with differential equations and can be used to perform, analysis, sensitivity analysis and user-friendly data visualizations. (2) development of validated computational models of CD4+ T cell differentiation and function [131-133], mucosal immune responses to H. pylori [134], modulation of CD4+ T cell responses to H. pylori by IL-21 [133]; (3) development of mouse and pig models of H. pylori infection [135, 136] and mouse models of EAEC infection [48] [137, 138]; (4) development of a methodology allowing the creation of dynamic models combining theoretical knowledge and time-course high dimensional datasets; (5) initial characterization of the ability of H. pylori to induce CD8+ T cell responses in the pig model; (6) determination of the role of H. pylori-infected mononuclear phagocytes on the modulation of mucosal immune responses to the bacterium, and (7) successful delivery of a summer school and symposium in Computational Immunology [139].

The web portal resources that can be used include the “Immune Modeling Community Web Portal” where news and relevant resources and news related to immune response modeling for infectious diseases are shared in order to facilitate collaboration and exchange of information between the investigators. Another workflow based modeling software designed for the use in immunology research is the Differential Equation Modeling Solution (DEDiscover) developed at the Center for Biodefense Immune Modeling. The software can be utilized to perform simulations, parameter estimation, sensitivity analysis, residual analysis and statistical analysis for a case study represented as a set of differential equations. However, DEDiscover does not fully support SBML and it cannot handle user-defined kinetics on import. Furthermore, it does not export SBML. These shortcomings make COPASI the preferred tool. The Program for Research on Immune Modeling and Experimentation (PRiME) is another multidisciplinary Immune Modeling Center that is focused on developing (1) mathematical and data based models for elucidating the viral mechanisms of category A-C viral pathogens and (2) bioinformatics components for data management and model development. The ImmuNet [140], developed by PRiME is a web interface that is aimed to provide immunology researchers with an easy to use resource that can be used to explore the immune related functional relationship networks. These functional relationship networks provide new mechanistic insights about the previously unknown gene-gene relations and can be used to predict the immune processes associated with any other pathway relevant components.

To date, most of the nutritional immunology studies have focused on characterizing the effect of nutritional components on individual parts of the immune system, whereas limited effort has been placed on elucidating and modeling the complex, interconnected massively interacting systems. Computational models can be trained using the known effects of the nutrients (data and theory) and the effect of interaction of different nutrients on the massively interacting system can be predicted using advanced machine-learning methods. These predictions can be tested in lab and the deviations can be used to update the models. A key aspect of the nutritional systems immunology cycle is the validation step. That step is inherently reductionist. However, if
performed in the broader context of systems nutritional immunology, then the validation studies are guided by the existing theory and data. Another way would include the use of multiscale models wherein the knowledge obtained from multi-omics studies regarding the regulatory mechanism of nutrition, can be studied with high levels of details ranging from the cellular level to whole body, population and policy level. The data obtained from the nutritional intervention studies integrated with –omics and targeted modeling-driven mechanistic studies will provide a comprehensive framework to simulate the physiological mechanisms and immunological changes in the body after the intake of nutrients.

A major goal in the advancement of systems nutritional immunology research and development is to build comprehensive, multiscale network models that will accurately predict global and local effects of nutrition-based interventions. The identification of the efficacy of these nutritional based interventions on the immune system using in silico experiments would lead to advancements in research for better treatment for disease mechanisms, assessment of disease risk, and prediction of optimal interventions for the immune mediated diseases, with an ultimate future goal of expanding the outputs for application in precision medicine. In this review, we specifically highlight the pressing need for the development of predictive systems modeling that provide a comprehensive mechanistic understanding of the system. The conceptual modeling approaches and the computational techniques need to be integrated with advanced big data analytics methods such as statistical and machine learning algorithms. In summary, the use of an iterative systems biology cycle of experimental simulation, data collection, along with mathematical and computational model building, simulations, prediction, calibration, refinement and validation have the potential to gain a systems level mechanistic understanding in order to guide nutrition-based precision medicine, health and wellness.

Author Contributions
Contributed to the design of the paper: JB-R, RH, and VA. Contributed to the writing and reviewing of the manuscript: MV, JB-R, VA, RH, AC, CP, AL, and NT-J. Contributed to the making of the figures: MV, AL, and NT-J.

Conflict of Interest Statement
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding
Funding provided by Nutritional Immunology and Molecular Medicine Laboratory.
Reference:


44. Philipson, C.W., et al., The role of peroxisome proliferator-activated receptor gamma in immune responses to enteroaggregative Escherichia coli infection. PloS one, 2013. 8(2).
74. Information, N.C.F.B., Medical Subject Headings.
Chapter 3: High-Resolution Computational Modeling of Immune Responses in the Gut

This work was published in *GigaScience*


**Abstract**

*Helicobacter pylori* causes gastric cancer in 1–2% of cases but is also beneficial for protection against allergies and gastroesophageal diseases. An estimated 85% of *H. pylori*–colonized individuals experience no detrimental effects. To study the mechanisms promoting host tolerance to the bacterium in the gastrointestinal mucosa and systemic regulatory effects, we investigated the dynamics of immunoregulatory mechanisms triggered by *H. pylori* using a high-performance computing–driven ENteric Immunity SImulator multiscale model. Immune responses were simulated by integrating an agent-based model, ordinary, and partial differential equations. The outputs were analyzed using 2 sequential stages: the first used a partial rank correlation coefficient regression–based and the second a metamodel-based global sensitivity analysis. The influential parameters screened from the first stage were selected to be varied for the second stage. The outputs from both stages were combined as a training dataset to build a spatiotemporal metamodel. The Sobol indices measured time-varying impact of input parameters during initiation, peak, and chronic phases of infection. The study identified epithelial cell proliferation and epithelial cell death as key parameters that control infection outcomes. *In silico* validation showed that colonization with *H. pylori* decreased with a decrease in epithelial cell proliferation, which was linked to regulatory macrophages and tolerogenic dendritic cells. The hybrid model of *H. pylori* infection identified epithelial cell proliferation as a key factor for successful colonization of the gastric niche and highlighted the role of tolerogenic dendritic cells and regulatory macrophages in modulating the host responses and shaping infection outcomes.

**Introduction**

Computational modeling of the immune response dynamics can provide novel insights and facilitate the systems level understanding of the interactions at the gastric mucosa during infection. Ordinary differential equation (ODE- based methods are deterministic and based on the average response of cells over time. Dynamical models are used in immunology for system-level analyses of CD4+ T cell differentiation [1], macrophage differentiation [2], immune responses elicited by *Clostridium difficile* infection [3], co-infections [4], and in cancer and immunotherapy [5]. However, ODE-based models lack the spatial aspects and the features to study the organ and immune cell topology
over time. Agent-based models (ABM) employ a bottom-up approach that focuses on the spatial and temporal aspects of individual immune cells, unlike the ODE-based methods. This rule-based method includes agents that act as local entities which interact locally with other agents, move in space, and follow set of rules representing their role in a given system and contribute towards generating an emergent behavior. Since, the immune system is a complex dynamical system [6] wherein the components i.e., the immune cells move in space and time changing their location, ABMs are useful tools that can be employed to understand biological mechanisms and the hidden insights.

*Helicobacter pylori* is a gram-negative bacterium that has persistently colonized the human stomach since early evolution [7] [8] and is currently found in over 50% [9] of the global population. *H. pylori* has co-evolved with humans for thousands of years, such that an estimated 85% of the *H. pylori*-colonized individuals, do not present any detrimental effects. Thus, the vast majority of carriers (i.e., up to 75%) remain asymptomatic, while only 15% develop ulcers, and less than 3% develop cancer. Further, growing and sometimes contradictory evidence from recent experimental, clinical studies and epidemiological studies suggest that *H. pylori* might provide protection against obesity-related inflammation and type 2 diabetes [10], esophageal, cardiac pathologies, childhood asthma and allergies [11] and autoimmune diseases. In this context, it is crucial to understand the mechanisms that promote host tolerance to the bacterium in the gastrointestinal mucosa and its systemic regulatory effects since these have been linked to the beneficial commensal aspects of *H. pylori*-human host interaction. Computational models provide a cost-effective and predictive way to study the complex and dynamic immune system interactions and form a non-intuitive novel hypothesis. Solving the complex puzzle of immunoregulatory mechanisms that include large spatiotemporal scales ranging from cellular, intracellular, tissue and organ level scales is a major unsolved challenge that requires applying computational modeling and data analytics.

An advanced hybrid model used to study the mucosal immune response during gut inflammation highlighted the mechanisms by which effector CD4+ T cell responses, contributing to tissue damage in the gut mucosa following immune dysregulation [12]. Other hybrid models with the integration of ABM, ODE, and PDE technologies, were developed to understand the dynamics of tumor development [13] and tumor growth models [14]. These combined techniques have been used to develop multi-organ models in various situations, including the study of granuloma formation [15] and pressure-driven ulcer formation in post spinal cord injury patients [16]. The summary of different agent-based simulators with immunology related applications are discussed and summarized in [17, 18]. The comparison between different multiscale modeling tools and agent-based immune simulators, are discussed in [12, 19].

In this study, we utilize a high-resolution ENteric Immunity SIMulator (ENISI)-based model of the stomach for simulating the mucosal immune responses to *H pylori* infection. The advanced hybrid multiscale modeling platform ENISI multiscale model (MSM) is capable of scaling up to $10^{12}$ agents [20]. The host immune responses initiated during *H. pylori* infection and the underlying immunoregulatory mechanisms are captured using the ENISI multiscale hybrid model. The underlying intracellular mechanisms that control cytokine production, signaling and differentiation
of macrophages and T cells are modeled by using ODEs, the diffusion of cytokine values is modeled using PDEs and the location and interactions among the immune cells, bacteria and epithelial cells are modeled by using ABMs. The hybrid model thereby represents a high-performance computing (HPC)-driven large-scale simulation of the massively interacting cells and molecules in the immune system, integrating multiple modeling technologies from molecules to systems across multiple spatiotemporal scales.

To understand the dynamics and emergent immunological patterns described by this hybrid model, we employed sensitivity analysis (SA), an important part of the model analysis used to explore the influence of varying model parameters on the simulation outputs. The influence of the effects of changes in parameter values on the model output explains the model dynamics that underlay the outputs [21, 22]. Furthermore, SA examines the robustness of the model output at a different range of parameter values that correspond to a range of different assumptions. We employed global SA and conducted a two-stage spatiotemporal global SA approach. First, we used a regression-based method such as the partial rank correlation coefficient (PRCC) and screened the important input parameters that were shown to have the most influence on the output cell populations obtained from the hybrid model. Second, the screened input parameters from the first stage were varied to build a second stage parameter design matrix, and the computer simulations were again run using the hybrid ENISI model. The outputs from both analytics stages were combined and used as a ‘training dataset’ to build a spatiotemporal Gaussian process based metamodel. Finally, variance-based decomposition global SA was used to compute the Sobol’ indices and the most influential parameters over the course of infection were identified. The data analytics methods conducted on the hybrid model identified the epithelial cell parameters such as epithelial cell proliferation as the most influential ones, required for the successful colonization of *H. pylori* in the gastric microenvironment.

**Methods**

**Hybrid multiscale *Helicobacter pylori* infection model**

We developed a multi-compartment, high-resolution, hybrid ABM/ODE/PDE model to capture the dynamics of the immune response during *H. pylori* colonization of the gastric mucosa. The model has a spatial discretization such that the dimension of the entire (two-dimensional, (2D)) grid is 30 mm x 10 mm. An individual lattice site for our simulation is 1 mm x 1 mm, however, this is a configurable run parameter and can be changed without modifying the model. An individual lattice site is a unit wherein all the agents located within that location have the same cytokine environment, *i.e.*, for all the agents in that location, ENISI-MSM would send the same concentration of the cytokines to COPASI. The entire grid is divided within into four functionally and anatomically distinct sized compartments: lumen, epithelium, lamina propria and gastric...
lymph node. In the model, there are multiple cells and cell types (i.e., agents) within this dimensional grid. At the beginning of each simulation cycle, the cells (agents) are randomly placed within the 2D grid. The separation of different types of agents, corresponding to different cell types, into compartments within the grid is based on the conceptual framework that underlines the model, which is based on author’s expertise and available information. Currently the individual agents do not have any physical size meaning such that there is no limit of agents within each individual spatial grid. The model is initialized with the concentration of different cell types (i.e. agents for e.g. macrophages) at the beginning of the simulation by the user.

The use of a border implementation permits the migration of agents (cells) across compartments and facilitates the unidirectional and bidirectional movement of the agents. At the cellular scale, ENISI MSM, simulated epithelial cells, macrophages, dendritic cells (DC), CD4+ T cells and bacteria that are implemented as agents in the model. At the intracellular scale, calibrated ODE-based models of T cells [23] and macrophages [2] were used to represent the intracellular pathways controlling cytokine production. The CD4+ T cell ODE model was calibrated using the experimental data provided in the Table S1 of [23]. The Particle Swarm algorithm implemented in COPASI was used to determine unknown model parameter values and fully calibrate the CD4+T cell ODE model, the details are described in [23]. The intracellular macrophage ODE model was calibrated using a combination of sourced and new data generated from in vitro macrophage differentiation studies, that were compiled into a dataset provided within S2 file of [2]. The parameter values are specified within the previously published manuscripts - CD4+ T cell ODE model Carbo et al. [2] and macrophages. The parameters of the calibrated ODEs were kept unchanged, and ABM parameters were calibrated by approximating the output simulations such that they qualitatively resembled the patterns observed in a mouse model of H. pylori infection [24], also described in detail in Results (see Results section, Hybrid model simulations produce similar immune dynamics observed in previously published experimental data). Cytokines secreted by immune cells and their change in concentration were modeled by PDE. The degradation value of the cytokines and the diffusion constant determines the spread of the cytokine value of one lattice site to its neighboring lattice site similar to as described in our previous work [12]. The features of ABM, ODE, and PDE were combined to create a multiscale modeling environment which spanned across different orders of spatiotemporal scales. The model output contains information about the x and y co-ordinate of the agents at every time point. The cytokines and internal signaling pathways that drive functional fates of cells are well mixed within a cell, i.e., we have only temporal resolution within the cell during a time step. Since, the model is capable of providing information regarding spatial co-ordinates over time, we claim the model to be a spatio-temporal model.

The code for the hybrid model is freely accessible and can be downloaded at (GitHub page for ENISI-MSM https://github.com/NIMML/ENISI-MSM Accessed April 30, 2019). The detailed instructions for the usability, instructions on ‘how to run a simulation’ and codes for creating specific examples presented here are presented in Additional file S1. The SciCrunch.org database assigned research identification initiative ID (RRID) for ENISI-MSM is RRID:SCR_016918.
design of the implementation of the code structure is depicted in the Additional file Fig S 2.1. The hybrid model is implemented in C++ and utilized the Repast HPC library [25]. For ODEs, we utilized COPASI [26], an ODE-based modeling tool used in computational biology. The rules in the model that described the interaction of *H. pylori* with the gastric mucosa and the immune responses resulting from the infection are derived from the findings in our previously published studies [1, 2]. Specifically, this hybrid model reproduced the immune responses generated by the interaction *H. pylori* and the resident macrophages as shown in the mouse model of *H. pylori* infection [24]. The rules for each cell type in the *H. pylori* infection are summarized in Table 2.1. A pictorial representation of the rules is depicted in *Figure 2.1*. These cell types represented as agents, act according to the rules (as in Table 2.1) that are updated at discrete simulation cycle.

**Table 2.1. A list of rules for all the agent types implemented in the hybrid model**

<table>
<thead>
<tr>
<th>Name of Agent</th>
<th>Agent Type</th>
<th>Rules</th>
</tr>
</thead>
</table>
| Helicobacter pylori | *H. pylori* | - Moves across the epithelial cell border if near damaged epithelial layer  
- Proliferates in the lumen and lamina propria  
- Dies (removed from the simulation) in lamina propria and in the lumen due to the damage of epithelial cells by Th1 or Th17 cells |
| Macrophages | Monocyte | - Proliferates in presence the of effector dendritic cells or damaged epithelial cells  
- Proliferates in the lamina propria  
- Differentiates to regulatory macrophage in based on the output from the Macrophage ODE  
- Differentiates to inflammatory macrophages in presence of IFN-γ  
- Dies naturally (removed from the model) |
| Resident | | - Proliferates in the presence of *H. pylori*  
- Secretes IL10  
- Dies naturally  
- Dies due to Th1 and Tr cells |
| Regulatory | | - Proliferates and removes bacteria  
- Dies  
- Secretes IL10 |
| Inflammatory | | - Proliferates in the presence of damaged epithelial cell  
- Dies naturally |
| Dendritics | Immature | - Moves from lamina propria to epithelium compartment and from the epithelium to the lamina propria  
- Differentiates to tolerogenic dendritic cell in the presence of tolerogenic bacteria, both in epithelium and lamina propria  
- Differentiates to effector dendritic cell in the presence of H. pylori  
- Proliferates in lamina propria and gastric lymph node  
- Dies naturally  |
|-----------|----------|--------------------------------------------------|
| Effector  | - Moves from lamina propria to gastric lymph node  
- Moves from epithelium to lamina propria  
- Secretes IL6 and IL12  
- Dies naturally  |
| Tolerogenic | - Moves from lamina propria to gastric lymph node  
- Moves from epithelium to lamina propria  
- Secretes TGF-b  
- Dies naturally  |
| T cells   | Naïve    | In the presence of effector dendritic cells:  
- Differentiates to Th1 in the presence of IFN-g or IL12  
- Differentiates to Th17 in the presences of IL6 or TGF-b  
In the presence of tolerogenic dendritic cells:  
- Differentiates to iTreg in the presence of TGF-b  
- Differentiates to Tr in the presences of IL10  
- Dies naturally  |
| Th1       | - Secretes IFN-g  
- Moves from gastric lymph node to lamina propria  
- Proliferates in lamina propria and gastric lymph node  
- Dies naturally  |
<table>
<thead>
<tr>
<th></th>
<th>Th17</th>
<th>iTreg</th>
<th>Tr</th>
<th>Epithelial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Secretes IL17</td>
<td>- Secretes IL10</td>
<td>- Secretes IL10</td>
<td>Healthy</td>
</tr>
<tr>
<td></td>
<td>- In the presence of tolerogenic dendritic cell, transition to iTreg cells</td>
<td>- In the presence of tolerogenic dendritic cell, transition to iTreg cells</td>
<td>- Dies naturally</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Moves from gastric lymph node to lamina propria</td>
<td>- Moves from gastric lymph node to lamina propria</td>
<td>- Proliferates in the lamina propria</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Proliferates in lamina propria and gastric lymph node</td>
<td>- Proliferates in lamina propria and gastric lymph node</td>
<td>- Dies naturally</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Dies naturally</td>
<td>- Dies naturally</td>
<td>- Proliferates in the lamina propria</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Damaged</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Transitions to healthy state in the presence of IL10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Dies naturally</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bacteria</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Dies due to Th1 or Th17 or inflammatory macrophages or damaged epithelial cells</td>
<td>Infectious</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Dies naturally</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Proliferates in the lamina propria</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tolerogenic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Moves from lumen to the epithelium in the presence of damaged epithelial cells</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Becomes infectious if moves in the lamina propria compartment</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Proliferates in lumen and lamina propria</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Dies naturally</td>
<td></td>
</tr>
</tbody>
</table>
**Figure 2.1. Helicobacter pylori infection schematic diagram of the hybrid ABM ODE model**

The model comprises four compartments, i) the lumen that contains *H. pylori* and bacteria, ii) epithelium that contains epithelial cells and dendritic cells, iii) lamina propria that contains a variety of immune cells including the infiltrating effector (eDCs) and tolerogenic (tDCs) dendritic cells, monocytes, regulatory macrophages (both resident and monocyte-derived macrophages), T helper cells and naïve CD4+ T cells (nT), Th1, iTreg, Th17, Tr cells. and iv) gastric lymph node compartment that contains eDCs, tDCs, Th1, Th17, iTreg and nT. The Tr cells in the lamina propria are the type 1 regulatory (Tr1) T cells with regulatory function whose expansion is largely dependent on environmental IL-10. These are different than iTreg which are T cells differentiated from naïve T cell in the presence of tolerogenic dendritic cells and TGF-β cytokine the two calibrated ODEs for T cells and regulatory macrophages are integrated as the ODE components in the hybrid model. The cellular agents are simulated in a two-dimensional grid space with their behavior defined by a set of rules during a course of *H. pylori* infection.

**Model description**

ENISI MSM is a multiscale agent-based modeling platform for computational immunology which was built on our previous works, ENISI-MSM [12] that integrated COPASI, the ODE solver, ENISI, an agent based simulator.

**Spatial discretization**

The model has a spatial discretization such that we define the area being simulated as a simulation environment with a two-dimensional grid whose size is 30 mm x 10 mm. An individual lattice site
is 1mm x 1mm, however, this is a configurable run parameter and can be changed without modifying the model. We further want to clarify that the above units in the model are annotations and purely aesthetic. The scales described in Table 1 in the previous version of ENISI-MSM [12] were kept unchanged.

The 4 functionally and anatomically distinct sized compartments are separated by border implementation such that the dimensions of the four compartments are lumen (2 mm), epithelium (1 mm), lamina propria (5 mm) and gastric lymph node (2 mm). The following compartments are adjacent to each other: lumen – epithelium, epithelium - lamina propria and lamina propria – gastric lymph node. A figure describing the spatial discretization is shown in Figure 2.2.

![Figure 2.2. A pictorial representation of the spatial discretization of the 2D grid.](image)

The parameters that define the initial concentration of the agents and the diffusivity of cytokines are obtained from a properties file (`model.props` in the “Howtorunasmulation” folder in the GitHub repository). All the values of the parameters as listed in Table S 3.1. The detailed mechanism that each parameter corresponds to is described in the second column, *parameter description*, of Table S 2.1. We demonstrate below how we obtain a count of thousands resident macrophages. For e.g., if the initial concentration of resident macrophages in the lamina propria is 30, the total number of these resident macrophages can be calculated by the equation described below -
\[ n(\text{resident macrophages}) = \text{size}_{\text{compartment}}(\text{lamina propria}) \times \text{concentration}_{\text{initial}} \text{ (resident macrophages)} \]
\[ n(\text{resident macrophages}) = (30 \times 5) \times 30 = 4500. \]

**Time Step size**
The time step size is 1 tick \( \sim 1 \) day which was obtained during the process of qualitatively comparing the output to the results from the mouse model of \( H. \ pylori \) infection. For e.g., the peak of resident macrophages in the lamina propria (refer Figure 2.3e and f) is observed at \( \sim 21 \) days which is similar to the results obtained in Fig 2A described in [24] (also described in detail in Results section as described in detail in Results section, Hybrid model simulations produce similar immune dynamics observed in previously published experimental data.)

**Updating**
Each agent has an ‘act’ function within the code that describes the rules implemented for each of the agent groups. At every simulation cycle, each agent inspects its location and updates its state. If the agents were T cells and macrophages, they obtained the cytokine concentration from the ValueLayers, sent that information to COPASI that calculated the differentiation subtype of the agent and cytokines to be secreted into the environment [12]. The input to the ODEs were the cytokine values at the agent’s location. Thus, the intracellular ODE models were utilized to determine and update the state. Each agent proliferated, died, changed its state and moved across the compartment, following the set of rules defined for them.

