

<b>Manuscript Number:</b>	GIGA-D-18-00435R1	
<b>Full Title:</b>	High-Resolution Computational Modeling of Immune Responses in the Gut	
<b>Article Type:</b>	Research	
<b>Funding Information:</b>	Defense Threat Reduction Agency (HDTRA1-18-1-0008)	Dr. Josep Bassaganya-Riera Dr. Raquel Hontecillas
<b>Abstract:</b>	<p>Background: Helicobacter pylori causes gastric cancer in 1-2% of cases, but it exerts beneficial health effects including protection against allergies and gastroesophageal diseases. The majority of H. pylori-colonized individuals, an estimated 85%, do not present any detrimental effects. The mechanisms that promote host tolerance to the bacterium in the gastrointestinal mucosa and systemic regulatory effects requires further study hence we investigated the dynamics of immunoregulatory mechanisms triggered by H. pylori infection using a high-performance computing driven ENteric Immunity Simulator (ENISI) multiscale model. The immune responses were simulated in a high-resolution model integrating agent-based model, ordinary and partial differential equations.</p> <p>Results: The outputs were analyzed using two sequential stages wherein the first stage used a partial rank correlation coefficient regression-based and the second employed 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 the time-varying impact of input parameters during the initiation, peak and chronic phases of infection. The data analytics methods 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.</p> <p>Conclusion: 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 both modulating the host responses and shaping infection outcomes.</p>	
<b>Corresponding Author:</b>	Raquel Hontecillas, Ph.D. Virginia Polytechnic Institute and State University Blacksburg, Virginia UNITED STATES	
<b>Corresponding Author Secondary Information:</b>		
<b>Corresponding Author's Institution:</b>	Virginia Polytechnic Institute and State University	
<b>Corresponding Author's Secondary Institution:</b>		
<b>First Author:</b>	Meghna Verma, M.S.	
<b>First Author Secondary Information:</b>		
<b>Order of Authors:</b>	Meghna Verma, M.S.	
	Josep Bassaganya-Riera, Ph.D.	
	Andrew Leber, Ph.D.	
	Nuria Tubau-Juni, M.S.	
	Stefan Hoops, Ph.D.	
	Vida Abedi, Ph.D.	
	Xi Chen, Ph.D.	

	Raquel Hontecillas, Ph.D.
<b>Order of Authors Secondary Information:</b>	
<b>Response to Reviewers:</b>	<p>Point by point response to the Reviewer reports</p> <p>We would like to thank the reviewers and editors involved for dedicating valuable time to review our manuscript entitled “High-Resolution Computational Modeling of Immune Responses in the Gut”. We are extremely grateful to both reviewers and the editors for their time and attention to our submission. The review was very helpful to us, and the recommended changes have allowed us to provide better documentation and clarity to the present work. The review process has been helpful in the improvement of our submission. We have considered the comments that were made and have prepared the following point-by-point response.</p> <p>Reviewer #1: In this manuscript, the authors constructed a multi-scale systems biology model of <i>Helicobacter pylori</i> infection to study the interaction between bacterial infection and the immune system. Some modifications could be considered to improve the quality of this manuscript:</p> <p>1. The model needs to be more clearly described in the text. Some details might be available from the code; nevertheless, it would be helpful for readers to understand if the authors can include more information regarding the model. For example:</p> <p>We thank the reviewer for their valuable suggestion. We agree with the reviewer’s comment and accordingly we updated the manuscript with the response described below in detail.</p> <p>a. Agent-based model:</p> <p>i. What is the spatial discretization? The authors mentioned it’s a 30*10 2D grid cell, but resident macrophages are in thousands. So multiple cells are allowed in the same grid location? How many?</p> <p>Response: i) The model has a spatial discretization such that the dimension of the entire (two-dimensional) grid is 30nm x 10 nm). An individual grid cell is 1nm x 1nm, however, this is a configurable run parameter and can be changed without modifying the model. An individual grid cell 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 resulting time series of cytokine concentrations will be used to update the cytokine value in the ABM/PDE system and COPASI would simulate a different model for each of the relevant cell type within that individual grid cell. Below is a figure describing the grid, also added in the Additional file Fig S2.</p> <p>The entire grid is divided within into 4 functionally and anatomically distinct sized compartments such that the dimensions of the 4 compartments are lumen (2nm), epithelium (1nm), lamina propria (5nm) and gastric lymph node (2nm). The following compartments are adjacent to each other:</p> <ul style="list-style-type: none"> <li>•Lumen - epithelium</li> <li>•Epithelium - lamina propria</li> <li>•Lamina propria – gastric lymph node</li> </ul> <p>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 agents were randomly placed separated by the four compartments 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 that there is no limit of agents within each individual spatial grid cell. 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. We demonstrate below how we obtain a count of thousands of 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 (1) described below -</p>

$n(\text{resident macrophages}) = \text{size}(\text{compartment}(\text{lamina propria})) \times \text{concentration}(\text{resident macrophages})(1)$ .

$n(\text{resident macrophages}) = (30 \times 5) \times 30 = 4500$ .

The manuscript has been updated with the above addressed points, please refer to Line129 - Line147, and L217-245.

ii. What is the time step size?

Response: The time step size is 1 tick ~ 1 day which was obtained during the process of fitting the output to the results from the mouse model of H. pylori infection. For e.g. the peak of resident macrophages in lamina propria (refer Fig 2b,d) is observed at ~21 days