The COPASi setup for the solver used the LSODA (Livermore Solver for Ordinary Differential Equations) differential equation solver. The default values for the setup such as the - relative tolerance (1e-6), absolute tolerance (1e-12) and maximum internal steps of 10000 were maintained. The ENISI MSM sends the current concentrations of the cytokines to COPASI. COPASI uses those values to integrate the deterministic model for one tick, i.e., 1 day. The resulting time series of cytokine concentrations are used to update the cytokine value in the ABM/PDE system. COPASI simulates different model for each relevant cell type. The ENISI MSM PDE solver uses a simple numerical scheme to solve the PDEs (GitHub page for ENISI-MSM diffuser. https://github.com/NIMML/ENISI-MSM/tree/master/src/diffuser. Accessed 6 October 2016) and process distributed value layer (GitHub page for ENISI-MSM ValueLayer https://github.com/NIMML/ENISI-MSM/blob/master/src/grid/ValueLayer.h Accessed 30 March 2016. The ValueLayer stores the value for a grid space and provides methods to change the values of individual lattice site. The Diffuser is used to diffuse the values of the ValueLayer using diffusion (d) and degradation (delta) constants as described in [12]. The diffusion constant determines the migration of values of a lattice site to its neighboring lattice site. The table below shows the constants used for each immediate neighbor during the PDE integration.
Let’s index the neighbors based on their position relative to its current lattice position, where the following table specifies the integration constants $K$ for each neighbor relative to any lattice point. For example, here $(0,0)$ is a lattice point and $(-1,-1), (-1,0), (0,-1), (0,1), (1,-1), (1,0)$ and $(1,1)$ are its relative neighbors. This leads to the equations shown below for each lattice point.

\[
v_n = v_{n-1} + \Delta t \left( c_d \sum K_{\text{neighbor}} v_{n-1} \text{neighbor} - c \ v_{n-1} \right)
\]

Note that summing over the neighbors includes the lattice point $(0,0)$. Here, $v_n$ is the value of the lattice site itself at step $n$, $\Delta t$ is the integration step, the values of $c$ and $c_d$ are degradation and diffusion constants, respectively. The PDE solver uses the above number scheme $K_{\text{neighbor}}$ for the diffusion process. The step size is automatically adjusted at the beginning of the simulation based on the degradation and diffusion constants to avoid underflow errors; i.e., multiple PDE steps are in general executed per tick. The grid size is identical with the spatial discretization for the agents.

**Movement**

The cells and bacteria agents presented in the model have Brownian motion and move randomly within the compartment. Brownian movement is an inherent property of a cell. Depending on cell phenotypes the movement can vary, but all cells with the same phenotype exhibit similar movements. Additionally, chemokine-driven movement is dependent on chemokine concentration in a tissue site. The capability of chemokine-driven movement exists in ENISI-MSM if the right chemokines are represented in the model. However, the focus of this model was to investigate changes in cell phenotype and not chemokine-driven movement of cells. Thus, the chemokines driving the movement are not represented in the current model. Cell migration is implemented in the code as the `move()` function for each of the cells and agents, which call the `moveRandom()` function from the (GitHub page for ENISI-MSM Compartment. [https://github.com/NIMML/ENISI-MSM/blob/master/src/compartment/Compartment.cpp](https://github.com/NIMML/ENISI-MSM/blob/master/src/compartment/Compartment.cpp). Accessed 6 September 2017) file.

The hybrid model simulations were run on an Ivy Bridge-EX E7-4890 v2 2.80 GHz (3.40 GHz Turbo) quad processor nodes. The code was parallelized such that the simulation time on a single node with four parallel tasks, varied between 9-10 minutes. This runtime was based on the model parameters at the initiation stage, which included the number of immune cells, bacteria, epithelial

<table>
<thead>
<tr>
<th>Relative location</th>
<th>-1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0.3</td>
<td>1.2</td>
<td>0.3</td>
</tr>
<tr>
<td>0</td>
<td>1.2</td>
<td>-6.0</td>
<td>1.2</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>1.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>
cells, number of time steps, and the size of the two-dimensional grid. To facilitate the investigation of the mechanisms underlying host responses during *H. pylori* infection, anatomical and functional compartments were spatially linked such that the agents had both unidirectional and bidirectional movement. All the agents worked in a synchronous format wherein the two agent populations (macrophages and T cells) made function calls to their respective ODE models [2] [23]. These agents used the varying cytokine concentration (*i.e.*, environment variable) in their grid spaces as inputs to the ODE model, and these models were run using COPASI [26]. Table 2.2 shows information on the agents and the states that they can acquire. All the agents can acquire at least 1 and at most 5 states. The names chosen for the acquired states are closely related to their functional properties based on the underlying “rules”.

**Table 2.2. List of all the agents and the states they can acquire.**

<table>
<thead>
<tr>
<th>Name of agents</th>
<th>States it can acquire</th>
<th>Name of the states in the hybrid model</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Helicobacter pylori</em></td>
<td>0</td>
<td><em>H. pylori</em></td>
</tr>
<tr>
<td>Macrophages</td>
<td>0</td>
<td>Monocyte</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Resident</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Regulatory</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Inflammatory</td>
</tr>
<tr>
<td>Dendritic cells</td>
<td>0</td>
<td>Immature</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Effector</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Tolerogenic</td>
</tr>
<tr>
<td>T cell</td>
<td>0</td>
<td>Naïve</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Th1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Th17</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>iTreg</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Tr</td>
</tr>
<tr>
<td>Epithelial</td>
<td>0</td>
<td>Healthy</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Damaged</td>
</tr>
<tr>
<td>Bacteria</td>
<td>1</td>
<td>Infectious</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Tolerogenic</td>
</tr>
</tbody>
</table>

Furthermore, we included the screenshots of 1 actual in silico simulation of *H. pylori* infection to highlight the spatiotemporal aspects of the modeling outputs. The time snapshots were created using VisIt version 2.12 [27], an interactive visualization and analysis tool. As shown in
Additional file, Fig S 2.2, the screenshots at time points 2, 4, 5 and 6 represent the spatial distribution of different agent cells over time distributed across the 2D grid.

Global sensitivity analysis

To conduct the global SA, we determined a list of 38 parameters to be varied that were selected based on the calibration process (wherein the parameters that did not show a lot of variation were not included). A range of values (maximum and minimum) was specified for each of the parameters (refer Additional file Table S 2.1) by expert judgment, summarized by bounded intervals. The practice of using expert judgment is known in the SA field as supported in [28]. As discussed in [29], one of the challenges encountered using ABM is the process of determining the parameter values, for e.g. this may include the lack of the availability of experimental techniques to measure such parameters. The values of the parameters for the model presented here are obtained via the best guess based on the qualitative comparison of the computer model outputs with that of the experimental results obtained from the mouse model of *H. pylori* infection [24] (as described in detail in Results section, *Hybrid model simulations produce similar immune dynamics observed in previously published experimental data*). Since, the source of the parameters is not known we estimated the values to fit the data obtained from the mouse model of infection.

The values of these parameters were normalized within the range of 0 and 1 for SA purposes. We employed a two-stage metamodeling methodology to determine the influence of each input parameter to the model output, in a high dimensional screening setting inspired by [30]. The step-wise procedure is described in the Additional file, Fig S 2.3. All the files for global SA are freely accessible and can be downloaded from GitHub page for Sensitivity Analysis - (https://github.com/NIMML/ENISI-MSM/tree/master/Sensitivity-Analysis. Accessed 18 January 2019.)

The 2-stage global SA is described in detail in the below section. To summarize, for the first stage the input parameter matrix was designed using the method described in Moon et al. [30] and simulations were run using the hybrid computer model. The simulation output from the first stage was analyzed using PRCC as it was computationally efficient, and the active inputs (significant effect) were screened to reduce the input parameter space. Second, the active parameters were varied whereas the inactive parameters from the first stage were maintained at a nominal value for the input parameter matrix design to be employed for the second stage. Third, the simulation outputs from both stages were combined and used as a training dataset to fit a spatio-temporal metamodel. Fourth, the unknown model parameters for the spatio-temporal metamodel were estimated using the maximum log-likelihood function. The spatio-temporal metamodel was used as a substitute for the hybrid computer model, and the variance-decomposition method was used to compute the Sobol’ total and first-order indices. Overall, we employed both approaches, PRCC based (for screening) and Sobol’ indices calculation to perform a complete global SA of the hybrid computer model. The following sections explain the procedure step by step.
Design of two-stage experiments and analysis

The input for the hybrid model are varying parameter values obtained from the design matrix and the output are the number of cells (agents) that vary over time. The first stage experiment was focused on the screening of the input variables to reduce the number of input parameters to vary for the SA and to limit the computational cost. Computational costs are often a limiting factor that play an important role in the inclusion of model parameters in the SA [21]. For the design, we assumed the total number of input parameters under consideration as $d$ (in our case, 38). With an assumption of a maximum of 50% active inputs that is aimed to improve the screening performance, the number of runs for stage 1, was fixed to $n_1 = 4d$, such that $n_1 > 5*d*0.5 = 2.5d$ as in [30]. To construct a $n_1 * (n_1-1)$ preliminary input parameter design matrix, $X^*$, needed to be constructed ([30]). The input parameter design matrix for first stage sampling was drawn from $X^*$. The algorithm for the first stage design generated a design matrix $X^{(1)}$ that satisfied the below three listed properties as in [30]

i) The columns of $X^*$ were uncorrelated thereby facilitating the independent assessments of the effects due to the input parameters.

ii) The maximum and minimum value in each input parameter column were ensured to be 0 and 1 respectively, thereby preventing any input values with larger values to have a larger influence on the response, induced by the design.

iii) The designs defined by $X^*$ had “space-filling” properties such that all the regions of the input space were exhaustively explored.

First stage sampling plan

The first stage input parameter design matrix $X^{(1)}$ was obtained by selecting the first $d$ columns of $X^*$, i.e. $X^{(1)} = (\xi_1, ..., \xi_d)$. The hybrid computer model was run and the simulation outputs at these $n_1$ design points were obtained.

In our case, the model comprised of $d = 38$ input variables. The total number of distinct input parameter design points obtained using the above procedure was $n_1 = 152 (4*d = 4*38)$. To account for the variability in the output, we run 20 replicates ($r$). Thus, the total number of simulations run using the hybrid model computer simulator with $X^{(1)}$ as input parameter design matrix, were $r \times n_1 = 20 \times 152 = 3040$.

First stage analysis

We analyzed the outputs from first stage analysis and screened the active inputs from using PRCC. To measure the effect of input parameter on output, we performed both PRCC and the spearman rank correlation coefficient (SRCC) analysis. PRCC and SRCC were chosen because they were computationally efficient (accounting for the low computational budget). A correlation analysis provides a measure of the strength of linear association between input and output variable [31]. A correlation coefficient between $x_j$ and $y$ is calculated as follows:

$$r_{x_jy} = \frac{\text{Cov}(x_j, y)}{\sqrt{\text{Var}(x_j)\text{Var}(y)}} = \frac{\sum_{i=1}^{N}(x_{ij} - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_{ij} - \bar{x})^2} \sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$$
\[ j = 1, 2, ..., k. \]

where \( \text{Cov}(x_j, y) \) stands for the covariance between \( x_j \) and \( y \), and \( \text{Var}(x_j) \) and \( \text{Var}(y) \) are the variance of \( x_j \) and \( y \) respectively.

PRCC is performed when i) a non-linear but monotonic relation exists between the input and outputs, and ii) when little or no correlation exists between the input variables (which is guaranteed by the property (i) of our input parameter matrix, \( X^{(1)} \) described above). As described in Marino et al. [31] the PRCC between rank transformed \( x_j \) and \( y \) is the CC between the two residuals \( (x_j - \bar{x}_j) \) and \( (y_j - \bar{y}_j) \) where \( \bar{x}_j \) and \( \bar{y}_j \) are rank transformed and follow the linear regression models as follows:

\[
\bar{x}_j = c_o + \sum_{p=j}^{k} c_p x_p \quad \text{and} \quad \bar{y}_j = c_o + \sum_{p=j}^{k} c_p x_p .
\]

We performed the PRCC analysis on the outputs obtained from the hybrid computer model with \( X^{(1)} \) as an input, using ‘epi.prcc’ package in R (https://cran.r-project.org/web/packages/epiR/epiR.pdf). The significance test evaluated the strength of influence each input parameter and assessed if the PRCC coefficients were significantly different than zero [31]. We run the PRCC analysis for 13 output cell populations (Figure 4 shows data for two output populations and the rest of the data not shown) and identified the active input parameters using the significance test. PRCC and SRCC produced identical outputs, hence results from SRCC are not shown here. If an input parameter was shown to be significant \( (P < 0.05) \) in one of the 13 output cell populations, it was considered as an active input for the second stage input parameter design matrix. Additionally, domain expert knowledge was employed to include additional parameters, based on the biological significance, that were otherwise shown to be non-significant. In all, based on the PRCC analysis performed on the outputs obtained from the first stage simulations and domain expert knowledge, we chose 23 input parameters as active inputs for the second stage (see Additional file Fig. S 3.4). Thus, PRCC screened inputs at significance level \( p < 0.05 \) and inputs based on expert knowledge were selected as active inputs to be varied for the second stage sampling plan.

**Second stage sampling plan**

The number of active inputs obtained from the first stage analysis amounted to 23 parameters out of the initial set of 38 parameters. We followed the design described in [30] for the second stage and the number of design points amounted to, \( n_2 = 100\% \times 5*a \) where ‘a’ stands for the number of active inputs from the first stage. This resulted into \( n_2 = 23 \times 5 = 115 \) parameters combinations for the second stage input parameter design matrix. Since outputs from both stages are to be combined for second stage analysis, per [30], the design for the second stage was chosen to build on top of \( X^{(1)} \). The sampling phase design algorithm ensured that the columns satisfied the properties (i) (uncorrelated design points) and (ii) (between values 0 and 1) as listed in the previous section. We constructed the 115 x 38 (115 parameter setting and 38 parameters) design matrix for the second stage that incorporated the 23 active inputs obtained from the PRCC screening in the first stage output analysis. After combining the design points from both the stages, the parameter design matrix \( X \) with space filling properties contained 267 (152 from the first stage and 115 from the second stage) design points.
**Second stage analysis**

We run the computer code for the hybrid model with the second stage input parameter design matrix (with 115 \((n_2)\) design points), for 20 \((r)\) replicates, which amounted to 115 x 20 (2300) runs. The outputs from the first stage (152 x 20 runs) and second stage (115 x 20 runs) were combined to provide the training data to build a spatio-temporal metamodel. For the second stage analyses, we utilized a metamodeling-based approach. Metamodels are surrogate models that can be used as a substitute for the simulation model [32]. The use of metamodels reduces the computational budget, cost of analysis, and are useful options in cases when the simulation model is expensive to run (in our case 9-10 minutes for 1 design point) [32]. The various metamodeling techniques used to build surrogates for a computer model output include linear regression models, neural networks, high dimensional model representation methods, Gaussian process (GP) regression models, polynomial chaos expansion and more that are discussed in length in [33, 34]. Amongst these, GPs are one of the most popular emulators as it allows modeling of fairly complex functional forms. The GPs not only provide prediction at a new point but also an estimate of the uncertainty in that prediction [33]. A GP is a stochastic process for which any finite set of y-variables has a joint multivariate Gaussian distribution [35] [33]. Suppose, \( y_j(w) \), the simulation response obtained on the \( j \)th simulation replicate, at a design point \( w = (X^T, t)^T \in \chi \times T \), it can be described as follows:

\[
y_j(w) = Y(w) + \varepsilon_j(w) = \beta_0 + M(w) + \varepsilon_j(w),
\]

where \( Y(w) \) represents the mean function of \( y_j(w) \), the quantity of interest that we intend to estimate at any design point \( w \). The \( \beta_0 \) is a constant trend term and is assumed to be unknown. The input parameter \( X \in \chi \subseteq \mathbb{R}^d \) and the time \( t \in T \subseteq \mathbb{R}^+ \); and \( X \) is independent of \( t \). The \( \varepsilon_j(w) \) are represents the sampling variability inherent in a stochastic simulation, that are that are assumed to be independent and identically distributed across the replications at any given design point [36].

The term \( M(w) \) represents a stationary Gaussian process with mean \( = 0 \) and covariance between any points was modeled as the Gaussian covariance defined in [37]. Thus, the covariance between any design points \( w_a = (X_a^T, t_a)^T \) and \( w_b = (X_b^T, t_b)^T \) in the random field can be modeled as

\[
\text{Cov}(M(w_a), M(w_b)) = \Gamma^2 \exp\left(-\sum_{r=1}^{d} \theta_r (X_{ar} - X_{br})^2 R(t_a - t_b; \gamma)\right),
\]

wherein, \( \exp\left(-\sum_{r=1}^{d} \theta_r (X_{ar} - X_{br})^2 \right) \) models the spatial correlation between two input design points \( X_a \) and \( X_b \) in the input parameter space, whereas \( R(t_a - t_b; \gamma) \) also given by \( \exp\left(-\sum_{r=1}^{d} \gamma_r (t_{ar} - t_{br})^2 \right) \) models the temporal correlation between time points \( t_a \) and \( t_b \). The parameters \( \theta \) and \( \gamma \) represents the rate at which i) spatial correlation decreases as the points move farther in space with the same time index, and ii) temporal correlation decreases as the time points are farther apart in time at the same input vector, respectively. Both the spatial correlation and temporal correlation are modeled using the Gaussian covariance. The parameter \( \Gamma^2 \) can be interpreted as the variance of \( M(w) \) for all \( w \). The input parameter design consists of \( \{(w_a, n_i)_{i=1}^k\} \) design points to run independent simulations with replicates applied to each of the design points. Let, \( k \times 1 \) denote a vector of sample averages of simulation responses given by \( \overline{y} = (\overline{y}(w_1), \overline{y}(w_2), \ldots, \overline{y}(w_k))^T \), where in \( \overline{y}(w_i) \) is the resulting estimate of performance measure obtained at design point \( w_i \) and \( \overline{\varepsilon}(w_i) \) is the sampling variability inherent in a stochastic simulation [36]. The equations associated with \( \overline{y}(w_i) \) and \( \overline{\varepsilon}(w_i) \) are described below in equation (3):
\[
\bar{y}(w_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} y_j(w_i) = Y(w_i) + \bar{e}(w_i) \quad \text{and} \quad \bar{e}(w_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} e_j(w_i), \quad i = 1, 2, \ldots, k \quad (3).
\]

Similar as in Ankenman et al. [36], shown below in equation (4), let \( \Sigma_M \) be the \( k \times k \) covariance matrix across all design points and let \( \Sigma_M(w_{0r}) \) be the \( k \times 1 \) vector, \( (\text{Cov}[M(w_0, w_1)], \ldots, \text{Cov}[M(w_0, w_k)]^T \) that contains spatial covariance between the \( k \) design points and a given prediction point \( w_0 \). Also, let \( \Sigma_e \) be the \( k \times k \) covariance matrix of the vector of simulation errors associated with the vector of point estimates \( \bar{y} \), across all design points. As described in [36], the best linear predictor \( Y(w_0) \) that has the minimum mean squared error (MSE) among all linear predictors at a given point \( w_0 = (X_o^T, t_o)^T \) can be given by equation (4):

\[
\hat{Y}(w_o) = \hat{\beta}_o + \Sigma_M(w_0, \cdot)^T [ \Sigma_M + \Sigma_e ]^{-1} (\bar{y} - 1_k \hat{\beta}_0), \quad (4)
\]

where, \( 1_k \) is the \( k \times 1 \) vector of ones and \( \hat{\beta}_o \) is estimated to be 1. The corresponding optimal MSE as in [36] is given by equation (5):

\[
\text{MSE} \left( \hat{Y}(w_o) \right) = \Sigma_M X_0, w_0 - \Sigma_M(w_0, \cdot)^T [ \Sigma_M + \Sigma_e ]^{-1} \Sigma_M(w_{0r}) \quad (5).
\]

To implement the metamodeling approach as described above, the unknown model parameters are estimated through maximizing the log-likelihood function. The underlying standard assumption is that \( (Y(w_0), \bar{y}^T)^T \) follows a multivariate normal distribution, for e.g., see [36] and [38]. The function implemented in the mlegp package in R [39] is used for the estimation of the parameters. Once the parameters are estimated the prediction then follows equations (4) and (5).

**Sensitivity index calculation**

To determine the effect of input variables on the output, we employed the variance decomposition method. These methods involve the decomposition of the variance of the output as a sum of the variance produced by each input parameter [35].

We independently generated 10,000 x 38 sampling matrices, such that the parameter combinations are generated via Latin Hypercube sampling and as described in [40]. Simulations were performed using the GP spatio-temporal model as described in the previous section, and the Sobol’ indices were computed as described in [41] [40]. The Sobol’ method quantitatively measured the contribution of each input parameter by computing the first order and total order index [40]. For output \( Y \), input parameter matrix \( X_i \) where, \( i \) is the input parameters of the model, the Sobol’ indices are computed as follows:

\[
SI^{Xi}_1 = \frac{V[E(Y|X_i)]}{V(Y)},
\]

and

\[
SI^{Xi}_{tot} = \frac{V[E(Y|X_{-i})]}{V(Y)}.
\]

The Sobol’ first order sensitivity index \( SI^{Xi}_1 \) measures the impact of one single parameter on the model output, whereas the Sobol’ total order index measures the influence of \( X_i \) including all the interactions with other parameters. The First-order indices were computed using the Sobol-Saltelli’s method as described in [40] whereas, the total order indices were computed using Sobol-Jansen as in [40, 43].
Results

Hybrid model simulations produce similar immune response dynamics observed in previously published experimental data

We first aimed to simulate the findings observed in previous gut models [24] to ensure that we obtained similar response dynamics from the hybrid ENISI model of *H. pylori* infection. As in [24], to demonstrate that the gastric mucosa harbors a system of macrophages that contribute to the outcome of *H. pylori* infection, we created an *in-silico* Peroxisome proliferator-activated receptor gamma (PPARγ) macrophage-specific knockout (KO) model. PPARγ is an important transcription factor that controls the expression of genes that contribute to the inflammatory response once this is initiated. To disrupt the downregulation of pro-inflammatory responses, we simulated a PPARγ KO system in either macrophage or T cell populations and compared the response to a wild-type system. In the model, we created three different macrophage populations, comprised of, “resident” macrophage agents that mimic the properties of the F4/80hi CD11b+ CD64+ CXCR1+ macrophages reported in [24], monocyte-derived (infiltrating) and macrophage populations with regulatory (M2, or alternatively activated) and pro-inflammatory function (M1 or classically activated) (see Table 2.2).

We simulated an *in-silico* *H. pylori* infection by creating four groups, i) a control - WT (representing a wild-type group), ii) CD4Cre (T cell specific PPARγ KO-lacks PPARγ gene in all CD4 T cells), iii) LysMCre (Myeloid cell specific PPARγ KO-lacks PPARγ gene in all macrophages) and Clodronate group (simulating the removal of macrophages by chemical depletion via clodronate treatment). To simulate the CD4Cre group, the probabilities of a naive T cell transitioning to an iTreg cell (*p_nTtoiTreg*) and Th17 cell differentiating to iTreg (*p_Th17toiTreg*) were reduced to 5% and 10% of the control value, respectively (refer to Table S1). As described in [23], to simulate the LysMCre experimental conditions, the probabilities of i) a monocyte transitioning to a regulatory macrophage (*p_Mregdiff*) and ii) immature dendritic cells switching to tolerogenic dendritic cells (*p_idCtotDC*) were reduced approximately to 60% and 30% of the control value, respectively (refer to Table S 3.1). A complete set of parameter for each of the biological KOs are included as separate columns in Table S 21. Lastly, the removal of macrophages by clodronate were simulated by decreasing the initial numbers of the macrophage population including the resident macrophages. The rationale to include the clodronate group (macrophage removal) was to evaluate if depletion of phagocytic cells (terminology with respect to model, i.e., monocytes, resident, monocyte-derived macrophages and inflammatory macrophages) would affect *H. pylori* colonization levels, as we have previously reported in an *in vivo* model [24]. Further, to simulate the myeloid cell PPARγ KO system, the initial population of resident macrophages were also reduced.

All the groups were initialized with equal loads of *H. pylori* agents. Ten replicates of the simulations were performed for each of the input parameter settings specific to each group. The outputs were averaged, and standard error of the means were plotted as ribbons (shaded regions) across the graphs. After running the ten replicates of the time series *in-silico* simulation, the hybrid model showed significantly (*p < 0.05*) higher levels of *H. pylori* in the WT and CD4Cre groups as compared to LysMCre KO and macrophage-depleted groups (Figure. 2.3b and c).
Figure 2.3. Time course simulations representing the immune response during Helicobacter pylori infection.

Time course simulations representing the immune response during Helicobacter pylori infection. The upper half of the plot in both panels shows the dynamics of the population cells over time representing the number of cells (y-axis) vs time (x-axis) in wild type (WT) (black), CD4Cre (green), clodronate (red), and LysMCre (blue) simulated in silico groups during H. pylori infection (see Results section, Hybrid model simulations produce similar immune dynamics observed in previously published experimental data for details about the groups). A side-by-side comparison with the bacterial load and macrophage population as observed in the mouse model of H. pylori infection is also included (a and d). The cell populations include (b) H. pylori, (e) the resident macrophages, and (g) monocyte-derived macrophages in the lamina propria compartment. The graphs in the lower half (c, f, h) of these panels show the
results for statistical comparison between the groups using ANOVA with the post hoc analysis. The letters “a,” “ab,” and “b” represent statistically significant differences (P < 0.05) between the groups obtained after running the Tukey honestly significant difference. The groups with letter "a" are statistically significant different than the group with letter "b", groups with same letter are not statistically significantly different than each other. The group with letter "ab" is not statistically significantly different than group "a" and "b". In the box plots, the horizontal line is the median of the respective cell population, the box contains the interquartile range, the whiskers show the 95% confidence interval, and the dots are the outliers. cfu: colony-forming units.

In addition to the increase in *H. pylori*, WT and CD4Cre *in-silico* experimental groups had a higher resident as well as monocyte-derived regulatory macrophages as compared to clodronate (macrophage depleted) and LysMCre groups (*Figure. 2.3e, f, g and h*). The results in the mouse model indicated that between weeks 2 and 3 post-infection a decrease in bacterial burden in the stomach of LysMcre mice was observed as shown in *Figure 1A* of Viladomiu et al. [24]. The decrease in bacterial burden led to a significant and sustained lower colonization levels when compared to WT and CD4Cre. Similar to the results observed in the mouse model, we observed a significant decrease (*Figure. 2.3b and c*) in the bacterial burden in the simulated LysMcre group as compared to the simulated WT and CD4cre groups. Furthermore, the results from the mouse model indicated that a significant increase in numbers of F4/80hiCD11b+ CD64+ CX3CR1+ cells (here referred to as resident macrophages in this paper), was observed in WT mice in comparison with LysMcre mice as shown in *Figure 2A, 2E* of Viladomiu et al. [24]. These cells accumulated in the stomach mucosa starting on day 14 post-infection in the WT mice but not in the LysMcre mice. We observed a similar increase (*Figure. 2.3e, f, g and h*) in the number of resident macrophages as well as monocyte derived macrophages in the simulated WT groups in comparison to the simulated LysMcre group. We estimated the parameter values to fit the data obtained from the mouse model of *H. pylori* infection. Thus, the observations were qualitatively similar to the findings in [24], where the stomach of WT mice was enriched in a population of F4/80+CD11b+CD64+ myeloid cells, compared to LysMCre mice.

Overall, with the results in *Figure. 2.3*, we showed the ability of the hybrid model to replicate the experimental results in [24], and this preliminary data was used as a base calibration setting for SA and other *in-silico* findings.

**Partial correlation coefficient analysis screened the influential parameters**

To reduce the computational complexity of varying an input parameter space of 38 parameters, we divided the SA process in two stages. For first-stage analysis, we utilized the PRCC regression-based SA method to screen the influential inputs and used it for the second stage design of the experiments (see Methods section, *Global Sensitivity Analysis*). Using PRCC, we determined the impact of the input parameters on the output cell populations in the model. The parameters with significant correlation with *H. pylori* in the gastric lamina propria compartment and resident macrophages are shown in *Figure. 2.4*, along with their PRCC values. The bars in blue, highlight the parameters that are significantly different than 0, at P < 0.05 compared to grey bars which are not significant. It is important to note that at this stage the analysis using PRCC was non-temporal.

The SA from first stage results showed that the epithelial damage due to infectious bacteria (*epinfbctdam*) with a coefficient value of (~0.2), was positively correlated with the colonization
of *H. pylori* in the lamina propria compartment, indicating the important role of epithelial cell damage during the course of infection, similar to our findings obtained in [44]. Another parameter included the probability of the release of IL-6 (IL6) with a coefficient value within the range (0.3-0.4).

Next, the epithelial cell damage parameters (epiinfbctdam = (0.2-0.3), epiTh17dam = 0-0.2) were shown to have positive influence on the resident macrophage cells whereas, the T cell type transition parameters (p_iTregtoTh17 = (0.3 - 0.4) and p_Th17toiTreg = (0.1 - 0.2)) showed a negative impact on the resident macrophages. Similarly, we performed the PRCC analysis for all the cell populations under consideration during the infection (not shown).

**Figure 2.4. Bar plots for the partial rank correlation coefficients.**

The magnitude of the bar-plot indicates the value of the partial rank correlation coefficient. The blue bar indicated the input parameters shown to be significantly different than 0, at P <0.05 as influential whereas the grey bars indicate the non-influential parameters on a) *H. pylori* and b) resident macrophages, in the lamina propria compartment. The detailed explanation of the abbreviations for the parameters are in Table S 2.1.

The significant parameters (marked in blue bars) obtained from the SA of the output from first stage design of experiments (152 parameter settings with 20 replicates, see Methods section, Global Sensitivity Analysis), were selected to be varied for the second stage design. All the selected inputs are shown in Additional file Fig. S 2.4. In all, we obtained 23 active inputs from the first stage.

**Metamodel based spatio-temporal sensitivity analysis**
The outputs obtained after running the first (152 x 20 runs) and second (115 x 20 runs) stage simulations, wherein x20 denotes the 20 replicates, were combined to be used as a training dataset. The combined output was utilized to build a Gaussian process based spatiotemporal metamodel (see Methods section, Global Sensitivity Analysis), using mlegp package in R [39].

The outputs from the training dataset were sub-divided into 6 datasets, corresponding to six time periods (Days 1-14, 15-21, 22-30, 31-42, 43-90, 91-201) and averaged across these periods. The sub-division of output across the time periods, aided the temporal analysis over the initiation (Day 1-14), peak of infection (Days 15-30) and chronic phase (post Day 31) stages as in [24]. We then fit a Gaussian process model (with nugget) and evaluated the performance of the fitting of the metamodel for H. pylori, resident macrophages, and monocyte-derived macrophages in lamina propria compartment, and tolerogenic DC in the gastric lymph node, using the diagnostic plots (see figures in Additional file, Fig. S 2.5). After fitting the models, we performed variance based global SA by computing the Sobol’ total order and first order sensitivity index (see Methods section, Global Sensitivity Analysis). The estimates of the Sobol’ total order indices for the input parameters calculated over the six time periods are shown in Figure. 2.5.
**Figure 2.5. Heat-maps of Sobol’ total order index for the input parameters across different output populations.**

The values in the heat-map indicate the Sobol’ total order sensitivity index obtained from the metamodel, for the 38 input parameters with respect to the cell populations. The values with darker color indicate a stronger influence on the cell population as compared to the ones with lighter shade that indicate non-influential parameters for the cell populations - a) H. pylori, b) monocyte-derived macrophages, c) resident macrophages, in the lamina propria compartment and d) tolerogenic DCs, in the gastric lymph node compartment. The indexes are calculated over six time points ranging across the three stages of infection, including initiation (Day 1-14), peak (Days 15-42) and recovery stages (Days 43-201). The detailed explanations of abbreviations for the parameters are in Table S1.
As shown in Figure 2.5a, the metamodel based global SA showed that the input parameters, epithelial cell proliferation (Epiprolifer) and epithelial cell death (Epicelldeath) had the strongest impact on the population of H. pylori in lamina propria compartment. As time progressed from initiation of the infection (Days 1-14), through peak (Days 15-30), the epithelial cell proliferation had a continued impact on the colonization of H. pylori. Next, the influence of the probability of epithelial cell death decreased over the course of infection. Furthermore, Figure 2.5b highlighted the impact of epithelial cell proliferation (Epiprolifer) and epithelial cell death (Epicelldeath) on the monocyte-derived macrophages.

For the resident macrophage population in the lamina propria, that have emergent properties similar to the one characterized in [24], we observed that the resident macrophage replication parameter (ResmMacRep) has an impact during the initiation and peak stages of the infection which indicates that these subsets of macrophages replicate during the course of H. pylori infection. This result highlights the reliability of the two-staged global SA method used here, as these findings are consistent with the ones in [24] wherein we observed that these subsets of macrophages expand in the gastric stomach lamina propria during the course of H. pylori infection.

Finally, for the tolerogenic DCs in Figure 2.5d, we observed that the epithelial cell death (Epicelldeath) seemed to have an impact. Another parameter that stands for the probability of naive T cell transitioning to iTreg cell (nTtoiTreg) was shown to have an impact on the tolerogenic dendritic cells. Tolerogenic dendritic cells are involved in the rule that transitions the naive T cells to iTreg cells in the gastric lymph node, and the stronger impact of the nTtoiTreg during the initiation and peak stages of the infection highlights the role of the tolerogenic dendritic cells during the course of infection.

The global SA data suggested that the main contributors of the chronic colonization of H. pylori in the lamina propria are the epithelial cells, specifically the epithelial cell proliferation parameter.

**Effect of different ranges of epithelial cell proliferation parameter.**

An interesting prediction derived from the metamodel based global SA is that epithelial cell proliferation is one of the parameters that has a strong impact on the size of H. pylori population. The biological hypothesis derived from this prediction is that the epithelial cell proliferation is responsible for the higher colonization of H. pylori. Prior to conducting any experimental studies, we wanted to explore the hypothesis using our hybrid computer model in silico and study the model outputs obtained after we changed the epithelial cell proliferation parameter. Thus, we varied the epithelial cell proliferation parameter across different ranges (0.1-0.9, with 0.6 being the value for baseline conditions) and ran the simulations using the hybrid model and studied its effect on the different output cell population (obtained after running the simulations). These outputs were the ones obtained after running the simulation using the hybrid computer model, as we varied the epithelial cell proliferation parameter. We analyzed the outputs from the hybrid computer model and interestingly, observed that upon decreasing the Epiprolifer from a range of values 0.9-0.1, the output cell populations with regulatory function, namely regulatory macrophages and tolerogenic dendritic cells were found to vary. We observed a decreasing effect (Figure 2.6) on H. pylori, monocyte-derived macrophages, resident macrophages in the lamina propria compartment and tolerogenic dendritic cells in gastric lymph node. Overall, these cell populations varied due to the variation in the epithelial cell proliferation parameter.

For clarification, such connection was not embedded in the mechanisms included in Table 2.1 but it represents an emergent behavior from the simulations predicting the involvement of
regulatory and tolerogenic dendritic cells in the mechanisms of immunoregulation during *H. pylori* infection. Finally, the simulations targeting the epithelial cell proliferation resulted in changes in regulatory and tolerogenic dendritic cell populations. This shows that the simulations indirectly targeted the regulatory and tolerogenic dendritic cell population. Thus, we hypothesize that epithelial cell proliferation might be responsible for the higher colonization of *H. pylori* through an immunoregulatory mechanism that involves regulatory macrophages and tolerogenic cells. This is in line with our own conclusions drawn from a previous paper [24] where we show that the presence of cells with regulatory phenotype favor higher levels of *H. pylori* colonization. The results from the sensitivity analysis presented in this paper suggest that epithelial proliferation might be a crucial part of the mechanisms by which these regulatory responses are induced and that there is a link between these parameters. The exact biological process however cannot be inferred from the current model and it will be investigated in follow-up in vivo studies.

![Graphs showing the effect of varying epithelial cell proliferation parameter on the cell populations.](image)

**Figure 2.6. In silico study of the effect of epithelial cell proliferation parameter on the cell populations.**

The plots show the effect of varying epithelial cell proliferation (p_Epiprolifer) parameter (with values 0.1, 0.5, 0.6(WT), and 0.9) on the output cell population of a) *H. pylori*, b) tolerogenic dendritic cells, c) resident macrophages and d) monocyte-derived macrophages. The parameter has a decreasing effect on the cellular populations under consideration, wherein a decrease in the parameter value, decreases the abundance of the cells over time. The lower half of the figures (a-d), show the results for statistical comparison between the groups using ANOVA with the post-hoc analysis. The letters ‘a’, ‘b’, ‘c’, and ‘bc’ represent statistically significant differences (P<0.05) between the groups obtained after running the Tukey’s Honestly Significant Difference. The groups with letter “a” are statistically significantly different (P < 0.05) than groups with letter “b”
and “c”. The group with letter “bc” are statistically significantly different than group “a” but they do not differ from groups “b” and “c”. The groups with same letter are not statistically significantly different than each other.

The in silico findings suggested the involvement of regulatory macrophages (both resident as well as monocyte-derived) and tolerogenic DC on the colonization of H. pylori in the gastric lamina propria. This highlighted and validated the role of epithelial cell proliferation as one of the main factor affecting H. pylori levels in the gastric niche.

Discussion

H. pylori is the dominant indigenous bacterium of the gastric microbiota. In the majority of individuals, H. pylori colonizes the stomach without causing adverse effects, with little to no activation of inflammatory pathways. However, certain members of the population lose immune tolerance to the bacterium thereby contributing to the development of chronic gastric diseases. The immunological mechanisms underlying its ability to persist in a harsh acidic gastric environment and its dual role as a pathogen and beneficial organism remain unknown. A subset of macrophages helps create a regulatory microenvironment that promotes the chronic colonization of H. pylori [24]. However, the immune regulatory mechanisms are incompletely understood. Computational models of the immune system featuring immune responses are powerful tools for testing the different ‘what-if’ scenarios. Multiscale models of the immune response are attractive in terms of modeling the responses at different spatiotemporal scales [45].

In this study, we developed a HPC-driven hybrid, high-resolution, multiscale model to simulate the complex immunoregulatory mechanisms during H. pylori infection. The hybrid model was integrated with two intracellular ODEs capturing the dynamics of CD4+ T cells and regulatory macrophages. The inputs to the hybrid model are the set of parameters whose variation governs the immune system dynamics during infection. The obtained outputs were emergent patterns of different cell types, cytokines, and bacterial levels for instance the levels of H. pylori, and that qualitatively matched the patterns observed in an in vivo infection model [1, 24]. We presented an in-silico framework that evaluated the global SA of the hybrid model and studied how the variation in the biological parameters affected the simulation outputs. The two-stage global SA indicated that epithelial cell parameters, specifically, the proliferation of epithelial cells affected the colonization of H. pylori in the gastric mucosa. These results were validated in silico, and highlighted the involvement of regulatory macrophages and tolerogenic DC in facilitating H. pylori colonization of the gastric mucosa. Previous studies highlighted H. pylori inhabits the apical surfaces of the epithelial cells and maintains a persistent infection [46].

Further, Mimuro et al. [47] demonstrated that H. pylori promotes epithelial gastric cell survival by attenuating apoptosis. These events showed how H. pylori regulated the gastric niche and utilized epithelial cells to facilitate its persistence within the stomach [47] [48]. Thus, the findings in the current study are in line with the literature that suggests epithelial cell proliferation favor the colonization of H. pylori in the stomach.

Our group also showed another mechanism used by H. pylori to create a gut microenvironment that involved the induction of IL-10-driven regulatory mechanism mediated by CD11b+F4/80hiCD64+CX3CR1+ mononuclear phagocytes, which facilitated bacterial colonization [24]. Additionally, in this paper, we reported that regulatory macrophages were involved in the process of colonization with H. pylori when we varied the epithelial cell proliferation parameter
in-silico. Zhang et al. [49] demonstrated that *H. pylori* directed active tolerogenic programming of DCs that favored chronic bacterial colonization, by altering the balance of Th17/Treg cells [49]. Rizzuti et al. [50] demonstrated *H. pylori*-mediated IL-10 release caused the activation of signal transducer and activator of transcription 3 (STAT3) in DC. This activation of STAT3 via IL-10 release was shown to induce the production of tolerogenic DC phenotype [50]. The findings from this paper also indicated the involvement of tolerogenic DCs in affecting the mucosal levels of *H. pylori*. Therefore, the literature combined with the results from this study, collectively suggest that during *H. pylori* infection, the epithelial cell favors the colonization of *H. pylori* by creating a regulatory microenvironment. This process is mediated by the regulatory macrophages and tolerogenic programming of DC. Based on the results from this paper and findings from the literature, this leads us to propose that the induction of IL-10 by the regulatory macrophages is potentially involved in directing the tolerogenic programming of DC. All experimental evidence combined with our model prediction suggest the action of an underlying biological mechanism that links the presence of *H. pylori* in the gastric mucosa with changes in the rates of epithelial cell proliferation which ultimately affects the levels of colonization. Our prediction points towards a link between epithelial cell proliferation and the action of tolerogenic dendritic cells and regulatory macrophages. The exact cellular mechanism induced during this process however cannot be inferred from the current model and it will be investigated in follow-up in vivo studies.

At its current stage, the hybrid ENISI model reproduces the overall immune system dynamics observed during an *H. pylori* infection. The parameters of calibrated ODEs were kept unchanged, whereas the ABM parameters were calibrated by qualitatively matching the patterns of the output simulations as observed in an in vivo model of *H. pylori* infection [24]. For ABM, its calibration and validation remain the major key issues, discussed elsewhere [21] [51] [52]. Further, developing targeted methods of SA have been identified as an important challenge in the field [21, 53, 54]. In this paper, we highlighted the use of SA methods with a two-stage global SA framework comprised of first, screening the input parameters (using PRCC) and second, building of a surrogate model (using GP) of the hybrid model, to understand the emergent behavior of the represented system. It is important to note that each *SA* method known, has its own merits and produces useful information however none provide a complete picture of the emergent model behavior [21]. First, we employed PRCC methods as the initial step in our two staged *SA* that aided the screening of active inputs and reduced the parameter space. The choice of PRCC was advantageous and justified by the low computational cost and low complexity in the computation of the coefficients. Another advantage of the regression-based PRCC method is that the complex output from our hybrid model was condensed into a descriptive relationship that can be described by statistical measures such as $R^2$ [21]. As described in [21] the results from PRCC are good descriptors of the outputs produced if the regression function constitutes a good fit to the output [21]. However, if the function does not yield a good fit, the regression-based *SA* are proven to be useful in screening the influential parameters for further analysis [21], as described in our analysis.

Furthermore, the interaction effects between the parameters are not considered in regression-based methods, and hence it was followed by the use of variance-based methods in later stage analysis. Second, we employed metamodeling-based approach and Sobol’ method as they provided information on the interaction between the input variable and the use of metamodels allowed to compute the sensitivity indices. One of the advantages of the Sobol’ method is that it is model-free and no fitting functions are used to decompose the output variance [32]. It considers the averaged effect of parameters over the whole parameter space but fails to explore the different patterns within the space [21]. Further, the method is not suitable for quantification of output
variability if the output distributions deviate from a normal distribution [21]. The detailed comparison of different SA methods used for the global SA of ABMs are described in detail in [21]. Thus, we performed both the PRCC and computation of Sobol’ indices approaches to evaluate the influence of the input parameter variation and identified the parameters involved in the successful colonization of the gastric niche by *H. pylori*.

Some limitations of the model include implementation through a two-dimensional grid system and including all cells of the same size. Although we parallelize the computation of the hybrid model output, the large number of simulations required for the global SA compensates for the benefits of parallelization. To improve the calibration process and overall usability of the model, the data required for model calibration would include tissue biopsies from people infected with *H. pylori* that can be used to quantify the cells and take into account their spatial arrangement. The current version is also limited in terms of the interactions that are based on epithelial cells and DC as they are strictly rule-based. The building of ODE models for these cells and integrating them with the ABM model will help capture the dynamics of epithelial cells and DC more in-depth. Overall the immunoregulatory mechanisms underlying the chronic colonization of *H. pylori* and the predictive capacity of the model can be further improved by incorporating cell-specific models for epithelial cells and DC.

In summary, a high-resolution, hybrid, multiscale spatiotemporal stochastic model of *H. pylori* infection was built and global SA was performed. The results from the global SA highlight the key role played by epithelial cells in affecting the levels of *H. pylori* colonization. The in-silico validation of varying the epithelial cell proliferation parameter demonstrated the involvement of regulatory macrophages and the tolerogenic DC. The next steps aimed to enrich the model will involve the validation of the findings in vivo to study the underlying mechanisms involved in the successful immune evasion by *H. pylori*. The computational modeling predictions will be further validated experimentally and clinically.

### Potential Implications

The computational model of the gut contains high-resolution information processing representations of immune responses that are generalizable for other infectious and autoimmune diseases. Complex diseases such as autoimmune disorders, infectious diseases, and cancer all require integration of the multiscale level data, information and knowledge, ranging from genes, proteins, cells, tissue to organ level. The ENISI model of the gut presented here can be generalized to other diseases by implementing the agents and rules specific to that disease, plus recalibrating the model based on data that are specific to the new indication. Since ABMs have modular architectures, an addition of new agent-types and modification of rules can be done without restructuring the entire simulation setup [19]. The use of ABM in such hybrid models not only facilitates the implementation of already known mechanisms but also helps validate and predict any unforeseen new mechanisms using data analytics methods such as global SA to analyze emerging behaviors at the systems level. The finer details regarding intracellular and intercellular interactions that contribute towards the nonlinear and complex behavior of the gut can also be studied by integrating the intracellular ODE models as implemented here.

### Supporting Files

The data sets and files supporting the results of this article are available in the ENISI-MSM GitHub repository, RRID: SCR_016918 [https://github.com/NIMML/ENISI-MSM](https://github.com/NIMML/ENISI-MSM). Further data
supporting this work and snapshots of our code are available in the *GigaScience* repository, GigaDB [55].

**Availability of source code and requirements**

- **Project Name:** ENISI MSM
- **Project homepage:** [https://github.com/NIMML/ENISI-MSM](https://github.com/NIMML/ENISI-MSM)
- **Operating system(s):** Linux, Mac OSX
- **Programming language:** C++, R, MATLAB
- **Other requirements:** CMake 3.7.2, ENISI Dependencies [https://github.com/NIMML/ENISI-Dependencies](https://github.com/NIMML/ENISI-Dependencies)
- **License:** Apache License 2.0
- **RRID:** SCR_016918

**Additional Files**

**File S 2.1** Instruction to Install ENISI MSM (Step I), Run a simulation (Step II) and Conduct Sensitivity Analysis (Step III).

This file contains the detailed instruction to **Install** ENISI MSM (Step I), **Run** a simulation (Step II) and **Conduct Sensitivity Analysis** (Step III). The jupyter (.ipynb) notebooks (*Fig2-Code.ipynb, Fig3-Code.ipynb, Fig4-Code.ipynb* and *Fig5-Code.ipynb*) in the GitHub repository include detailed instructions on how to create the specific figures presented in the paper.

**A. How to install ENISI MSM**

1. Create a folder for the hybrid computer model: `mkdir ENISI`
2. Change directory to the newly created folder: `cd ENISI`
3. Clone the dependencies required from the ENISI-Dependencies from the NIMML GitHub repository -
   1. `git clone --recursive https://github.com/NIMML/ENISI-Dependencies`
4. Change the path to the ENISI-Dependencies folder: `cd ENISI-Derpendencies`
5. Create a directory build within the folder: `mkdir build`
6. Change directory to the directory created in step 5: `cd build`
7. Start the installation: `cmake ./`
   
   `make`
8. Change the directory `cd`
9. Change the directory to the one created in step 1: `cd ENISI`
10. Clone the ENISI-MSM model from the NIMML GitHub repository –
    1. `git clone --recursive https://github.com/NIMML/ENISI-MSM`
11. Change the directory to ENISI-MSM: `cd ENISI-MSM`
12. Create a directory build within the folder: `mkdir build`
13. Change the directory to the directory created in step 12: cd build
14. Start the installation:
   cmake -DENISI_MSK_DEPENDENCY_DIR=PATH TO ENISI-Dependencies FOLDER/install ..
   make

B. How to run a simulation

1. Create a folder FolderName to save the simulation results. It is important to place all the results of every experiment and its respective files in different folders.
2. Place the files i) config.props ii) run.props iii) job.sh (required only if running on cluster) iv) CD4.cps v) MregDiff.cps vi) model.props all in the folder where you want the output files to be saved (i.e FolderName).
3. model.props is the parameter file wherein you can change the parameters.
4. run.props and config.props are the configurable files where you can change the number of TICKS (that is a measure of computational time, i.e stop.at = number of TICKS) and the size of the grid (in the current model that is set to 1nm).
5. For running locally, use run.sh
6. To run on a cluster, use job.sh.
7. For the -output folder path, change the CONFIG variable and provide path to your folder i.e /home/username/FolderName.
8. ENISI executable to be used in the job.sh file is located in /PATH: ENISI/ENISI-MSM/bin folder that is created in the (installation step, Section A).
9. Run your job by typing -> sh run.sh (OR) ./ run.sh "path of the folder where you want the results or sh job.sh (specify the CONFIG variable within).
10. After the job is completed, you will have .log files, .tsv files for all the compartments.
11. The .log file will contain debugging statements if there are any issued in the code. Additional statements can be added to the source code for confirmation and monitoring the output.

C. Sensitivity Analysis

Stage 1 Initialization

1. Parameters.xlsx -> Contains the maximum and minimum values of the input parameters and information about which parameters are fixed.
2. Generate the Input parameter design matrix (P1) using - design_matrix_generation.m; (NOTE: Comment out the Stage 2 part of the code).
3. Each row in P1 corresponds to the different values of the parameters to be used in the model.props files.
4. Run the simulation using the hybrid computer code as described in Section B.
**Stage 1 Analysis**

1. Run the simulations (152 x 20 replicates) for each input parameter setting obtained from **P1** (see above, step 2 in the initialization stage).
2. Convert the data into `.csv file` format:
   a. 1st column: **time points information (i.e. Ticks)**,
   b. 2nd column **mean values** and
   c. 3rd column **standard deviations**

   All the information will be obtained from the ENISI-MSM output runs.
3. Run **Stage1-PRCC.ipynb** - Formats the data to be used for the PRCC analysis and calculates the PRCC coefficients. (The code generates a data frame with rows from the Parameters.xlsx file and average of the output obtained for that parameter setting in the last column).
4. Plot the PRCC graphs using **Stage1-PRCC_barplots.R**
5. Alternatively, use Fig3-Code.ipynb jupyter notebook to recreate the figures in the paper.
6. Create an excel sheet with information about the active and inactive inputs from **PRCC_ACTIVEINACTIVEINPUTS-ADDED.xlsx**.

**Stage 2 Initialization**

1. Generate the Input parameter design matrix (**P2**) using – i) **design_matrix_generation.m** (NOTE: Comment out the Stage 1 part of the code) and ii) information regarding the active and inactive inputs present in **PRCC_ACTIVEINACTIVEINPUTS-ADDED.xlsx**.
2. Run the simulation using the hybrid computer code as described in Section B.

**Stage 2 Analysis**

1. Run the simulations (115 x 20 replicates) for each input parameter setting obtained from **P2** (see above, step 1 in the initialization stage).
2. Convert the data into `.csv file` format:
   a. 1st column: **time points information (i.e. Ticks)**,
   b. 2nd column **mean values** and
   c. 3rd column **standard deviations**.

   All the information will be obtained from the ENISI-MSM output runs.
3. Combine all the outputs obtained from P2 and P1. (outputs obtained after running simulation for P1 from Stage 1, Section C and for P2 from Stage 2, Section C).
   Create folders for each of the cell (cells are represented as agents in each compartment) populations and save the files from step 2, Sage 2, Section C.
4. Run `Stage2-inputfilegeneration.m` and save the output as .mat file to be used to build a temporal metamodel.
5. Build a temporal metamodel using `Stage2-BuildTempMM.R` and save the output as .Rdata dataset.
6. Calculate the Sobol Indices using `Stage2-SA-temporal6tps.R`. The input to the code includes the .Rdata obtained from the previous step 6 (stage 2 Analysis, Section C) and the datasets obtained after running SobolIndex_data_generation.m.

**Fig S 2.1.** Design implementation of the hybrid multiscale model used to simulate Helicobacter pylori infection

![Diagram](image)

**Fig S 2.1.** Design implementation of the hybrid multiscale model.
The figure shows the class structure used in the ENISI MSM hybrid agent based-ODE model. Each group consists of an act() function that includes the implemented rule for each agent. The previously published ODE models for T cells and Macrophage are used to integrate in the ABM code.

**Table S 2.1.** Parameters
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Baseline (WT)</th>
<th>Min</th>
<th>Max</th>
<th>CD4 Cre</th>
<th>LysMC re</th>
<th>Clodronate</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>p_epiinfbacterdamaged</td>
<td>Epithelial cell damage due to infectious bacteria</td>
<td>0.9</td>
<td>0.3</td>
<td>1</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_epith1damaged</td>
<td>Epithelial cell damage due to Th1 cells</td>
<td>0.8</td>
<td>0.3</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_epith17damage</td>
<td>Epithelial cell damage due to Th17 cells</td>
<td>0.8</td>
<td>0.3</td>
<td>0.9</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_EpiProliferation</td>
<td>Epithelial cell proliferation</td>
<td>0.6</td>
<td>0.2</td>
<td>0.9</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_EpiCellDeath</td>
<td>Epithelial cell death</td>
<td>0.6</td>
<td>0.1</td>
<td>0.9</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_epiIL10h</td>
<td>Epithelial cell healing due to IL10 cytokine</td>
<td>0.2</td>
<td>0.01</td>
<td>0.9</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_nTrep</td>
<td>Naïve T cell replication</td>
<td>0.022</td>
<td>0.01</td>
<td>0.5</td>
<td>0.022</td>
<td>0.022</td>
<td>0.022</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_naiveTcelldeat</td>
<td>Naïve T cell death</td>
<td>0.000025</td>
<td>0.0001</td>
<td>0.001</td>
<td>0.000025</td>
<td>0.000025</td>
<td>0.000025</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_allTrep</td>
<td>Replication of all the types of T helper cells</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_iTregtoTh17</td>
<td>Plasticity of induced Treg (iTreg) to Th17 cells</td>
<td>0.001</td>
<td>0.0001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_Th17toiTreg</td>
<td>Plasticity of Th17 to induced Treg (iTreg) cells</td>
<td>0.001</td>
<td>0.0001</td>
<td>0.001</td>
<td><strong>0.0001</strong></td>
<td>0.0001</td>
<td>0.001</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_nTtoTr</td>
<td>Differentiation of naïve T cell to type 1 regulatory T cell</td>
<td>0.4</td>
<td>0.04</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_nTtoiTr</td>
<td>Differentiation of naïve T cell to induced Treg cell</td>
<td>0.02</td>
<td>0.001</td>
<td>0.04</td>
<td><strong>0.001</strong></td>
<td>0.02</td>
<td>0.02</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_nTtoTh17</td>
<td>Differentiation of naïve T cell to Th17 cell</td>
<td>0.02</td>
<td>0.01</td>
<td>0.04</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>Per time step</td>
</tr>
<tr>
<td>Event Name</td>
<td>Description</td>
<td>Values</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Th1deadth</td>
<td>Th1 cell death</td>
<td>0.00001, 0.001, 0.03, 0.00001, 0.0001, 0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Th17death</td>
<td>Th17 cell death</td>
<td>0.03, 0.02, 0.1, 0.03, 0.03, 0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_iTregdeath</td>
<td>iTreg cell death</td>
<td>0.001, 0.001, 0.01, 0.001, 0.001, 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Trdeath</td>
<td>Type 1 regulatory T cell death</td>
<td>0.001, 0.001, 0.01, 0.001, 0.001, 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IL10Tr</td>
<td>IL10 release by type 1 regulatory T cell</td>
<td>1.2, 1, 2, 1.2, 1.2, 1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dummy</td>
<td>Dummy</td>
<td>0.3, 0.3, 0.6, 0.3, 0.3, 0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Bacteriakill</td>
<td>Bacteria removed due to inflammatory responses</td>
<td>0.5, 0.05, 0.6, 0.5, 0.5, 0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Bacteriadeath</td>
<td>Bacteria death</td>
<td>0.4, 0.04, 0.5, 0.4, 0.4, 0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_HPdeathdueToTcells</td>
<td>Helicobacter pylori death caused due to T cells</td>
<td>0.00001, 0.00001, 0.001, 0.0001, 0.0001, 0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_HPyloriDeath</td>
<td>Helicobacter pylori death</td>
<td>0.0001, 0.0001, 0.001, 0.0001, 0.0001, 0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_DCDeath</td>
<td>Dendritic cell death</td>
<td>0.0003, 0.0001, 0.003, 0.0003, 0.0003, 0.0003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Monocytedeath</td>
<td>Monocyte death</td>
<td>0.04, 0.01, 0.1, 0.04, 0.04, 0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_resmacdeath</td>
<td>Resident macrophages death</td>
<td>0.005, 0.001, 0.05, 0.005, 0.005, 0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Trmackill</td>
<td>Macrophage killed due to type 1 regulatory T cells</td>
<td>0.05, 0.01, 0.09, 0.05, 0.05, 0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_MregDiff</td>
<td>Differentiation of monocytes to regulatory macrophages</td>
<td>0.3, 0.1, 0.6, 0.3, 0.2, 0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_resmacrep</td>
<td>Proliferation of resident macrophages</td>
<td>0.2, 0.1, 0.6, 0.2, 0.2, 0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_monorep</td>
<td>Proliferation of monocytes</td>
<td>0.04, 0.01, 0.5, 0.04, 0.04, 0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IFNg</td>
<td>Release of IFNg</td>
<td>0.5, 0.1, 0.8, 0.5, 0.5, 0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Values</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IL10</td>
<td>Release of IL10</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IL17</td>
<td>Release of IL17</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IL21</td>
<td>Release of IL21 (not implemented, can be considered as dummy variable 2)</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IL6</td>
<td>Release of IL6</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_TGFb</td>
<td>Release of TGFb</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IL12</td>
<td>Release of IL12</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_nTrep</td>
<td>Naïve T cell proliferation</td>
<td>0.022</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_epicyto</td>
<td>Cytokine produced due to epithelial cells</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_TotalTc</td>
<td>Maximum amount of T cells allowed within a cell</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_random</td>
<td>Random damage of epithelial cell</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_nTtoTh</td>
<td>Differentiation of naïve T cell to Th1 cell</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_tcellmo</td>
<td>T cell movement</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Th17cyt</td>
<td>Release of cytokine from Th17 cells</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Th1cyt</td>
<td>Release of cytokines from Th1 cells</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_TroriTr</td>
<td>Release of cytokine from type 1 regulatory cell or induced regulatory T cell</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_BacteriaLPProl</td>
<td>Proliferation of bacteria in the lamina propria</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Bacteri aLumiProl</td>
<td>Proliferation of bacteria in the lumen</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_HPepit oLP</td>
<td>Movement of <em>H. pylori</em> from epithelium to lamina propria</td>
<td>0.125</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.125</td>
<td>0.125</td>
<td>0.125</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_Hpylori cap</td>
<td>Maximum amount of <em>H. pylori</em></td>
<td>30</td>
<td>Fixed</td>
<td>Fixed</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_HPylori rep</td>
<td>Proliferation of <em>H. pylori</em></td>
<td>0.01</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_iDCtoe DCLP</td>
<td>Differentiation of immature dendritic cell to effector dendritic cell in lamina propria</td>
<td>0.6</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_iDCtot DCLP</td>
<td>Differentiation of immature dendritic cell to tolerogenic dendritic cell in lamina propria</td>
<td>0.3</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.3</td>
<td>0.095</td>
<td>0.3</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_iDCtoe DCE</td>
<td>Differentiation of immature dendritic cell to effector dendritic cell in epithelium</td>
<td>0.6</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_iDCtot DCE</td>
<td>Differentiation of immature dendritic cell to tolerogenic dendritic cell in epithelium</td>
<td>0.5</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>Per time step</td>
</tr>
<tr>
<td>p_iDCrep</td>
<td>Proliferation of immature dendritic cell</td>
<td>0.5</td>
<td>Fixed</td>
<td>Fixed</td>
<td></td>
<td></td>
<td></td>
<td>Per time step</td>
</tr>
<tr>
<td>p_eDCcyt o</td>
<td>Release of cytokine from effector dendritic cell</td>
<td>0.9</td>
<td>Fixed</td>
<td>Fixed</td>
<td></td>
<td></td>
<td></td>
<td>Per time step</td>
</tr>
<tr>
<td>p_tDCcyt o</td>
<td>Release of cytokine from tolerogenic dendritic cell</td>
<td>0.9</td>
<td>Fixed</td>
<td>Fixed</td>
<td></td>
<td></td>
<td></td>
<td>Per time step</td>
</tr>
<tr>
<td>p_iDCmoveLPtoEp i</td>
<td>Movement of immature dendritic cells from lamina propria to epithelium</td>
<td>0.6</td>
<td>Fixed</td>
<td>Fixed</td>
<td></td>
<td></td>
<td></td>
<td>Per time step</td>
</tr>
<tr>
<td>p_iDCmoveEpitoLP</td>
<td>Movement of immature dendritic cells from epithelium to lamina propria</td>
<td>0.6</td>
<td>Fixed</td>
<td>Fixed</td>
<td></td>
<td></td>
<td></td>
<td>Per time step</td>
</tr>
<tr>
<td>p_DCLPtoGLN</td>
<td>Movement of (effector and tolerogenic) dendritic cells from</td>
<td>0.4</td>
<td>Fixed</td>
<td>Fixed</td>
<td></td>
<td></td>
<td></td>
<td>Per time step</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Value</td>
<td>Unit</td>
<td>Time Step</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------</td>
<td>-------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_DCEpit_oLP</td>
<td>Movement of (effector and tolerogenic) dendritic cells from epithelium to lamina propria</td>
<td>0.6</td>
<td>Fixed</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_DCbasal</td>
<td>Maximum amount of dendritic cells</td>
<td>10</td>
<td>Fixed</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_MinfDiff</td>
<td>Differentiation of monocyte to inflammatory macrophages</td>
<td>0.3</td>
<td>Fixed</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_resmaccyto</td>
<td>Release of cytokine from resident macrophage</td>
<td>0.9</td>
<td>Fixed</td>
<td>0.9 0.9 0.9</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Monocap</td>
<td>Maximum amount of monocytes</td>
<td>35</td>
<td>Fixed</td>
<td>35 35 35</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_resmacCap</td>
<td>Maximum amount of resident macrophages</td>
<td>35</td>
<td>Fixed</td>
<td>35 35 35</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Mregcyto</td>
<td>Release of cytokine from regulatory macrophage</td>
<td>0.9</td>
<td>Fixed</td>
<td>0.9 0.9 0.9</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Minfcyto</td>
<td>Release of cytokine from inflammatory macrophage</td>
<td>0.9</td>
<td>Fixed</td>
<td>0.9 0.9 0.9</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Hpregeomv</td>
<td><em>H. pylori</em> promotes the creation of regulatory environment</td>
<td>0.7</td>
<td>Fixed</td>
<td>0.7 0.7 0.7</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_EpiCap</td>
<td>Maximum amount of epithelial cells</td>
<td>20</td>
<td>Fixed</td>
<td>20 20 20</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_BacCap</td>
<td>Maximum amount of bacteria</td>
<td>20</td>
<td>Fixed</td>
<td>20 20 20</td>
<td>Per time step</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lamina_propria.eIL6_initialValue</td>
<td>Initial value</td>
<td>10</td>
<td>Fixed</td>
<td>10 10 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lamina_propria.eIL6_diffusion</td>
<td>Diffusion of IL6 cytokine</td>
<td>0.6</td>
<td>Fixed</td>
<td>0.6 0.6 0.6</td>
<td>mm$^2$/d day</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lamina_propria.eIL6_degradation</td>
<td>Degradation of IL6 cytokine</td>
<td>0.05</td>
<td>Fixed</td>
<td>0.05 0.05 0.05</td>
<td>day$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>Parameter</td>
<td>Value</td>
<td>Fixed</td>
<td>Value</td>
<td>Value</td>
<td>Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lamina_pr_opria.TGFb.initialValue</td>
<td>Initial value</td>
<td>50</td>
<td>Fixed</td>
<td>Fixed</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>lamina_pr_opria.TGFb.diffusion</td>
<td>Diffusion of TGFb cytokine</td>
<td>0.6</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>lamina_pr_opria.TGFb.degradation</td>
<td>Degradation of TGFb cytokine</td>
<td>0.02</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>lamina_pr_opria.eIL12.initialValue</td>
<td>Initial value</td>
<td>20</td>
<td>Fixed</td>
<td>Fixed</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>lamina_pr_opria.eIL12.diffusion</td>
<td>Diffusion of IL12 cytokine</td>
<td>0.6</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>lamina_pr_opria.eIL12.degradation</td>
<td>Degradation of IL12 cytokine</td>
<td>0.02</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>lamina_pr_opria.eIL17.initialValue</td>
<td>Initial value</td>
<td>0.4</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>lamina_pr_opria.eIL17.diffusion</td>
<td>Diffusion of IL17 cytokine</td>
<td>0.6</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>lamina_pr_opria.eIL17.degradation</td>
<td>Degradation of IL17 cytokine</td>
<td>0.02</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>lamina_pr_opria.eIL10.initialValue</td>
<td>Initial value</td>
<td>4</td>
<td>Fixed</td>
<td>Fixed</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>lamina_pr_opria.eIL10.diffusion</td>
<td>Diffusion of IL10 cytokine</td>
<td>0.6</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Value1</td>
<td>Value2</td>
<td>Value3</td>
<td>Value4</td>
<td>Value5</td>
<td>Unit</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------------------------------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td><strong>Degradation of IL10 cytokine</strong></td>
<td></td>
<td>0.02</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>day⁻¹</td>
<td></td>
</tr>
<tr>
<td><strong>Initial value</strong></td>
<td></td>
<td>4</td>
<td>Fixed</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Degradation of IFNγ cytokine</strong></td>
<td></td>
<td>0.02</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>day⁻¹</td>
<td></td>
</tr>
<tr>
<td><strong>Diffusion of IFNγ cytokine</strong></td>
<td></td>
<td>0.6</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>mm²/day</td>
<td></td>
</tr>
<tr>
<td><strong>Degradation of IFNγ cytokine</strong></td>
<td></td>
<td>0.02</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>day⁻¹</td>
<td></td>
</tr>
<tr>
<td><strong>Diffusion of IL6 cytokine in gastric lymph node</strong></td>
<td></td>
<td>0.6</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>mm²/day</td>
<td></td>
</tr>
<tr>
<td><strong>Degradation of IL6 cytokine in gastric lymph node</strong></td>
<td></td>
<td>0.02</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>day⁻¹</td>
<td></td>
</tr>
<tr>
<td><strong>Diffusion of IL6 cytokine in gastric lymph node</strong></td>
<td></td>
<td>0.6</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>mm²/day</td>
<td></td>
</tr>
<tr>
<td><strong>Degradation of IL6 cytokine in gastric lymph node</strong></td>
<td></td>
<td>0.02</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>day⁻¹</td>
<td></td>
</tr>
<tr>
<td><strong>Diffusion of IL12 cytokine in gastric lymph node</strong></td>
<td></td>
<td>0.6</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>mm²/day</td>
<td></td>
</tr>
<tr>
<td><strong>Degradation of IL12 cytokine in gastric lymph node</strong></td>
<td></td>
<td>0.02</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>day⁻¹</td>
<td></td>
</tr>
<tr>
<td><strong>Diffusion of IL12 cytokine in gastric lymph node</strong></td>
<td></td>
<td>0.6</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>mm²/day</td>
<td></td>
</tr>
<tr>
<td>gastric_lymph_nod_e.eIL12.degradatio_n</td>
<td>Degradation of IL12 cytokine in gastric lymph node</td>
<td>0.02</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>day⁻¹</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------------------------------------------------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>gastric_lymph_nod_e.eIL17.initalvalue</td>
<td>Initial value</td>
<td>5</td>
<td>Fixed</td>
<td>Fixed</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>gastric_lymph_nod_e.eIL17.idiffusion</td>
<td>Diffusion of IL17 cytokine in gastric lymph node</td>
<td>0.6</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>mm²/d ay</td>
</tr>
<tr>
<td>gastric_lymph_nod_e.eIL17.degradatio_n</td>
<td>Degradation of IL17 cytokine in gastric lymph node</td>
<td>0.02</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>day⁻¹</td>
</tr>
<tr>
<td>gastric_lymph_nod_e.eIL10.initalvalue</td>
<td>Initial value</td>
<td>4</td>
<td>Fixed</td>
<td>Fixed</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>gastric_lymph_nod_e.eIL10.idiffusion</td>
<td>Diffusion of IL10 cytokine in gastric lymph node</td>
<td>0.6</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>mm²/d ay</td>
</tr>
<tr>
<td>gastric_lymph_nod_e.eIL10.degradatio_n</td>
<td>Degradation of IL10 cytokine in gastric lymph node</td>
<td>0.02</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>day⁻¹</td>
</tr>
<tr>
<td>gastric_lymph_nod_e.eIFNg.initalValue</td>
<td>Initial value</td>
<td>4</td>
<td>Fixed</td>
<td>Fixed</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>gastric_lymph_nod_e.eIFNg.idiffusion</td>
<td>Diffusion of IFNg cytokine</td>
<td>0.6</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
<td>mm²/d ay</td>
</tr>
<tr>
<td>gastric_lymph_nod_e.eIFNg.degradation</td>
<td>Degradation of IFNg cytokine</td>
<td>0.02</td>
<td>Fixed</td>
<td>Fixed</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>day⁻¹</td>
</tr>
</tbody>
</table>
**Fig S2.2.** Time screenshots of a simulation of Helicobacter pylori infection modeled in a two-dimensional grid.

The thickness of the compartment is shown on the y-axis, such that: lumen spans (0 to 2) units, epithelium spans (2 to 3) units, lamina propria spans (3 to 8) units and gastric lymph node across (8-10) units on the scale. Two-dimensional distribution of different cell subsets over the time steps (ticks) 2, 4 (top panels), 5 and 6 (bottom panels) are shown. The insets in each image shows a zoomed in portion of the respective grids across the time steps 2, 4, 5 and 6. The agents represented in the screenshots below are only for visual representation and do not represent the actual size of the biological cells.
S2.3 Flowchart for the two-staged global sensitivity analysis.
### Fig S 2.4. Active and inactive inputs selected from the stage 1 analysis

<table>
<thead>
<tr>
<th>Name of parameters (Rows)</th>
<th>Cell Population (Columns)</th>
<th>Resident macrophages - LP</th>
<th>Monocyte derived macrophages - LP</th>
<th>Inflammatory macrophages - LP</th>
<th>Helicobacter pylori - LP</th>
<th>Th1-LP</th>
<th>Th17-LP</th>
<th>iTreg-LP</th>
<th>Tr-LP</th>
<th>Th1-GLN</th>
<th>Th17-GLN</th>
<th>Effector dendritic cells - GLN</th>
<th>Tolerogenic dendritic cells - GLN</th>
<th>Number of active inputs</th>
<th>Present at least once</th>
</tr>
</thead>
<tbody>
<tr>
<td>p_epiinfbactdamage</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>p_epith1damage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>p_epith17damage</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>p_EpiProliferation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>p_EpiCellDeath</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>p_epiIL10h</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>p_nTrep</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>p_naiveTcelDeath</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>p_allTrep</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>p_iTregtoTh17</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>p_Th17toiTreg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>p_nTtoTr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>p_nTtoiTreg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>p_nTtoTh17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>p_Th1death</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>p_Th17death</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>p_iTregdeath</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>p_Trdeath</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>p_IL10Tr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>dummy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>p_BacteriaKill</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>p_BacteriaDeath</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>p_HPdeathduetoTcells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Parameter</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_HPyloriDeath</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_DCDeth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Monocytedeath</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_resmacdeath</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_Trmackill</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_MregDiff</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_resmacrep</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_monorep</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IFNg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IL10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IL17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IL21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IL6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_TGFb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_IL12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sum (active inputs present at least once) = 23

1 – Active input
Empty – Inactive input

* Recommendation from experts to include

Fig S 2.4. The rows represent the input parameters and columns represent the output cell populations. The green boxes highlight the ‘active’ input parameters (row) that are shown to have a significant influence (calculated based on the results obtained from partial correlation coefficient analysis), on an output cell (columns) under consideration.
Fig S 2.5. Diagnostic and residual plots obtained for the Gaussian processes fitted metamodels

Fig S 2.5. The upper panel represents the diagnostic Q-Q plots where the open circles represent the cross-validated predictions; solid black lines represent observed response. The “observed simulations” data in the first half of the lower panel, refer to the observed output values of the simulations obtained after running the hybrid computer model, whereas the y axis refers to the predicted simulation values obtained from the Cross-validated model. Each point represents 1 output point obtained as an output from the simulation. The second half of the lower panel, refers to the standard residual plot wherein the x-axis represents the observed simulation values obtained from the simulation and the y-axis refers to the residual error ((error (predicted values – observed values) / standard deviation (error))) obtained. The diagnostic plots denote the black circles which are the cross-validated prediction. Cross-validation is in the sense that for predictions made at design point x, all observations at design point x are removed from the training set. The lower panel represents the residual plots for the cell populations –(a) Helicobacter pylori; (b) Resident macrophages; (c) Monocyte-derived macrophages in the Lamina propria and (d) Tolerogenic dendritic cells in the Gastric lymph node compartment.

List of abbreviations
ABM – Agent based model
DC – Dendritic cells
ENISI MSM – Enteric Immunity Simulator Multi-scale Modeling
GLN – gastric lymph node
GP - Gaussian process
H. pylori – Helicobacter pylori
HPC – High performance computing
LP – Lamina propria
ODE – Ordinary Differential Equation
PDE – Partial Differential Equation
SA – Sensitivity analysis
PRCC - Partial rank correlation coefficient
Competing interests
The authors declare that they have no competing interests.

Funding
This work was supported by the Defense Threat Reduction Agency (DTRA) grant HDTRA1-18-1-0008 to JBR and RH and funds from the Nutritional Immunology and Molecular Medicine Laboratory (www.nimml.org). The funding body had no role in the design of the study, data collection, analysis, interpretation of data and writing of the chapter.

Author contributions
MV, RH and JBR formulated the model, implemented, performed the simulations, analyzed model-generate outputs, made the figures and wrote the manuscript. MV, AL, JBR, RH, and SH formulated the model. SH, AL and VA implemented the code architecture and benchmarked the parallel version of the hybrid model. XC and MV wrote the codes for global sensitivity analysis and generated the design matrices. NTJ generated macrophage and H. pylori experimental data. JBR, VA, and RH supervised the project. JBR and RH edited the manuscript. JBR, AL, NTJ, SH, VA, XC and RH participated in discussions on the model and results. All authors provided critical feedback on the project.

Acknowledgements
We thank Dr. Luis S. Mayorga from IHEM (Universidad Nacional de Cuyo, CONICET) for his help in the visualization of the spatiotemporal simulation and critical feedback on the regression based sensitivity analysis.
References


Chapter 4: Modeling the Mechanisms by Which HIV-Associated Immunosuppression Influences HPV Persistence at the Oral Mucosa

This work was published in PLOS One

* denotes equal contribution

Abstract

Human immunodeficiency virus (HIV)-infected patients are at an increased risk of co-infection with human papilloma virus (HPV), and subsequent malignancies such as oral cancer. To determine the role of HIV-associated immune suppression on HPV persistence and pathogenesis, and to investigate the mechanisms underlying the modulation of HPV infection and oral cancer by HIV, we developed a mathematical model of HIV/HPV co-infection. Our model captures known immunological and molecular features such as impaired HPV-specific effector T helper 1 (Th1) cell responses, and enhanced HPV infection due to HIV. We used the model to determine HPV prognosis in the presence of HIV infection, and identified conditions under which HIV infection alters HPV persistence in the oral mucosa system. The model predicts that conditions leading to HPV persistence during HIV/HPV co-infection are the permissive immune environment created by HIV and molecular interactions between the two viruses. The model also determines when HPV infection continues to persist in the short run in a co-infected patient undergoing antiretroviral therapy. Lastly, the model predicts that, under efficacious antiretroviral treatment, HPV infections will decrease in the long run due to the restoration of CD4+ T cell numbers and protective immune responses.

Introduction

Infection with the human immunodeficiency virus (HIV) afflicts over 35 million people worldwide and results in impaired immune responses which may affect defenses against other pathogens. In the absence of protective vaccines, current management of HIV consists of administration of combination antiretroviral therapy (cART) - which suppresses viral replication and, consequently, drastically reduces morbidity and mortality [1, 2]. cARTs are 99% effective; however, antiviral drug resistance (mainly caused by non-compliance) against some of the anti-HIV drugs has been reported in up to 60% of the patients [3]. Moreover, a cure for HIV is challenging due to the strict need for lifelong treatment to avoid virus rebound from activation of latent reservoirs. Apart from the acquired immunodeficiency syndrome (AIDS), HIV increases the risk of developing opportunistic infections by other infectious agents, including viruses: papillomavirus,
herpesviruses, flaviviruses; and bacteria: Helicobacter pylori, Salmonella typhimurium, Chlamydia pneumonia [4]. Epidemiological data suggests that HIV patients have an increased risk for developing human papillomavirus (HPV)-induced cancers such as oropharyngeal cancer, cervical cancer, anogenital cancer and anal cancers [5-10]. However, the cellular and molecular mechanisms explaining the correlation between increased susceptibility of HPV-associated diseases and HIV-induced immune suppression remain largely unknown.

The oropharyngeal cancers are a subset of head and neck cancers (HNCs) which account for approximately 4% of all the cancers in the United States. The incidence of disease was twice as high in men than in women in 2015 [11]. The oropharyngeal cancers, in their clinically distinct form of squamous cell carcinoma, are commonly detected in HIV patients with a higher number of lifetime oral sexual partners, immunosuppression, smoking, and current tobacco use. The causal role of oral HPV infection is supported by substantial molecular and cellular evidence [5-7, 12]. Recent studies suggested that the interaction between HIV and HPV might be responsible for the increased risk of cervical cancers [13, 14]. Similarities in risk factors for the acquisition of HIV and HPV infections, such as high-risk sexual behavior, multiple sexual partners, and disease-related immunosuppression makes the demarcation between the HPV and HIV-associated malignancies challenging. It has been hypothesized that HIV patients who have been infected for a long period of time have a higher prevalence of oral HPV infection and subsequently are at a higher risk for HPV-associated HNCs [15]. HIV-induced immunosuppression also increases the risk of HPV-associated cancer. Other factors that can increase the risk of severe HPV infections in HIV patients include immune senescence, aging, impaired immune response to HPV, and direct interaction between the two viruses [13, 14, 16]. The immunological changes caused by HIV create a permissive immune environment, thereby decreasing the overall immune responses against HPV. However, the mechanisms by which HIV-induced reduction in CD4+ T cell levels impairs the immune response against HPV or other pathogens remain largely unknown.

Besides immunological factors, interactions at the molecular level between HIV genes tat, rev and vpr and HPV have also been reported. Multiple studies indicate an up-regulation of HPV oncogenic genes (E6 and E7) expression by tat [14, 17, 18]. Tat increases HPV shedding which suggests that HIV infection may contribute to the pathogenesis of HPV-associated disease by molecular interactions through tat [16].

Currently, there are two commercially available HPV preventive vaccines, the quadrivalent human papillomavirus vaccine (QHPV) (Gardasil; Merck) against HPV types 6, 11, 16 and 18 and a bivalent vaccine (Cervarix, Glaxosmithkline) against HPV types 16 and 18 [19]. Both vaccines are known to be 98-100% effective against HPV types 16 and 18 [20]. Recent studies demonstrated the efficacy of the anti-HPV vaccine Cervarix in HIV-infected-oral-HPV-negative patients where more than 90% of patients produced high antibody titers against HPV [21]. However, more data is needed to address cross-reactivity between the induced antibody and other HPV strains [22]. Moreover, additional efficacy trials of QHPV in HIV infected individuals are needed to properly determine the correlation between vaccines dose, timing and protection against all HPV genotypes in immunocompromised HIV patients. The mechanistic investigation of co-infection scenario is experimentally challenging. The disadvantages of interrupted treatment of HIV and limited efficacy for the HPV vaccine against all the HPV strains further adds to the complexity. The limited clinical information about treatment and prevention options against HPV in an HIV/HPV co-infection has lead us to an alternative mode of investigation.
To gain new insights into the underlying mechanistic interactions between HIV and HPV, in an HIV/HPV co-infection we developed a mathematical model of HIV and HPV interactions. Building on previous models of HIV [23] and HPV [24] single infections, this novel model captures the molecular interactions between HIV and HPV due to tat and the effect of progressive depletion of CD4+ T cell due to HIV infection. Using the model, we aim to investigate why the prevalence of oral HPV infection is increased in HIV-infected individuals. We demonstrate how the dynamics of HPV changes in an HIV/HPV co-infection when the patient undergoes combined antiretroviral therapy. The findings can be used to further advance our understanding of the mechanisms underlying oral immune plasticity. Lastly, modeling can help propose new hypotheses for reversing residual inflammation in individuals following the start of cART and guide clinical practice.

**Methods**

**Mathematical model of HIV infection**

We model the interaction between HIV and CD4+ T cells as in [23, 25]. Briefly, we consider the interaction between three populations: i) target CD4+ T cells (T), ii) productively infected CD4+ T cells (I), and iii) HIV (V). Target cells are produced at rate s, die at rate d, and become productively infected at rate β proportional to the interaction between target cells and the virus. Infected cells produce $N_1$ virions throughout their lifetime, which are released through bursting, and die at rate δ. The virus is cleared at a rate $c_1$ per day. The following system of ordinary differential equations (ODE) represents these dynamics:

$$\begin{align*}
\frac{dT}{dt} &= s - dT - \beta TV, \\
\frac{dI}{dt} &= \beta TV - \delta I, \\
\frac{dV}{dt} &= N_1 \delta I - c_1 V,
\end{align*}$$  \hspace{1cm} (1)

with initial conditions $T(0) = T_0$, $I(0) = I_0$, and $V(0) = V_0$.

The effect of cART has been modeled as a reduction of the virus infectivity in the presence of reverse transcriptase inhibitors to $\beta (1 - \varepsilon_{RT})$ and a reduction in the production of infectious virions in the presence of protease inhibitors to $N_1(1 - \varepsilon_{PI})$. Here $0 \leq \varepsilon_{RT}, \varepsilon_{PI} \leq 1$ are the drug efficacies [25, 26].

The model in the presence of cART becomes:

$$\begin{align*}
\frac{dT}{dt} &= s - dT - (1 - \varepsilon_{RT}) \beta TV, \\
\frac{dI}{dt} &= (1 - \varepsilon_{RT}) \beta TV - \delta I, \\
\frac{dV}{dt} &= (1 - \varepsilon_{PI}) N_1 \delta I - c_1 V,
\end{align*}$$  \hspace{1cm} (2)

with initial conditions $T(0) = T_1$, $I(0) = I_1$, and $V(0) = V_1$. Note that models (1) and (2) do not take into account the HIV latent reservoirs in the form of resting long-lived memory CD4+ T cells with integrated HIV in their genome [27].
Mathematical model of HPV infection

We model HPV in-host dynamics as in [24]. We consider the interaction between four populations: i) HPV infected basal epithelial cells ($Y_1$), ii) the HPV infected transit-amplifying cells, in the suprabasal epithelial layer ($Y_2$), iii) HPV ($W$) and iv) HPV-specific cytotoxic T lymphocytes (CTL) ($E$). We assume that $N_2$ is the total concentration of epithelial cells at the beginning of HPV infection and the basal layer is formed of uninfected basal epithelial cells, targeted by HPV. Upon HPV infection, the basal epithelial cells become infected $Y_1$ when HPV interacts with uninfected cells at rate $\psi$. We denote the difference $N_2 - Y_1$ as the concentration of uninfected epithelial cells. The basal infected cells, $Y_1$ traverse up through the epithelial column and transform into $Y_2$ cells, which move further into the suprabasal epithelial layer [28, 29]. The $Y_2$ cells become transit-amplifying cells which start assembling virions to be released at the surface [30, 31]. Therefore, both $Y_1$ and $Y_2$ cells are HPV infected cells but differentially located in the epithelial cell layer, wherein $Y_2$ cells are assumed to have higher expression of the oncogenes $E6$ and $E7$ compared to $Y_1$ [24]. For simplification, we assume that the uninfected cells and infected cells have an equal probability of interaction with the HPV virions irrespective of the spatial architecture of the tissue. A more generalized model which takes into account the infectivity and layer transition terms, or one which would consider spatial structures for epithelial cells in different layers requires extensive knowledge of numerous parameters, which are currently unknown. We assume that the infection is density dependent with $\varphi$ representing the uninfected cell concentration where the infection is half-maximal. We assume that infected cell populations $Y_1$ and $Y_2$ differ in terms of the oncogene expression such that the $Y_2$ (located in the suprabasal epithelial layer) have higher oncogene expression compared to $Y_1$ cells (located in the basal epithelial layer) [32]. The rate of oncogene expression of the HPV type present in an infected cell, given by $\varepsilon$ controls the conversion of $Y_1$, into the transit-amplifying infected cells, $Y_2$. Cells, $Y_2$, grow at rate $r\varepsilon$, proportional to their own density and die at rate $\mu$. Due to higher expression of oncogenes, the transit-amplifying cells, $Y_2$ divide more before death, compared to the basal infected cells $Y_1$. Since, both infected cell population have an expression of oncogene, as in [24], both types of infected cells produce free virions ($W$), at production rates $k_1$ and $k_2$ that are released through bursting. For simplicity, we consider an equal virion production rate of $k_1 = k_2 = k$. The HPV virions are cleared at rate $c_2$ [24]. The $c_2$ clearance rate captures the antibody clearance rate implicitly.

The clearance of HPV in the infected cells, is associated with a successful immune response that includes the trigger of innate immune responses targeted against the virions released from the surface as well as infected cells [31]. In addition to the innate immune responses, the HPV-specific CTLs recruited during the adaptive immune response aid in the elimination of the infected basal cells [33]. Here, we assume that, after encountering transit-amplifying infected cells $Y_2$, effector cells specific to HPV, $E$, expand with a maximum per capita rate $\omega$ and carrying capacity $K$. This carrying capacity is an addition to the original work [24]. In the current model, the CTL response $E$ is initiated only by $Y_2$ cells which have higher oncogene $E6$ expression [34] [35] compared to $Y_1$.

We disregard the differential CTL response against the infected cell populations and consider that HPV-specific CTL population $E$ kills both classes of infected cells at the same rate $\alpha$, since both infected cells populations express oncogenes $E6$ and $E7$ [31, 33]. Additionally, the model does not take into account the virus specific gene expression at any particular epithelial site. Finally, the functional differences in $E6$ and $E7$, which are major determinants of HPV pathogenicity between HPV types [36], are also neglected.

The following system of differential equations represents these dynamics:
\[
\frac{dY_1}{dt} = \psi W \frac{N_2 - Y_1}{\theta + N_2 - Y_1} - \varepsilon Y_1 - \mu Y_1 - a Y_1 E,
\]
\[
\frac{dY_2}{dt} = \varepsilon Y_1 + r \varepsilon Y_2 - \mu Y_2 - a Y_2 E,
\]
\[
\frac{dW}{dt} = \mu k (Y_1 + Y_2) - c_2 W,
\]
\[
\frac{dE}{dt} = \omega Y_2 E \left(1 - \frac{E}{K}\right),
\]

with initial conditions \(Y_1(0) = Y_{10}, Y_2(0) = Y_{20}, W(0) = W_0\) and \(E(0) = E_0\).

**Co-infection Model**

**The effect of tat protein.**
HIV-protein \(tat\), secreted from HIV-infected intraepithelial immune cells, is known to play an important role in the disruption of epithelial tight junctions, thereby facilitating the entry of HPV into the mucosal epithelium [37]. We model the \(tat\)-induced increased likelihood of HPV infection through increasing the total available epithelial cells from \(N_2\) to \(N_2(1+pV)\), where \((p)\) is the effect of \(tat\) protein secreted by an HIV virion \((V)\), given in equation (4). The term \((1+pV)\) incorporates the HIV-associated epithelial disruption as one of the major underlying mechanisms that increases the susceptibility of epithelial cells to the HIV/HPV co-infection [37].

\[
\frac{dY_1}{dt} = \psi W \frac{(1+pV)N_2 - Y_1}{\theta + (1+pV)N_2 - Y_1} - \varepsilon Y_1 - \mu Y_1 - a Y_1 E. \tag{4}
\]

**The effect of immunosuppression.**
In HIV-infected individuals, the loss of CD4+ T cells leads to consecutive loss of CD4+ T cell-mediated immune responses against other pathogens such as HPV. Naïve CD4+ T cells are depleted in mucosal tissues in all the stages of HIV infection [27, 38], and progressive decline of CD4+ T cells affects the differentiation process of naïve CD4+ T cells into the different subsets. Such a subset, Th1, is known to play a major role in immune responses against HPV [35] through induction of cell-mediated immunity in the presence of IL-2, IL-12 and IFN-\(\gamma\) cytokines [39].

To model the decrease in the availability of CD4+ T cell population due to HIV; and the subsequent effect of such loss on HPV-specific CTL \((E)\) responses, we assume that the carrying capacity of the \(E\) population decreases in an immunosuppressed patient. In particular, we represent \(K\) as the carrying capacity of CTLs and thus the maximum \(E\) population. We make \(K\) a function of CD4+ T cell population, such that \(K\) is given by, \(K = K(T) = b T\), where \(T\) are the uninfected CD4+ T cells in the model (1). When \(T\) decreases during the progressive loss of CD4+ T cells, \(T, K(T)\), the maximum carrying capacity decreases at a linear rate. We assume that the CTL carrying capacity is directly proportional to the amount of CD4+ T cells. Other modeling options, such as a T dependent source with a death term were explored, however the maximum proliferation term \(K(T)\) best explained the homeostatic mechanistic behavior of the CTLs.

The co-infection model becomes (see Fig. 3.1):
\[
\begin{align*}
\frac{dT}{dt} &= s - dT - (1 - \varepsilon_{RT})\beta TV, \\
\frac{dI}{dt} &= (1 - \varepsilon_{RT})\beta TV - \delta I, \\
\frac{dV}{dt} &= (1 - \varepsilon_{PI})N_1\delta I - c_1 V, \\
\frac{dY_1}{dt} &= \psi W \frac{(1+pV)N_2-Y_1}{\varphi+(1+pV)N_2-Y_1} - \varepsilon Y_1 - \mu Y_1 - a Y_1 E, \\
\frac{dY_2}{dt} &= \varepsilon Y_1 + r \varepsilon Y_2 - \mu Y_2 - a Y_2 E, \\
\frac{dW}{dt} &= \mu k (Y_1 + Y_2) - c_2 W, \\
\frac{dE}{dt} &= \omega Y_2 E \left(1 - \frac{E}{K(T)}\right),
\end{align*}
\]

(5)

with initial conditions \( T(0) > 0 \), \( I(0) > 0 \), \( V(0) > 0 \), \( Y_1(0) = Y_{10} \), \( Y_2(0) = Y_{20} \), \( W(0) = W_0 \) and \( E(0) = E_0 \) where \( t = 0 \) is the time of co-infection.

Fig 3.1. HIV HPV Diagram.
A diagram for the co-infection model (5). The left side of the figure represents the HIV dynamics wherein the interaction between target CD4+ T cells (T), productively infected CD4+ T cells (I) and HIV (V) are shown. The figure also includes the effect of reverse transcriptase (RT) and protein inhibitor (PI) (shown by red line - inhibition). The right side of the figure represents the HPV dynamics wherein the interaction between infected basal cells (Y_1), suprabasal transit-amplifying cells (Y_2), HPV specific (E) cells and HPV (W) are shown. The systems biology markup language (SBML) compliant network of interactions between HIV (V) and HPV (W) is created using CellDesigner [40] (S 3.1 Fig.).
The complete SBML compliant model was deposited in BioModels [41] and assigned the identifier MODEL 1605030001.

Results

Analytical results

Analytically, we can find a necessary condition for the HPV infection to be cleared (data in S 3.1 File). We assume that we have a chronically infected HIV subject who reached steady state values ($\bar{T}$, $\bar{T}$, $\bar{V}$). Under these conditions, the carrying capacity for population E becomes $\bar{K} = b\bar{T}$. System (5) reduces to:

$$\begin{align*}
\frac{dY_1}{dt} &= \psi W \frac{(1+p\bar{V})N_2-Y_1}{\varphi(1+p\bar{V})N_2-Y_1} - \varepsilon Y_1 - \mu Y_1 - aY_1 E, \\
\frac{dY_2}{dt} &= \varepsilon Y_1 + r\varepsilon Y_2 - \mu Y_2 - aY_2 E, \\
\frac{dW}{dt} &= \mu k (Y_1 + Y_2) - c_2 W, \\
\frac{dE}{dt} &= \omega Y_2 E \left(1 - \frac{E}{\bar{K}}\right).
\end{align*}$$

(6)

Then, HPV will clear when:

$$\frac{\psi k \mu (1+p\bar{V})N_2}{c_2 (\varphi(1+p\bar{V})N_2)} < \frac{(\varepsilon + \mu + aE)(-\varepsilon + aE + \mu)}{(\varepsilon + \mu + aE - \varepsilon)}$$

(7)

where $E$ is any CTL level (data in S 3.1 File). Biologically, this means that when the product between HPV infection rate and the HPV production rate (in the presence of HIV) is less than the combined effect of effector cells and natural death rate of HPV, clearance of HPV will be observed.

Numerical results

Using the co-infection model (5), we numerically simulated disease scenarios in order to understand the dynamics of HPV infection in a co-infected individual. A recent clinical trial has investigated the effect of HIV in HPV infection in the presence and absence of combination antiretroviral therapy [42]. The levels of oral HPV DNA in the co-infected patients, which was monitored for 24 weeks after the start of cART, remained elevated throughout therapy. To determine the possible mechanisms of HPV persistence, we investigate models (5) and (6) for the relative contributions of co-infection factors: tat, as given by $p\bar{V}$, and immunosuppression as given by $K(T)$.

Parameter values.

Parameter values from previously published studies are utilized here, as follows. Equilibrium values for HIV RNA per ml and HIV-specific uninfected CD4+ T cells per ml were reported in an HIV/HPV co-infection study to be $\bar{V} = 4.8x10^4$ virions per ml and $\bar{T} = 3.3x10^5$ cells per ml [42]. Since the patient is in a chronic HIV steady state, we derive $\bar{T}$, $\beta$ and $s$ from steady state conditions

$$\bar{T} = \frac{c_1 V}{N_1 \delta}, \beta = \frac{c_1}{N_1 \bar{T}} \text{ and } s = \frac{d\bar{T} + \bar{T}V}{\bar{T}}.$$ 

to be $\bar{T} = 2.4x10^3$ cells per ml, $\beta = 1.5x10^{-7}$ ml per cells per day and $s = 5.6x10^3$ cells per ml per day. The remaining parameters are summarized in Table 3.1.
Table 3.1. Parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s$</td>
<td>$5.6 \times 10^3$ cells ml$^{-1}$ day$^{-1}$</td>
<td>CD4+ T cell recruitment rate</td>
<td>See text</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$1.5 \times 10^{-7}$ ml virions$^{-1}$ day$^{-1}$</td>
<td>HIV infection rate</td>
<td>See text</td>
</tr>
<tr>
<td>$d$</td>
<td>0.01 day$^{-1}$</td>
<td>Uninfected CD4+ T cell death rate</td>
<td>[43]</td>
</tr>
<tr>
<td>$\delta$</td>
<td>1 day$^{-1}$</td>
<td>Infected CD4+ T cells death rate</td>
<td>[44, 45]</td>
</tr>
<tr>
<td>$N_1$</td>
<td>467 virions cells$^{-1}$</td>
<td>HIV burst size</td>
<td>[46]</td>
</tr>
<tr>
<td>$c_1$</td>
<td>23 day$^{-1}$</td>
<td>HIV clearance rate</td>
<td>See text</td>
</tr>
<tr>
<td>$\varepsilon_{RT}$</td>
<td>varied</td>
<td>Reverse transcriptase efficacy</td>
<td>See text</td>
</tr>
<tr>
<td>$\varepsilon_{PI}$</td>
<td>varied</td>
<td>Protease inhibitor efficacy</td>
<td>See text</td>
</tr>
<tr>
<td>$\bar{T}$</td>
<td>$3.2 \times 10^7$ cells ml$^{-1}$</td>
<td>Uninfected CD4+ T cells at equilibrium</td>
<td>[42]</td>
</tr>
<tr>
<td>$\bar{I}$</td>
<td>$2.4 \times 10^4$ cells ml$^{-1}$</td>
<td>Infected CD4+ T cells at equilibrium</td>
<td>See text</td>
</tr>
<tr>
<td>$\bar{V}$</td>
<td>$4.8 \times 10^5$ virions ml$^{-1}$</td>
<td>HIV at equilibrium</td>
<td>[42]</td>
</tr>
<tr>
<td>$N_2$</td>
<td>$10^4$ cells</td>
<td>Total concentration of epithelial cells</td>
<td>[24]</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>$10^6$ cells</td>
<td>Epithelial cell concentration for which infection is half maximal</td>
<td>[24]</td>
</tr>
<tr>
<td>$\Psi$</td>
<td>$0.0067$ cells virions$^{-1}$ day$^{-1}$</td>
<td>HPV infection rate</td>
<td>[24]</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.048 day$^{-1}$</td>
<td>Epithelial cell death rate</td>
<td>[24]</td>
</tr>
<tr>
<td>$r$</td>
<td>0.1</td>
<td>Transit-amplifying cells recruitment rate</td>
<td>[24]</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>varied</td>
<td>Oncogenic expression</td>
<td>See text</td>
</tr>
<tr>
<td>$\omega$</td>
<td>$10^3$ cell$^{-1}$ day$^{-1}$</td>
<td>CTL expansion rate</td>
<td>[24]</td>
</tr>
<tr>
<td>$K$</td>
<td>varied</td>
<td>CTL carrying capacity</td>
<td>See text</td>
</tr>
<tr>
<td>$a$</td>
<td>0.01 day$^{-1}$ cells$^{-1}$</td>
<td>CTL killing rate</td>
<td>[24]</td>
</tr>
<tr>
<td>$k$</td>
<td>1000 virions cells$^{-1}$</td>
<td>HPV burst size</td>
<td>[24]</td>
</tr>
<tr>
<td>$c_2$</td>
<td>0.05 day$^{-1}$</td>
<td>HPV clearance rate</td>
<td>[24]</td>
</tr>
</tbody>
</table>

We assume that the tat-effect, given by $p\bar{V}$, ranges between zero and 20, to account for up to a 20-fold increase in the target epithelial cell population due to co-infection. The immunosuppression factor, given by $K(\bar{T})=b\bar{T}$ ranges between $K(\bar{T})=35$ cells and $K(\bar{T})=1$ cell to account for changes in available CTL concentrations between an HPV infection and HIV/HPV co-infection. Lastly, the oncogene expression $\varepsilon$ ranges from zero to one.

The viral dynamics of HPV infected individuals.

We first study the dynamics of HPV infection in the absence of HIV, as given by model (6) with $\bar{I}=\bar{V}=0$ cells per ml, and $\bar{I}=10^6$ cells per ml. We let $\varepsilon = 0.5$ per day, $K(\bar{T})=35$ cells, $b=3.5\times10^{-5}$ and the other parameters are listed in Table 1. Under these assumptions, model (6) predicts HPV and CTL levels similar to those in [24]. In particular, HPV reaches a maximum of $1.4\times10^5$ copies at day 174 and eventual clearance (see Fig. 3.2, panel b, green solid line). The CTL expansion is delayed by 80 days, and reaches an equilibrium value of 27 cells by day 240 (see Fig. 3.2, panel b, purple dashed line). Transit-amplifying cells, $Y_2$ with high oncogenic gene expression, are 12-
times higher than cells with low oncogenic expression, $Y_1$ (see Fig. 3.2, panel a). This result is dependent on the oncogene expression rate $\epsilon$ (not shown).

![Graph](image)

**Fig 3.2. HPV infection.**
(a) Infected basal cells $Y_1$ (blue solid line) and suprabasal, transit-amplifying cells, $Y_2$ (red dashed line); (b) HPV W (green solid line) and CTL E (purple dashed line) for $\epsilon = 0.5$ per day and $K(\overline{T}) = 35$ cells. All the other parameters are listed in Table 3.1.

**The viral dynamics of HIV/HPV co-infected individuals.**
We start by assuming that the *tat* effect leads to the doubling of available target epithelial cells, *i.e.* $N_2(1 + p\overline{V}) = 2N_2$. We then account for HIV induced immunosuppression in an HIV/HPV co-infected individual, by changing $K(\overline{T})$ as follows. We have shown in the previous section that an HIV-naive individual has a CTL carrying capacity $K(\overline{T}) = 35$ cells, where $\overline{T} = 10^6$ CD4+ T cells per ml and $b = 3.5 \times 10^{-5}$. We keep the $b = 3.5 \times 10^{-5}$ and decrease the $\overline{T}$ number to i) $\overline{T} = 5 \times 10^5$ cells per ml, corresponding to average chronic HIV CD4+ T cell numbers [48]; ii) $\overline{T} = 3.3 \times 10^5$ cells per ml as in the HIV/HPV co-infection study [42]; and iii) $\overline{T} = 2 \times 10^5$ cells per ml, corresponding to AIDS.

Under these assumptions and parameters in Table 1, model (6) predicts HPV clearance in cases (i) and HPV persistence in cases (ii) and (iii). In case (i), HPV levels reaches a maximum of $2.4 \times 10^5$ copies at day 128 and clears by day 1050. In cases (ii) and (iii), HPV reaches steady state values of $3.5 \times 10^6$ and $6.7 \times 10^7$ DNA cells after 20 and 2.1 years, respectively (see Fig. 3.3, panel a). CTL levels decrease to 17.5, 11.5 and 7 cells per ml for cases (i), (ii) and (iii), respectively (see Fig. 3.3, panel b).
Fig. 3.3 HIV/HPV co-infection.
(a) HPV W and (b) CTL E as given by model (6) for $\varepsilon = 0.5$ per day, parameters are listed in Table 1, and $T = 10^6$ cells per ml (blue solid lines); $T = 5 \times 10^5$ cells per ml (red dashed lines); $T = 3.3 \times 10^5$ cells per ml (green dotted lines); and $T = 2 \times 10^5$ cells per ml (purple dashed-dotted lines).

To determine the relative contributions of the tat-effect and immunosuppression in the transition between HPV clearance and HPV persistence, we derived a bifurcation diagram showing the asymptomatic dynamic of HPV as given by model (6) when both $p$ and $K$ are varied. As expected, an increase in the available epithelial cells requires a larger CTL population for the clearance to occur (see Fig. 4, red dashed lines). In particular, if the tat effect is increased to 100% such that $(1 + p) = 2$, then the CTL carrying capacity has to be $K > 11.9$ cells for clearance to occur. Moreover, a carrying capacity as low as $K = 7$ cells is enough to ensure HPV clearance in the HIV-naïve case (80% lower than the considered base value of $K = 35$ cells).
Fig 3.4. Varying oncogene expression rates.
(a) Bifurcation diagram showing cleared $W$ (area below the curve) versus chronic $W$ (area above the curve) as the tat effect $pV$ and CTL carrying capacity $K(\overline{T})$ vary. Here, the criterion for HPV clearance is given by equation (7); (b) HPV $W$; and (c) CTL $E$ as given by model (6) for parameters listed in Table 1 and $\varepsilon = 0.1$ (blue solid lines), $\varepsilon = 0.5$ (red dashed lines), and $\varepsilon = 0.9$ (green dotted lines).

Changing oncogene expression rates.
We have considered that the oncogenic expression is $\varepsilon = 0.5$. In an HIV-naïve host, this corresponds to transit-amplifying cells, $Y_2$ exceeding the infected basal cells, $Y_1$ by 12-times (see Fig. 3.2, panel a). In the mathematical model from [24], the authors showed that in an HIV-naïve, HPV-unvaccinated individual, a decrease in the oncogenic expression $\varepsilon$ leads to a slower growth of $Y_1$, $Y_2$ and $W$, a delayed and weak CTL response $E$ and, consequently, a delayed HPV clearance.

To determine whether this effect is carried over in an HIV/HPV co-infected individual, we compared clearance regions for $\varepsilon = 0.1$ per day, $\varepsilon = 0.5$ per day and $\varepsilon = 0.9$ per day for varying $pV$ and $K(\overline{T})$ values (see Fig. 3.4, panel a). We find that the clearance regions (defined as the area under the curve) are higher for low $\varepsilon$ values, similar to the results from an HIV-naïve patient [24].

To determine the timing of clearance, we fixed the tat-effect to $(1+p\overline{V}) = 2$ and the CTL carrying capacity to $K(\overline{T}) = 17.5$ cells and determined the changes in $W$ and $E$ dynamics for various values of $\varepsilon$ (see Fig. 3.4, panels b and c). We find that HPV levels are slightly higher for high oncogenic expression, $\varepsilon$, and they take significantly longer to get cleared (see Fig. 3.4, panel b). This happens in spite of the fact that CTL levels grow faster for high oncogenic expression (see Fig. 3.4 panel c).

The effect of cART on an HIV/HPV co-infection.
A recent trial has investigated the dynamics of oral HPV DNA in HIV/HPV co-infected individuals undergoing cART. They found that 28% of the co-infected individuals had a persisting infection with at least one of the HPV genotypes. Moreover, 42% of co-infected individuals experienced either a persisting infection with the same genotype, or an acquired infection with a different genotype 24 weeks after the start of therapy [42]. We use model (5) to determine the tat effect $pV$, CTL numbers $K(T)$, and oncogenic expression $\varepsilon$ that can explain this observation.
The patients in [42] have $\overline{T} = 3.3 \times 10^5$ uninfected CD4+ T cells per ml and $\overline{V} = 4.8 \times 10^4$ virions per ml at day $t = 0$, when cART begins. We assume that the drug efficacies are $\varepsilon_{RT} = 0.95$, $\varepsilon_{PI} = 0.5$ and the oncogenic factor is $\varepsilon = 0.5$. If the co-infection with HPV is not included i.e. $pV = 0$ and $K(\overline{T}) = 35$ cells, then HIV RNA levels decrease to below limit of detection (of 50 copies per ml) in 6.5 days. CD4+ T cell concentration increases to a maximum of $5.6 \times 10^5$ cells per ml by day 329 (161 days after the end of the study).

We next add co-infection into our model and apply it to the setup of the oral co-infection trial [42]. If $pV = 1$ and $K(T) = 11.5$ cells (corresponding to CD4+ T cell concentration of $\overline{T} = 3.3 \times 10^5$ per ml), then HPV is cleared under the cART conditions $\varepsilon_{RT} = 0.95$ and $\varepsilon_{PI} = 0.5$. The timing of clearance depends on two factors: the HPV stage and the level of CD4+ T cells at the start of cART. If HPV infection occurs at the same time as the start of cART, then the HPV levels increase to a peak value of $1.4 \times 10^5$ DNA cells and stay elevated throughout the 24 weeks of the study (see Fig. 3.5, panel a). HPV starts to decay at day 180 (see Fig. 3.5, panel a, zoomed out graph) when CD4+ T cells reached $5.2 \times 10^5$ cells per ml (see Fig. 5, panel b) which is the low level of CD4+ T cell counts for healthy individuals.

In contrast, if HPV infection reached its chronic state at the start of cART, then cART helps to initiate HPV decay by day 8 (see Fig. 3.6, panel a), when the CD4+ T cell population is still low, i.e., $\overline{T} = 3.5 \times 10^5$ cells per ml (see Fig. 3.6, panel b). It is worth noting that cART removes HIV, and consequently the tat effect, but it does not control how fast CD4+ T cells rebound.

Fig 3.5. HIV/HPV dynamics when cART and HPV infection coincide.
(a) HPV W; (b) CD4+ T cells (T) as given by model (5) under cART. Here, $\varepsilon = 0.5$, $\varepsilon_{RT} = 0.95$, $\varepsilon_{PI} = 0.5$, and all other parameters are listed in Table 1. Initial conditions are $T_0 = 3.3 \times 10^5$ cells per ml, $I_0 = 2.4 \times 10^5$ cells per ml, $V_0 = 4.8 \times 10^4$ virions per ml, $Y_{10} = 1$ cells, $E_0 = 0.01$ cells, $Y_{20} = W_0 = 0$, and $t = 0$ is the start of cART. Over the first 24 weeks HPV persists (panel a), and in the long term HPV is cleared (zoomed out panel a).

In contrast, if HPV infection reached its chronic state at the start of cART, then cART helps to initiate HPV decay by day 8 (see Fig. 3.6, panel a), when the CD4+ T cell population is still low, i.e., $\overline{T} = 3.5 \times 10^5$ cells per ml (see Fig. 3.6, panel b). It is worth noting that cART removes HIV, and consequently the tat effect, but it does not control how fast CD4+ T cells rebound.
Fig 3.6. HIV/HPV dynamics under chronic HPV at the start of cART

(a) HPV W; (b) CD4+ T cells T as given by model (5) under cART. Here, $\varepsilon = 0.5$ per day, $\varepsilon_{RT} = 0.95$, $\varepsilon_{PI} = 0.5$, and all other parameters are listed in Table 1. Initial conditions are $T_0 = 3.3 \times 10^5$ cells per ml; $I_0 = 2.4 \times 10^5$ cells per ml; $V_0 = 4.8 \times 10^4$ virions per ml; $Y_{10} = 3.2 \times 10^3$ cells; $Y_{20} = 1.6 \times 10^4$ cells; $W_0 = 1.8 \times 10^7$ virions; $E_0 = 0.01$ cells, and $t = 0$ is the start of cART. Moreover, we found two instances when HPV infection stays chronic in the presence of cART, namely strong drug efficacy, $\varepsilon_{RT} = 0.95$ and $\varepsilon_{PI} = 0.5$, and AIDS level CD4+ T cells, $T \leq 1.7 \times 10^5$ cells per ml; and, inefficient drug therapy, $\varepsilon_{RT} = 0.2$ and $\varepsilon_{PI} = 0$ and intermediate CD4+ T cell levels $T \leq 2.6 \times 10^5$ cells per ml.

Lastly, we investigated how the dynamics of HPV infection in a co-infected individual undergoing cART change with the oncogenic expression $\varepsilon$. We found that HPV levels stay high throughout the first 24 weeks of cART, but are eventually cleared for all levels of oncogenic expression (see Fig. 3.7, panel a). This is due to the fact that the dynamics of CD4+ T cells are not affected by the oncogenic expression (see Fig. 3.7, panel b), and, in all cases, the patients return to healthy CD4+ T cell levels. HPV has lower peak levels but longer time until clearance for low oncogenic expression, $\varepsilon = 0.1$ (see Fig. 3.7, panel a, blue solid line) compared to high oncogenic expression $\varepsilon = 0.9$ (see Fig. 3.7, panel a, green dotted line).
Fig 3.7. Effect of oncogene expression rates and cART on HIV/HPV co-infections
(a) HPV W; (b) CD4+ T cells T as given by model (5) under cART. Here, ε = 0.1 (blue solid line); ε = 0.5 (red dashed line); ε = 0.9 (green dotted line), εRT = 0.95, εPI = 0.5, and all other parameters are listed in Table 1. Initial conditions are $T_0 = 3.3 \times 10^5$ cells per ml, $I_0 = 2.4 \times 10^5$ cells per ml, $V_0 = 4.8 \times 10^4$ virions per ml, $Y_{10} = 1$ cells, $E_0 = 0.01$ cells, $Y_{20} = W_0 = 0$ and $t = 0$ is the start of cART.

An intriguing finding in [42] showed higher frequency of HPV DNA in individuals with the strongest rebound in absolute CD4+ T cell count post cART [42]. To investigate possible underlying mechanisms explaining this unexpected finding, we considered two virtual patients: patient 1 has a rebound to $6.5 \times 10^5$ cells per ml CD4+ T cell count as in [41], and patient 2 has a rebound to $5.6 \times 10^5$ cells per ml. We further assume that patient 1 has high oncogenic expression level $\varepsilon = 0.9$ per day, and patient 2 has low oncogenic expression level $\varepsilon = 0.1$ per day. This increase in oncogene expression leads to higher HPV DNA production in patient 1 (see Fig. 3.8 panel a, solid blue vs dashed green line) in spite of its better cART outcome (see Fig. 3.8, panel b, solid blue versus dashed green lines).
Fig 3.8. Effect of CD4+ T cell levels on HIV/HPV co-infections.
(a) HPV W; (b) CD4+ T cells, T, as given by model (5) under cART. Here, $\varepsilon = 0.1$ (green dashed line), $\varepsilon = 0.9$ (blue solid line), $\varepsilon_{RT} = 0.95$, $\varepsilon_{PI} = 0.5$, and all other parameters are listed in Table 1. Initial conditions are $T_0 = 3.3 \times 10^5$ (green dashed line) or $T_0 = 4.5 \times 10^5$ (blue solid line) cells per ml, $I_0 = 2.4 \times 10^6$ cells per ml, $V_0 = 4.8 \times 10^4$ virions per ml, $Y_{10} = 1$ cells, $E_0 = 0.01$ cells, $Y_{20} = W_0 = 0$ and $t = 0$ is the start of cART.

We investigated the coinfection model using the full model from [24] and found that it gives similar results (data in S 3.2 File) to the coinfection model (6) and the differences are negligible (Fig. A in S 3.2 Fig. and Fig. B in S 3.2 Fig.).

**Discussion**

The model presented here is a mechanistic ordinary differential equation (ODE)-based model that studies the dynamical interaction between the host and two virus populations: HIV and HPV. The model is aimed towards determining the mechanistic interactions leading to HPV clearance or persistence in HIV/HPV co-infected individuals post cART, and increased risk of HPV infection in HIV infected individuals as reported clinically [42][49][15, 37]. Indeed, a recent study reported an increased prevalence of oral HPV infection in an HIV-infected cohort, where HPV DNA levels in the patients were not reduced following treatment with antiretroviral therapy [42]. This result is corroborated by other studies, suggesting that HPV may be present chronically in oral sites among HIV infected individuals undergoing combination antiretroviral therapy [50, 51].

To address the possible interactions leading to HPV persistence, we highlighted specific scenarios presenting an increased persistence of HPV due to the permissive immune environment created in an HIV-infected individual. Our model predicted that among the HIV infected individuals, those who had CD4+ T cells corresponding to average chronic HIV CD4+ T cell levels were more likely to clear HPV than those who had the CD4+ T cell levels reported in the coinfection clinical study [42], or those who had CD4+ T cell levels corresponding to AIDS. These results are dependent on the oncogenic expression levels, with HPV DNA levels increasing and taking a longer time to clear for high oncogenic expression levels. Interestingly, for high oncogenic
expressions, HPV clearance is delayed despite the faster expansion of CTL levels. This is due to an increase in the amount of HPV transit-amplifying cells.

We used the model to study the impact of cART on HPV persistence in HIV/HPV co-infected individuals and compared the findings with those from the clinical study [42], which showed that 28% of the co-infected patients had at least one detectable HPV DNA genotype 12-24 weeks after the start of cART. Our mechanistic model predicts that, for the median CD4+ T cell levels in [42], HPV levels decrease in co-infected individual receiving cART by 24 weeks. In addition, our model showed persistence of HPV DNA levels between 12-24 weeks post cART in the patients with the highest CD4 T cell rebound, as in [41]. The HPV DNA levels may be higher in the patients with the higher CD4 T cell rebound after cART if these patients have a higher HPV oncogenic expression. However, for these individuals with restored high levels of CD4+ T cells post cART, we predict clearance for HPV levels at 48 weeks. These findings, which differ from those in [42], can be explained by the absence of consideration for latent HIV reservoirs and latent HPV infection. Thus, reactivation of the latent HIV reservoirs and latent HPV infection is an integral part of immune reconstitution in the co-infected individuals and is guaranteed to impact HIV/HPV co-infection dynamics. Some studies have investigated the prevalence of anal human papillomavirus infection in HIV-infected patients receiving long-term cART. Piketty et al, showed that the incidence of anal cancer was higher in HIV-infected patients particularly in MSM (men who have sex with men) and the incidence of anal HPV infections did not reduce despite the increased CD4+ T cell count following cART [52], suggesting that cART- associated immune restoration does not play a role in reduction of the incidence of anal cancers [53]. Other studies have analyzed the effects of HIV and cART on HPV persistence and cervical squamous intraepithelial lesions. Blitz et al reported higher rate of acquisition and reduced clearance of oncogenic high-risk HPV types among HIV positive women. The study reported that cART improved clearance of high risk HPV type other than oncogenic HPV type 16/18 [54]. Interestingly, the findings from the study are in accordance with the findings from our model, wherein the oncogenic high risk HPV types persisted in the patients restored to high CD4+ T cell level post 24 weeks of cART.

Our study predicts that the timing of viral clearance is determined by the timing of cART compared to the timing of HPV co-infection, as well as the CD4 T cell level. When cART is started shortly after HPV infection, HPV will expand and will not be controlled in the 24 weeks of cART as described for the 28% cases in [42]. This can be explained by the fact that CTL expansion and control of HPV is delayed due to both recruitment and size limitation observed when CD4+ T cell reconstitution following cART is delayed. If, however, the cART starts at the peak anti-HPV CTL response, then HPV is cleared as soon as CD4+ T cell reconstitution allows for the adequate CTL levels to control HPV infection. Thus, our findings support the conclusion that increased risk of infection [49] due to immunosuppression may play a role in the reduced clearance of HPV. Specifically, we showed that chronic HPV levels were maintained at: i) AIDS level CD4+ T cell count and ii) inefficient drug therapy, $e_{RT} = 0.2$ and $e_{PI} = 0$ with intermediate CD4+ T cell levels such that $T \leq 2.6 \times 10^5$ cells. These results provide further evidence to support the findings from other studies that show immunosuppression plays a role prior to cancer diagnosis [48] and facilitates HPV related head and neck carcinogenesis [15,30]. The results from our model are independent of the molecular effects induced by tat, which are removed when HIV is removed, and therefore do not influence HPV clearance. Finally, our results show that when HPV infection occurs at the same time as the start of cART, HPV levels increase and stay elevated throughout
the 24 weeks of study. The findings support previous reports that predicted an increased risk of oral HPV infection among immunosuppressed individuals, which can be explained by the reactivation or reacquisition of a previously acquired infection [49]. We further investigated the possibility of HPV persistence and found that weak cART or AIDS level CD4+ T cells, which do not rebound to high enough levels, are needed for HPV to remain chronic following cART. These results highlight the important role played by CD4+ T cells in the resolution and control of HPV infection. It also demonstrates the importance of cART in controlling molecular mechanisms such as tat and CD4+ T cell rebound. A finding from our model highlights the need of a higher carrying capacity in a co-infected individual for HPV clearance to occur. This suggests a potentially clinically testable hypothesis that - if a HIV/HPV co-infected individual is immunosuppressed, then they should be treated for immunosuppression first. Once the CD4+ T cell levels are restored, the individual can be treated against HPV. Moreover, if a person treated for HIV has restored CD4+ T cell levels, HPV treatment should be tailored towards type specific HPV. In particular, our simulation result show that individuals with high CD4+ T cell levels post cART, produce more HPV when their oncogenic expression is high, compared with a patient with a weaker CD4+ T cell restoration but a lower oncogenic expression. Hence, the treatment (vaccine) against the type specific HPV, may aid the viral clearance.

This work is one of the first HIV/HPV co-infection models that investigates the dynamics of HPV in HIV-infected individuals. The only other published HIV/HPV co-infection model is a transitional probability-based model [55], which was used to study the relationship between immune status and the probability of the type of HPV clearance in HIV infected patients. The model [55] showed that HPV clearance was mainly based on the level of CD4+ T cell count. The main difference between our findings compared to [55] stems from the fact that the current model takes into account that HPV clearance not only depends on the levels of CD4+ T cell count but also the stage of HPV infection. Our results are consistent with observed associations between immunosuppression and HPV persistence in several clinical studies [42] [49] [15, 37].

To understand the mechanisms responsible for progression to AIDS and for CD4+ T cell rebound following cART, it is necessary to understand the impact of HIV in the dynamics of each functional CD4+ T cell subsets [56-59]. Th17 cells [60] [61] appear to play a central role in the HIV pathogenesis. Th1/17 CD4+ T cells have a role as a long-term reservoir for HIV-1 infection, and are unaffected by cART [62]. Tfh serve as reservoirs of virus-infected cells [63] in the lymph node and peripheral blood. The central memory (T_{CM}) are the major cellular reservoirs for HIV in the peripheral blood [64]. Furthermore, long-term cART is known to only partially restore the CD4+ T cell pool [65, 66] in the oral mucosal sites. Thus, investigating the T cell subsets involved in the HIV/HPV co-infection in our model would aid in better understanding of these mechanisms and the development of approaches for their therapeutic manipulations.

The present modeling study should be evaluated in the context of several limitations. First, it does not take into account the spatial structure of the epithelial tissue. A generalized model that takes in to account the infectivity and layer transition terms described by probabilities or one which would consider spatial structures for epithelial cells in different layers, requires extensive knowledge of numerous parameters, that are currently unknown. The second limitation concerns the immune clearance of lesions caused by HPV infections, which can lead to asymptomatic or latent infections with possibility of increased virion production upon immunosuppression [67-70]. This further necessitates the need for consideration of both the cellular environment and the site of infection which are important determinants of virus activity [36]. Third, the simplistic modeling
approach employed in the current modeling study does not take into account the feedback from the HPV to HIV infection. Due to absence of feedback from HPV to HIV, we disregard the effects of cART induced immune reconstitution. Additionally, our study does not consider the activation of latent HIV reservoirs post cART, which may contribute to the emergence new HPV genotype infections in co-infected individuals as shown in [42].

The next step would be investigating the T cell subsets involved in the HIV/HPV co-infection in our model. Additionally, the findings of the modeling study and associated limitations guarantees and necessitates inclusion of latent T cell reservoirs which are involved in the activation of residual cART induced immune reconstitution. Furthermore, incorporating the immune activation in the T cells under the effect of cART during a HIV/HPV co-infection would corroborate the findings regarding the presence of new detectable HPV DNA.

In summary, we developed a novel mathematical and computational model of HIV/HPV co-infection and used it to present hypotheses for the mechanisms underlying HPV persistence in HIV/HPV co-infected individuals. Our model can be applied to studying interactions between HIV and other widespread microbes to gain a better mechanistic understanding, guide the rational for the design of clinical trials, and accelerate the path to safer and more effective vaccines and therapeutics. We use this study as an alternative approach to determining how overall CD4+ T cell levels influence HPV prognosis in an HIV-infected individual. Overall, a better understanding of the cell specificity of HIV infection integrated with the cellular environment in HPV infections would facilitate the development of more effective therapeutic strategies in HIV/HPV co-infections. The model was deposited in BioModels [41] and assigned the identifier MODEL 1605030001.

Author contributions

MV, SE, VA, RH, SH, AL, JBR and SA contributed to model conceptualization. MV, SE, SH and SC contributed towards data curation, software and validation. MV, SE and SC contributed towards formal analysis and visualization aspects. MV, SE, VA, RH, AL, JBR and SC were involved in the writing, reviewing and editing process.

Acknowledgments

We thank Dr. Isaac Rodriguez-Chavez from National Institute of Dental and Craniofacial research who provided meaningful feedback on the co-infection model.

Supporting Files
Fig. S 3.1 Systems biology markup language (SBML) compliant network of interactions between HIV (V) and HPV (W) created using CellDesigner. The left side of the figure represents the HIV dynamics wherein the interaction between target CD4+ T cells (T), productively infected CD4+ T cells (I) and HIV (V) are shown. The figure also includes the effect of reverse transcriptase (RT) and protein inhibitor (PI) (shown by red line - inhibition). The right side of the figure represents the HPV dynamics wherein the interaction between infected basal cells (Y1), suprabasal transit-amplifying cells (Y2), HPV specific (E) cells and HPV (W) are shown.
(A)

(a) HPV DNA cells

(b) HPV DNA cells

(B)

(a) HPV DNA cells

(b) HPV DNA cells
Fig S 3.2. Full model comparison. (A) HIV/HPV co-infection model comparison. (a) HPV W and (b) CTL E as given by model (5), solid blue lines and model (26), dashed red lines, for $\varepsilon = 0.5$ per day, parameters are listed in Table 1 for different $T$ levels: $T = 5 \times 10^5$ cells per ml (first row); $T = 3.3 \times 10^5$ cells per ml (third row); and $T = 2 \times 10^5$ cells per ml (fourth row). (B) HIV/HPV dynamics when cART and HPV infection coincide. (a) HPV W; (b) CD4+ T cells ($T$) as given by model (5) solid blue lines and model (26) dashed red lines under cART. Here, $\varepsilon = 0.5$, $e_{RT} = 0.95$, $e_{PI} = 0.5$, and all other parameters are listed in Table 1 and $t = 0$ is the start of cART. Over the first 24 weeks HPV persists (panel a), and in the long term HPV is cleared (zoomed out panel a).

File S 3.2 Analysis of the coinfection model using the full model

We investigate the properties of the co-infection model using the large model developed in the original HPV work’s supplementary material from Murall et al 2015. The extended model includes the population of uninfected basal epithelial cells, $X$, which are born at rate $\lambda(t)$ and die at rate $\mu$. HPV, $W$, interacts with uninfected basal epithelial cells, $X$, at rate $\psi$ to produce infected cells, $Y_1$. We assume that the infection is density dependent with $\phi$ representing the uninfected cell concentration where the infection is half-maximal. Below is the expanded co-infection model wherein we included the effect of tat, $pV$, with the birth of uninfected basal epithelium. The term $(1 + pV)\lambda(t)$ accounts for the production of epithelial cells that are susceptible for HPV infection. The rest of the co-infection dynamics are described in the main text.

The expanded co-infection model becomes:

\[
\begin{align*}
\frac{dT}{dt} &= s - dT - (1 - \varepsilon_{RT})\beta TV, \\
\frac{dI}{dt} &= (1 - \varepsilon_{RT})\beta TV - \delta I, \\
\frac{dV}{dt} &= (1 - \varepsilon_{PI})N_1 \delta I - c_1 V, \\
\frac{dX}{dt} &= (1 + pV)\lambda(t) - \mu X - \psi W \frac{X}{\phi + X}, \\
\frac{dY_1}{dt} &= \psi W \frac{X}{\phi + X} - \varepsilon Y_1 - \mu Y_1 - aY_1E, \\
\frac{dY_2}{dt} &= \varepsilon Y_1 + r\varepsilon Y_2 - \mu Y_2 - aY_2E, \\
\frac{dW}{dt} &= \mu k(Y_1 + Y_2) - c_2 W, \\
\frac{dE}{dt} &= \omega Y_2E \left(1 - \frac{E}{K(T)}\right),
\end{align*}
\]

with initial conditions $T(0) > 0$, $I(0) > 0$, $V(0) > 0$, $X(0) > X_1$, $Y_1(0) = Y_{10}$, $Y_2(0) = Y_{20}$, $W(0) = W_0$ and $E(0) = E_0$ where $t = 0$ is the time of co-infection.

Since, tat is known to play an important role in the disruption of epithelial tight junctions, thereby facilitating the entry of HPV into the mucosal epithelium [30]. We compare the dynamics of model (26) against those of main model (5) for the same tat effect $(1 + pV) = 2$.
When considering HIV induced immunosuppression in HIV/HPV co-infected individuals with different CD4+ T cells levels of i) $\overline{T} = 10^6$ cells per ml, corresponding to a healthy patient ii) $\overline{T} = 5 \times 10^5$ cells per ml, corresponding to average chronic HIV CD4+ T cell numbers [41]; iii) $\overline{T} = 3.3 \times 10^5$ cells per ml as in the HIV/HPV co-infection study [35]; and iv) $\overline{T} = 2 \times 10^5$ cells per ml, corresponding to AIDS; the extended model (26) has similar results as the reduced model (5) (see (A) in Fig. S 3.2). Similarly, when we considered the effects of co-infection under the setup of the oral co-infection trial [35] for $P \nu = 1$ and $K(T) = 11.5$ cells (corresponding to CD4+ T cell concentration of $\overline{T} = 3.3 \times 10^5$ per ml) we found that both model (6) and (26) give similar results. In particular, HPV is cleared under cART conditions $\varepsilon_{RT} = 0.95$ and $\varepsilon_{PI} = 0.5$ and the timing of the clearance depends on two factors: the HPV stage and the level of CD4+ T cells at the start of cART (see (B) in Fig. S 3.2).
References


Chapter 5: Challenges in Personalized Nutrition and Health

This work was published in *Frontiers in Nutrition*


**Abstract**

Personalized nutrition innovation has the promise to help identify diet and food combinations that maximize health outcomes, quality of life and longevity at an individual and personalized level. Identifying optimal personalized dietary regimens requires capturing the information regarding individual human diversity and inter-individual variation. This information should be at a resolution sufficient to observe and clinically validate the predictive biomarkers of health and disease responsive to food. The challenges to making personalized nutritional recommendations for health promotion and disease prevention are discussed here. Integration of data-driven methods can facilitate the process. These include artificial intelligence and modeling integrated with large-scale medical records, real-time data collected through wearables and input on nutritional intake. The paper highlights that informed decisions by health practitioners, patients and consumers alike, requires successful implementation of novel personalized nutrition approaches through necessary data management, informatics, and real-time data analytics platforms.

**Introduction**

**Personalized nutrition and approaches employed**

Personalized nutrition refers to tailored nutritional recommendations aimed at the promotion, maintenance of health and prevention against diseases [1]. These recommendations take into account differential responses to certain individualized food-derived nutrients that arise due to the interaction between nutrients and biological processes. [2]. These include the interactions between internal factors such as genetics, microbiome, metabolome interactions as well as external factors such as dietary habits and physical activity [3]. In contrast to precision medicine defined by the Precision Medicine Initiative as an approach towards the treatment and prevention of disease for an individual, the goal of personalized nutrition is to promote the health and well-being through diet.

A balanced diet promotes good health as it provides adequate amounts of energy, proteins, vitamins, minerals, essential fats, micro and macronutrients for the metabolic needs of the body to function properly at each stage of the lifespan. The absence of balanced food and nutrition security leads to health problems such as diabetes, obesity, and malnutrition [4]. Although, the importance of nutrition and beneficial effects of food are well established, the mechanisms underlying their role in disease prevention or health benefits are incompletely understood [5, 6]. Further, there exists an inter-individual response to dietary intervention due to which a sub population may benefit more than others. This underlying variability can be attributed to genetics, age, gender, lifestyle, environmental exposure, gut microbiome, epigenetics, metabolism nutrition derived from diet and foods. The inter-individual variability to treatments and nutritional recommendations is largely reflected in biomarker values [7].
Reductionist approaches fail to demonstrate how the cellular and molecular responses due to food produce health benefits [6]. Current approaches used to study the inter-individual response to diet include –omics technologies such as genomics, metabolomics, proteomics integrated with the systems biology programs. These approaches are focused on integrating and analyzing complex datasets generated during dietary intervention association studies [3, 8, 9]. Systems biology approaches are impacting the field of nutrition [10-12] and immunology [13], however, significant challenges still remain in the translation and application of these advances to human studies [8]. A comprehensive systems-wide mechanistic understanding of the interplay between nutrition and health benefits requires the knowledge of network dynamics in the context of health, pre-disease and disease states. This requirement gives rise to the demand for new approaches and methods that could not only quantify the effects of dietary interventions in healthy individuals but also facilitate comparison to diseased patients [6].

Need for integrated personalized predictive models for use in personalized nutrition and health

To understand the underlying health dynamics while considering inter-individual variability and implementing personalized nutrition-driven interventions, efforts should focus on devising predictive methods that timely monitor the individual’s health responses to food. A systems science perspective can help physicians tailor targeted treatment, comprehend the variability in response to treatments and design personalized nutrition approaches [14]. Personalized nutrition approaches have the potential to spearhead the creation of information-processing representations of digestion, absorption and metabolism. These provide linkages between molecular events and health outcomes through: i) integration of data at all salient scales; ii) combination of multiscalar models with health outcomes through advanced machine learning (ML) models; iii) generation of non-intuitive hypotheses; and iv) experimental validation using preclinical and clinical trials with standardized nutritional interventions. With the advent of big-data era, data specific to consumption of standardized meal, functional food and beverage sales reports can be extracted. Health informatics enabled initiatives can be applied to conduct data mining and extraction from the electronic health records (EHRs) and insurance claims database. The EHR data can be combined with knowledge derived from nutritional and data sciences to build computational models and synthetic patient cohorts. These synthetic patients can be used as avatars that reflect inter-individual variation to preform predictive analysis and evaluate the system-level responses to the personalized food recommendations. These predictive insights can be utilized in order to elucidate the complex regulatory mechanisms of nutritional interventions at the interface of immunity, metabolism and gut microbiome. Overall, advanced computational methods and data analytics platforms could help shape the development of health platforms, tailor future nutritional recommendations for promoting health and accelerating the translation of the recommendations into the clinic.

Recent studies demonstrated successful application of providing personalized dietary advice at an individual level [15, 16]. Although the prediction method used by Zeevi et al demonstrated the effectiveness of personalized diet regimes to reduce levels of glucose, the results failed to connect to health outcomes. A web-based pan-European, Food4Me study [17] aimed to evaluate whether personalized advice caused more changes in dietary behavior as compared to “one size fits all”
An automated dietary feedback system was used to deliver personalized dietary advice and its comparison with manual system demonstrated complete agreement. The study demonstrated that personalized nutrition advice was more effective compared to a population-based nutritional advice.

These integrated data-driven approaches that build predictive computational models could be trained with additional features including phenotypic changes due to nutritional factors and changes due to interaction between genotypic and nutrient derived metabolic factors. The predictions from these models that include genotypic and metabolic features can aid the design of personalized diets based on the feature attributes that capture human diversity and variation. Thus, a unique comprehensive strategy that can automate data driven analytical model building, could be employed by focusing on the unique iterative integration of large-scale clinical record mining, -omic analysis, hypothesis-based modeling, simulation, and advanced ML approaches, to make tangible progress towards personalized nutrition and precision medicine. The vision for personalized nutrition has stimulated an immense interest for advancements in the diagnostics and decision support systems that allow continuous assessments of nutritional status. However, the advancement in the predictive technologies and their integration has posed numerous challenges. The focus of this review article is to put forth the major challenges (as shown in Figure 1) encountered in the process of revolutionizing the personalized nutrition health-care information technology including i) limitation in the reductionist approaches and opportunities for adoption of advanced computational data-driven technologies; ii) need for personalized nutrition computational infrastructure; iii) data standardization and the requirement for training individuals; iv) data sparsity, missing data and need for improved imputation methods. We further discuss the possible solutions for above listed challenges to achieve preventive, personalized and predictive approaches that aid the process of making decisions about diet and foods at an individual personalized level.
Figure 5.1. Challenges in personalized nutrition.
The challenges encountered in the path of making tailored recommendations toward personalized nutrition and health include – i) limitations due to the reductionist approaches that can be overcome by employing data-driven technologies such as AI and ML; adaptation to existing data-driven technologies raises, ii) the need for building a personalized nutrition computational infrastructure; the lack of standardization in format of the data utilized in electronic health records raises, iii) the need for data standardization and updated training programs for the users; the inconsistencies and missing values in the electronic health records results in, iv) the data sparsity and missing data problem that emphasizes the need for the development of new methods for data imputation.

Challenges

Challenge 1: Limitation in reductionist approaches and opportunities for adoption of advanced computational data-driven technologies

The study of biological mechanisms at a single gene or protein level in nutritional studies is largely outdated. The dynamic interaction among nutrition-metabolism-microbiota-host at the cellular and molecular level and health outcomes at the individual and personalized level are not completely understood [5, 6]. A holistic understanding of health requires evaluating the interactions among diet, genes, gene products, health and environmental exposures as opposed to focusing on the interaction of nutrients or macromolecules on specific gene or gene products. Current experimental techniques are limited in their ability to - i) allowing researchers to quantitatively manipulate diet in a controlled manner in humans and animals and ii) being able to trace events at the tissue level back to specific cellular and molecular level interactions and or signaling mechanisms. Data-driven approaches that utilize multi-parameter measures such as the influence of the nutrients on gene expression, genetic variations and interaction with environmental factors such as influence of lifestyle measures on gut microbiome interaction are capable of providing a comprehensive understanding of an individual’s health.

The current health care system evaluates static measurements of an individual’s health which includes the EHR data contents and physical examination results. In addition to the static measurements continuous measurements of health determinants such as the daily meal intake, microbiota composition, metabolomes, sleep and stress levels can allow for stratification of patients in sub-groups. An establishment of a baseline range for every parameter of a healthy individual is crucial [18]. The above listed multi-parameter measures will not only capture the dynamic relationship between the healthy parameter range of an individual but also aid in the process of early detection of the diseases and provide support in the decision-making processes. Advances in computational modeling and tools, data analytics methods and a systems science approach [4] can be employed to design, update strategies for nutrition-based health care and enhance disease preventive management.
The use of electronic health records, data from wearables and health apps integrated with individual specific variables - a “big data” approach

The EHRs are clinical repositories wherein longitudinal patient health information generated in any healthcare delivery setting is updated in real time. The data within are comprised of physiological measure outcomes, patient’s demographics, progress notes, past medical history, laboratory reports, medical prescriptions, radiology reports and administrative information. These data differ from those in the disease registries, claims records, or prescription databases and are specifically designed for patient care, billing purposes and prove to be important pertaining to health research [19]. Until 2011, the US healthcare system reached 150 exabytes of data and at this rate the healthcare big data are estimated to reach up to zetabyte ($10^{21}$ gigabytes) [20] level. These EHR repositories are sufficiently large and can be integrated with other -omics based databases to unravel phenotypic links between the data and other genetic risk factors [21]. The worldwide EHR adoption rates has increased and it is suggested that there will soon be a billion patient visits recorded yearly in the EHR systems [22] [23]. However, it is important to note that the amount of nutritional information in EHRs is limited. The collection of information regarding the daily dietary intake, meals and food content information and integration of consumer products with the massive amounts of clinical records already stored in the EHR systems could open up a new avenue for development of precision medicine and personalized nutrition pipelines. The wearable sensors integrated with the mobile technology have become increasingly popular. The real-time parameters recorded by these wearable sensors include physical activity, calories burnt and blood glucose levels that can also be leveraged to derive precise health outcomes for each individual. The derived data integrated with the EHRs could aid the process of designing automated data analytic pipeline to tailor personalized recommendation with real time input [24] [23]. With an emerging need for collection of real-time data, collaboration within the healthcare firms, insurers, and hospitals has become equally crucial. In addition exposome [25] and social determinants of health could be used to assist in the personalized nutrition based recommendations (as shown in Figure 5.2).
Figure 5.2. A pipeline for personalized nutrition and health. The figure represents the integration of the data derived from health determinants such as diet, gut microbiome, data from electronic health records, physical activity measures and data collected from wearable sensors can be used to train the artificial intelligence algorithms. The outputs can be used to make targeted personalized nutrition recommendations.

Use of advanced artificial intelligence methods

The healthcare sector generates a large amount of data that would promote wide studies in terms of risk evaluation, disease management, and patient stratification. The major drivers of the increase in U.S. data analytics market will include the initiative to increase adoption of EHR systems, availability of healthcare information technology big data infrastructure in healthcare, and technological advancements in genome sequencing. To derive predictive insights from patient data we would need to employ data-driven modeling because systems wide advanced computational methods and data analytics platforms can help in the organization, interpretation and pattern extraction from the health data [5] [6]. These modeling technologies would include the use of AI systems, ML classifiers, and mathematical models including differential equations (DEs), rules and agent based modeling [26, 27] [5, 6, 18, 28, 29].

With the wider application of AI methods large data repositories have ceased to be a data warehouse and have become true brains for information and knowledge extraction. Advanced ML approaches and AI are transformative technologies [30] that can be used to – i) develop synthetic patient populations from large-scale clinical data, ii) conduct in silico clinical trials to optimize clinical trial design and iii) compare the response to various treatment options and health outcomes. The application of ML techniques and AI technologies to build in silico pipelines with data analytics capabilities have the potential to tailor recommendations to achieve personalized nutrition and guide the design of human studies to improve success rate. These technologies are promising in terms of investigating the linkages between nutritional regimens and modulation of
whole genome-scale molecular signatures predictive of: i) health, ii) future health deterioration, iii) pre-disease and iv) diseased states.

During the adoption of the data analytics platform as a decision support system it is important to consider not only its scientific and functional capabilities but also the management and technology issues associated with it. The compatibility of EHRs software versions with the ML and AI technology is crucial. Further, it is important to have evidence based research regarding the success of the analytics platforms. Pilot and published studies that demonstrated improved outcomes due to the use of analytics platform (similar to the reporting process on drugs and medical devices) is required. The lack of the above listed points, were some of the reasons why the MD Anderson Cancer Center’s efforts to use IBM Watson cognitive computing system, in the clinical decision making process were put on hold (https://www.healthnewsreview.org/2017/02/md-anderson-cancer-centers-ibm-watson-project-fails-journalism-related/).

**Case study**

Application of advanced ML methods generated predictive insights in 10,000 virtual patients with Crohn’s disease [31]. Information on changes in the immunological parameters that drive response to treatment and knowledge based on experimental immunological insights were applied to create an advanced ML model. Leber et al developed a computational modeling pipeline to use preclinical data to test and predict the efficacy of existing and novel treatments against *Clostridium difficile* infection and make predictions about clinical outcomes [32, 33]. The computational pipeline included mechanistic ordinary DE based models [34] with stochastic simulations and an ensemble of advanced ML methods. Leber et al generated replicates through stochastic simulation of the model which are similar to virtual patients from seed populations of actual patients. This approach facilitated a complete coverage of parameter and response space [32, 33] as compared to responses obtained from averaging the data. The modeling pipeline examined the dosage effects and predicted synergisms or antagonisms of combination therapies [33]. Another data driven study in personalized nutrition [15] involved the metabolic phenotyping of 800 individuals and provided personalized dietary advice. This study was highly successful for clinically useful predictive modeling [15]. Zeevi et al investigated the intra and inter-individual variability in the glycemic response to standardized meals measured in an 800-participant cohort over 7 days. A ML algorithm was devised to predict the postprandial glycemic response and a combination of dietary, anthropometric, physical activity and gut microbiota data were used as input training dataset and the output was validated from the trained ML model in an independent cohort of 100 participants. Zeevi et al reported that different people have widely different post meal responses to the same standardized meal. Further, they provided individually tailored dietary advice to a new group of 26 individuals based on a ML derived insight or an expert opinion from clinical dietician and similar improvements in the glycemic response were reported in both the groups [15].

Another study conducted by Mathias et al. [35] evaluated how healthy Brazilian children and teens (aged 9-13) responded inter-individually to multi-micronutrients supplement intervention. Mathias et al. [35] took into account that all the individuals were genetically and environmentally unique and utilized a comprehensive approach wherein the aggregated data from all the individuals were analyzed along with the analyses of individual responses using a variation of n-of-1 trial design. The n-of-1 was designed such that each participant were their own control accounting for their inter-individual variability. Mathias et al assessed the effectiveness of
nutritional intervention by comparing the changes in –omics and clinical variables recorded at baseline, 6 weeks post intervention followed by a 6-week washout period. They employed an elastic net regression model and evaluated whether multiple variables such as the baseline clinical biochemistry, blood vitamin levels and dietary intakes could explain the variation in response to the intervention at each clinical endpoint [35]. The results demonstrated that multi-micronutrients mediated the physiology of the systems associated with metabolism. This result was based upon the response of the total cholesterol, LDL-cholesterol, mean corpuscular volume and circulating levels of nine vitamin metabolites to the 6-week intervention within a duration of 2 consecutive years [35].

Thus, these case studies demonstrated that ML architecture integrated with large-scale clinical data are capable of taking into account the inter-individual variability to an assigned diet and could help in moving away from the standardized approach of making general nutrition related recommendations. During a patient visit the integration of AI powered healthcare decision support systems could provide the clinicians with the EHR information, record of physical activity, microbiome information and aid in forming tailored health recommendations. The aim of utilizing the healthcare information technology driven framework will be used to provide a predictive, personalized and secured solution based on personal health record.

Integration of electronic health records with artificial intelligence-based methods and need for new biomarkers
Along with advances in ML models, deep learning has set a new trend [36] and advancements in computer vision has revolutionized clinical image analysis and widened the possibilities of complex tasks that the human brain can perform. For example, a deep convolutional neural network trained on thousands of clinical images of skin lesions outperformed dermatologists in the classification task of skin cancers [37]. Similar deep neural network trained on thousands of images outperformed ophthalmologists in the detection of diabetic retinopathy [38]. This demonstrates the capacity of advanced AI systems to be employed as smart clinical decision support system. Similar to the application of AI in interpretation of medical images, the AI technology can be leveraged towards tracking the nutritional contents in the daily meals [18] and the data can be connected to health outcomes measures. Food tracking is an essential component of disease management for patients with chronic metabolic disorders such as pre-diabetes, diabetes, obesity, metabolic syndrome or for people who aim to reach their recommended body mass index. The advent of mobile applications and wearable sensors can ease the tedious process of manual data entry by capturing an image of the meal that can be used to train a ML system along with input from EHRs. The ML system could be used for further processing such as the calorie content calculation [18, 39-41] that can aid the monitoring of food intake and change in health outcomes. One such study used a ML algorithm, trained the model with a dataset of images obtained from 23 different restaurants, and accurately determined the contents and calories in a meal [40]. However, it is important to note that these methods need to be fine-tuned and improved for meals that are highly complex in terms of multiple ingredients and spices [42]. In healthcare one of the most important components of AI system are physicians that drive and review the predictions regarding patient healthcare. The personal health data can be used to train the AI platforms and the prediction can be reviewed by the clinicians along with patient reports to aid in the decision-making process. After the doctors review the AI recommended suggestions they can then prescribe or make alternative dietary or medication suggestions. Once the doctor determines the recommendations, the AI system can be designed to send the prescriptions directly to patient’s
pharmacy and notify the patients in real time regarding any abnormalities or alternative medication suggestions.

Further, modeling systems that integrate the patient characteristics from different sources of data ranging from molecular to clinical cohort scales can aid the predictive \textit{in silico} testing of new nutritional interventions. These advanced analytics approaches can be employed to identify new molecular bio signatures as biomarkers which would be capable of defining individual responses, diversity and variation to a specific diet. The biomarker analysis tends to vary per individual especially based on age and other physiological factors. For instance, an increase in age changes the expression of gene or protein over time, for \textit{e.g.}, a disease biomarker for a younger adult will be different compared to an elderly. Although, there are available biomarkers for nutrition which includes the measure of essential dietary acids in plasma, protein intake, sodium and potassium levels [43], new biomarkers that would reflect the overall dietary pattern are needed. These new biomarkers and biomarker variation linked to dietary changes can be utilized as additional features for training the ML models to make accurate predictions of a personalized diet for an individual.

\textbf{Challenge 2: Need for personalized nutrition computational infrastructure}

In spite of nutrition research gaining enormous support due to the revolutionary -omics and data science era, there is an urgent need to deploy personalized nutrition computational infrastructure. When considering this it is important to take into account the challenges involved in the food intake databases. The food intake databases are comprised of the ability to capture complex eating patterns in an organized manner translating chemicals constituents in the food to intake of energy and nutrients. The existing tools for monitoring the food intake include the food diaries but those are challenging in terms of converting the food descriptions to the energy. An effort of moving towards an informatics infrastructure can be advantageous in terms of tailored nutrition databases that can create an environment of standard formats, annotations and network based systems to enhance food monitoring and intake processes [44]. However, differences can exist in terms of food description and methods used to collect the information. These include the methods used to generate compositional values when the food is collected from different sources such as the laboratory, patient or hospital-based sources.

Personalized nutrition and health research can benefit from building research infrastructures [45]. The development of personalized nutrition user-friendly platforms can enable the interrogation of data from different resources, multiple studies, research groups at different levels comprised of molecular, cellular, tissue, organ and population level. This personalized nutrition platform is required to take into account a systems-based approach for the identification of components involved in the human well-being and optimal health. The infrastructure build should enable the collection of information from each individual with a focus on n-of-1, as opposed to population-level data wherein an average response data limits its translatability [46]. It is known that the variation in response to nutritional factors can be explained by a set of identifiable factors such as genetic, environmental and behavioral factors that are specific to any individual [47]. The advent of self-monitoring devices that facilitate real-time recording of health data can facilitate this process. Although, most of the self-monitoring devices could potentially be affected due to personal bias and do not contain the scientific rigor of n-of-1 trials, they can be improved if the individuals are well informed and are made aware of the n-of-1 trial methodologies [47]. The outputs from n-of-1 studies are promising ways to advance individualized medicine and gain
insights regarding the comparative treatment effect among a group of patients [48]. The use of n-of-1 trails can play an important role in facilitating the process of making individualized diet recommendations as well. This would require the collection of data for one person every day or periodically over the duration of months or years. [49]. The data collected from these n-of-1 trials can be mined, analyzed for trends and pattern unique to an individual [47] and help determine personal risk factors that is otherwise not possible due to the averaging of data that may come from the analysis of a group of individuals. Such inclusion of n-of-1 data in the personalized nutrition infrastructure development can benefit researchers, stakeholders, clinicians and policy makers by providing access to individual health data and knowledge. The infrastructure can facilitate evidence based research by increasing access to the data derived from successfully implemented nutrition strategies [50]. The increased access to data can provide opportunities for transdisciplinary collaboration between industrial and academic institutions and help in creating public and private partnerships.

The core requirement of the infrastructure is that it needs to be classified and identified as a food and health infrastructure (FHI). The need for research infrastructures in the specific areas of food and nutrition area were recently highlighted by the EuroDISH project [45] which mapped the existing research infrastructures and identified the gaps. The project emphasized the need for infrastructure in the Determinants of food choice, Intake of foods and nutrients, Status and functional markers of nutritional health and Health and disease risk (DISH model) [50]. The management and implementation of the FHIs should be driven from the user level and comprise of i) nutrition bioinformatics structures including nutrition databases for e.g. the Nutritional phenotype database [44] that is designed to facilitate the storage of biologically relevant, preprocessed omics as well as descriptive and study participant phenotype data; ii) data management; iii) data processing; iv) data sharing capabilities and v) platforms for publishing the data derived from the studies to a bigger community for e.g. web portals. The FHIs should aim to develop methods related to dietary assessments of the food intake and accommodate the user input regarding the daily food intake in an electronic format. The FHIs should be adaptive such that the statistical effects of the nutritional interventions can be evaluated on a continuous basis since the earlier stages.

Nutrition quality is influenced by the environment, how and where the food is grown, transported and stored [46]. Additionally, the responses to diet, micro- and macronutrients differ and results from interactions of individual genetic makeup and the environment [51]. These interactions between nutrition and environmental factors emphasizes the need to collect a wide variety of measures of environmental variables across the globe known alter the health outcomes. The added information about environmental variables to the FHI databases will help in understanding how the interactions between diet and human and microbial genetic diversity is affected under the environmental influence. Further, the chemical constituents in the diet are known to alter the omics and microbial profile in humans and animal studies which emphasizes the need to include that information in the databases [46]. Therefore, the establishment of a FHI will ensure that the data related to food constituents, intake, environmental variables, health determinant data, energy expenditure and disease risk are all in one place. This FHI data can help reveal the determinants of behavior which can be utilized in the development of nutritional interventions [50]. The above mentioned added measures of environmental variables and the integration of data with prior knowledge of the health relevant interactions will ultimately aid novel hypothesis generation, interpretation and validation of results [46, 52]. It is important to note
that the standardization of collected data and usage of a FHI could facilitate the integration of an individual’s data from other settings including the outpatient, follow-up evaluations, and discharge [53]. Overall it will aid the understanding of personalized health and well-being, and predict the disease risk based on the environment, current eating habits and health status.

The utilization of data from already existing infrastructures that can be applied and extended to the food, nutrition and health interface is crucial. These include the application based infrastructures provided by LabKey (used to store laboratory based research projects) and REDCap (used to store patient oriented research) that can be utilized to build data linkages related to food and nutrition [5]. Adherence to data standards and quality control is essential for data sharing, integration, reproducibility and reduced query response times [5]. An ideal infrastructure for personalized nutrition should standardize data formats, use standard vocabulary and ontology. Therefore, it is crucial to develop data and software platforms that collect the food consumption data in a standardized format such as the EuroFIR [50]. Further, the infrastructure should be interoperable, regularly maintained and the tools and software need to be updated with new versions. Finally, the infrastructure build should provide technical support, documentation and training to facilitate its utility.

Challenge 3: Data standardization and the requirement for training individuals

Standardization of the data

The use of EHRs and FHIs could ease the accessibility to patient health records that contain information determining patient care. The transition of patient records from paper to electronic format maintained in EHRs is beneficial in terms of proper diagnoses based on the patient history and cost saving thereby giving rise to better patient outcome and healthcare decisions. However, digitization in electronic format can cause – i) improper standardization formats, ii) erroneous documentation of diagnostic codes resulting into incorrect reporting and denial from insurance companies, iii) lack of user engagement if not trained and additionally iv) poorly designed technology can lead to error and give rise to issues that can cause the insufficiencies within the EHRs [54]. Furthermore, the healthcare providers may report the diagnostic and billing codes inappropriately to ensure the insurance coverage for the services otherwise considered medically unnecessary by the medical insurance policy. Despite the worldwide adoption of EHR systems the processes of i) extracting the data and transforming into standard formats, ii) loading data from the EHRs and iii) reporting data to the billing department are not standardized [23, 55]. It is imperative that the organization works towards – i) standardizing the data formats, ii) ensuring transparent communication between the physicians and billers, iii) providing the billing staff access to provider’s documentation to investigate the diagnostic codes in case of discrepancy [54]. These practices would reduce the chances of medical errors and help maintain the regulatory standards. Also, the patient data sets in the records should be dynamic to facilitate the regular review and modification upon availability of additional information. Processing the data in a standard format would help the organization in the data retrieval process and is essential to streamline the process of clinical data collection to derive predictive insights. Another advantage of data standardization is that it will aid the communication across different hospitals, physicians, research institutions, and data scientists analyzing the data. The common format would ease the understanding of patient data emerging from different sources including the inpatient, outpatient or clinic and office visits for regular check-ups. Other advantages include consistent and easy identification of missing data information across different patients. Finally, data standardization across the healthcare
institutions will ensure the quality of patient health care and help keep track of their records facilitating evidence based recommendations. Although, there are some initiatives including the electronic medical records and genomics that mapped phenotypic information from the EHR data to standardized formats and mapped clinical data to single nucleotide polymorphisms [56, 57].

**Making sense of the data and training of the individuals**

With the emergence of the new data driven technologies, the demand for individuals trained at the interface between computational and clinical or translational approaches has risen exponentially. Application of variety data analysis skill sets are required to interpret large and complex databases which would not only include EHRs but also external data such as claims data, public repositories, and curated data sources such as ClinVar [58] among others. In spite of the availability of open source frameworks and tools, numerous challenges are encountered in the installation, configuration and administration of services included in the data analytics pipelines [20]. The training of the individuals with integrated data science skills knowledge of biology, nutrition, biomedicine, computer science, statistics and mathematics is required. The training curriculum for students, trainees and employees in the field of data science that is comprised of concepts from statistics, ML, bioinformatics, mathematics, and computer science needs to be regularly updated.

In 2014, under the *Modeling Immunity to Enteric Pathogens* project, a summer school and symposium on computational immunity was organized to provide experiential learning to individuals from varying backgrounds ranging from experimentalists, mathematicians, bioengineers, physicist, and nutritionists [5]. The participants learned how to employ computational tools and mathematical models to deepen their analysis of the experimental data and derive new non-intuitive novel insights. Furthermore, similar regular educational boot camps and training programs, should be organized for physicians, clinicians, nutritionists and dieticians who will directly interact with the patients. These workshops and training session will bring researchers from a diverse background and perspectives and aid in the improvement of the predictability of the responses [28]. The training of the new professionals should reflect the cutting-edge knowledge guided by the change in day to day informatics challenges [59]. The training, tutorial and boot camp sessions can aid the trainers by providing them an overview of the latest technology available, data standards and the methodologies used to use the infrastructure services [50].

Although, the AI technologies are revolutionizing it is crucial to understand that it is a means to an end (or tools to facilitate and enable decision making) as opposed to a replacement for the human experts. The most important part of the system is the ‘user’ that is required to make sense of large complex datasets. Humans and their expertise would be an integral part of the data knowledge discovery process. Thus, in order to utilize the advanced predictive capabilities of the AI system, a partnership of the experts with AI advanced capabilities is crucial. These can include experts in the fields of nutrition, bioinformatics, computer science, statistics, immunology, biochemistry, physiology, endocrinology, exercise science and mathematics. Further, since the predictive models are trained on data that needs to be cleaned before the analysis, domain specific knowledge experts are an inevitable part of the discovery process. The AI technologies have accelerated to a point where it should be accessible to all [60]. There are systems, such as the PennAI that provide a user-friendly interface ranging from uploading datasets, running ML analyses, visualizing the results in an intuitive manner and using the results to refine the knowledge derived. Thus, an availability of these interactive discovery environment can give access to ML technologies to health practitioners, healthcare providers and researchers that can aid the users in
understanding the data [60]. However, it is important to note that these web tools will not replace
the human interface.

**Challenge 4: Data sparsity and need for improved methods**
When handling clinical data pertaining to EHR, one of the main challenges is the high
dimensionality and sparsity of the data. The EHR store every clinical event during patient visit or
stay in the hospital. Zeevi et al examined the fluctuation in glucose levels in response to nutrition,
physical activity, and sleep; with a high-density data matrix on glucose levels that was available
through continuous glucose monitoring [15]. However, often times there are missing data in the
EHRs, the types of data available in EHRs are heterogeneous, complex and are in a mixed format
of structured and unstructured form. The structured data includes data entry in template
information such as patient’s demographic information, clinical measurements, drug prescription,
diagnoses whereas unstructured data included physicians handwritten notes. The reasons why the
data are missing are classified into three major categories: missing completely at random, missing
at random and missing not at random [61]. The missing completely at random arise when the
differences between missing and observed values are negligible, this could be due missed
measurements due to medical device breakdown. The missing at random arises when there is a
systematic difference between the missing and observed values and the predicted value may be
higher than actual measurement if the factors such as age or gender is not taken into account. The
data are classified as missing not at random when the patients miss the appointment due to ill
health or they expire. The duplications in the data can arise due to a patient experiencing the same
event multiple times and being prescribed the same drug. Thus, multiple challenges stem not only
from analyzing the EHR data but also from the sparsity and duplicated values in the available
information. Dealing with the missing data is important as the missingness can lead to biased and
misleading results. In cases where there are no missing values, regressions and principal
component analysis methods are used. The implementation of these methods removes variables
with missing values or remove patient datasets with missing values only when the number of
missing data patients is a small number. The employed estimation of missing value comprises of
deterministic methods such as mean or median imputation, K-nearest neighbors [62], and
stochastic methods such as the multiple imputation using chained equations process methods [63,
64] [62].

The imputation with mean and mean is easy to implement wherein the missing values are
substituted with the mean or median values from the distribution of data, however, that introduces
biases and large errors if the missing value belonged to the tail of the distribution [62]. In case of
the K-nearest neighbors, the values of the missing data can be estimated based on the values from
the individuals that are clustered together in a group [62]. The values from the grouped individuals
can be averaged and assigned to the missing variable. However, the K-nearest neighbors methods
may fail in cases where individuals cannot be well separated in groups based on their clinical
record values [62]. The stochastic method including the multiple imputation by chained equations,
is a framework for applying various imputation algorithms. The missing value of a variable for an
individual is imputed by considering the value of other observed variables within the individual
dataset, the relation between the variables and the value of the variable of interest observed in other
individuals. The process is repeated for number of iterations and the imputed values are used as
training set to update the estimates for second iteration [62]. Lastly, it is important to remember
that the stochastic processes are not free of biases as well. The assumption in multiple imputation
methods is that the data are considered to be normally distributed, thus excluding non-normally
distributed variables can add bias [61]. Further, multiple imputation methods are computationally intensive and the algorithms require the run length proportional to the volume of missing data [61]. The decision making process involved in selecting the data imputation method comprises of the data dimensionality, number of individuals, relationship among the variable, amount and pattern of missingness time and performance of a method [62].

Since, the high dimensionality and rich volume of EHR data are valuable for personalized nutrition research it is crucial to develop improved methods to deal with the missingness. The currently available imputation methods introduce bias errors and are computational intensive. Also, the required run time for the imputation methods increases with the volume of data. Along with improved methods, better documentation along with well-versed knowledge in statistical methods regarding missing data can help.

**Concluding Remarks**

An individual’s nutritional status can be determined by the integration of various factors including food intake, physiological health, diet and nutrition, -omics, metabolism, and physical activity measures. To make accurate personalized nutrition recommendations and accelerate the goal of better and health well-being, advanced computational technologies such as AI, ML and deep learning are promising in terms of providing an integrated framework. The use of data-driven methods will require the development of a personalized food and health infrastructure system comprised of advanced computational technologies with data storage, processing and sharing capabilities. The integrated and standardized infrastructure system will strengthen and enhance the patient care based on the collection of longitudinal data related to physiological measures, gut microbiome and other relevant biomarker measures. From legal and ethical consideration, it is important to take into who will have access to an individual’s personal data. It is important to protect the privacy of data and prevent discrimination in terms of eligibility for health insurance from the insurance companies, services from the hospitals and hiring decisions or terms from the employers. Overall, the standardized personalized nutrition framework approaches with protection of patient privacy can help establish preventive and predictive guidelines for promotion of health and better disease management.

**Conflict of Interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Author contributions**

MV, RH, VA and JBR designed the architecture of the review. MV, NT, RH and JBR performed the searches for the review. MV, VA, RH and JBR helped with the edits. MV and JBR wrote the manuscript.
Funding

None.

List of abbreviations

ML - Machine learning
AL - Artificial Intelligence
EHR - Electronic health record
FHI - Food health infrastructure

Acknowledgments

None.
References


60. Langley P. Lessons for the computational discovery of scientific knowledge. 2002.


Chapter 6. Conclusion

In this dissertation work, we developed computational models that can simulate host immune responses in two infectious diseases - i) *Helicobacter pylori* infection and ii) human immunodeficiency virus - human papilloma virus co-infection. The goal of this dissertation was to adopt a systems biology based approach methods to study and analyze the host immune responses in infectious diseases. The work presented here highlight the importance of modeling approaches that could lead to the identification of novel host immune factors in a predictive way.

In Chapter 2, we highlight the need for predictive systems modeling based methods to gain a comprehensive understanding of the complex interactions of a massively interacting systems such as the immune system. We discuss – i) equation-based methods, agent-based modeling methods, ii) tools developed in-house such as the Enteric Immunity Simulator Multi Scale Model (ENISI-MSM) and iii) the need to integrate these methods with advanced data analytic methods such as statistical and machine learning algorithms. The review highlights the potential of iterative systems biology cycle comprised of experimental techniques, data collection, computational model building, modeling prediction, sensitivity analysis and validation to gain a systems level understanding and guide the future precision medicine based health and wellness.

In Chapter 3, we developed a high-resolution model of the stomach to stimulate the immune responses against *H. pylori* infection using high-performance computing using a (HPC)-driven ENteric Immunity SImulator (ENISI)-multiscale. Equations-based methods such as ODEs modeled the intracellular mechanisms involved in signaling, cytokine production and differentiation of CD4 T cells and macrophages. Partial-differential equations were used to model and study the diffusion of cytokine values. ABM was used to model the location and interactions among the host cells and bacteria. A “hybrid” model combining the features of an ABM with equation-based methods (ODE/PDE/ABM) was developed to study the emergent patterns of the gut immune system during a *H. pylori* infection. Further, the work presented here focused on global sensitivity analysis (SA) and developed a spatiotemporal global SA method. A two-stage global SA method comprised of using a regression-based method to screen and reduce the input parameter space from 38 to 23 parameters. With this reduced parameter space, a spatiotemporal Gaussian-process (GP) based metamodel was constructed. A variance-based decomposition was performed and the Sobol’ indices were computed. The data analytics methods employed in this study identified epithelial cell proliferation as the most important input, which was required for successful colonization of *H. pylori* in the gastric niche. We validated the role of epithelial cell proliferation in silico and observed the involvement of regulatory macrophages and tolerogenic dendritic cells affecting the *H. pylori* levels in the gastric niche.

In Chapter 4, we developed a novel mathematical model of HIV/HPV coinfection that helped us gain insights on the interaction between the two viruses in a coinfection setting. This work was built using previously published models of HIV and HPV single infections, and it captured the progressive reduction of CD4+ T cells due to HIV infection as well as the molecular interactions between the two viruses due to tat. The findings from the model demonstrated how the dynamics of HPV changed when a co-infected patient underwent combined antiretroviral
therapy (cART). We observed that HIV infected patients with average chronic HIV CD4 T cells levels were more likely to clear HPV than the ones with AIDS CD4 T cells levels. The factors responsible for this include the oncogene expression level such that higher the oncogene expression level, the longer it takes to clear the HPV infection. Following combined antiretroviral therapy, cART, the model suggested improved clearance of HPV types other than oncogenic HPV types 16/18. The findings of the co-infection model highlighted that high risk HPV type persisted in infected individuals whose CD4 T cell levels are restored after 24 weeks of cART. Further, our model predicted that the timing of cART determined the viral clearance as opposed to the timing of co-infection or CD4 T cell level. Overall the study highlighted that immunosuppression played a major role in HIV/HPV coinfection and the results from the model were shown to be independent of the induced molecular effects by tat. However, there are few important limitations of the study: the effect of latent HIV reservoirs after cART are disregarded and the spatial structure of epithelial tissue layer is not taken into account. The latent HIV reservoirs can potentially contribute towards the emergence of HPV infections in such co-infected individuals.

In Chapter 5, we highlighted the need for integration of factors including health, physical activity, diet, nutrition, -omics and metabolism measures. It suggests the need for advanced computational modeling techniques to make accurate personalized health recommendations. The chapter allows us to consider the challenges that we encounter ranging from a need of computational infrastructure, need for data standardization and supports the need of adaptation to data-driven machine learning methods, some of which are used in Chapter 3.

Taken together, the research presented in this dissertation highlights the use of computational and mathematical modeling to study the host immune responses in the course of viral and bacterial infectious diseases. We firstly review -i) the need for computational modeling, discuss two case studies where we use, ii) a hybrid (ODE/PDE/ABM) model to study H. pylori infection and iii) an ODE model to study HIV/HPV coinfection model. Lastly, we discuss the challenges faced in the process of adopting the computational modeling technologies and how advanced data analytics and machine learning based approaches can help in the process. The findings from the two case studies presented here contribute towards understanding the host immune system dynamics underlying infectious diseases.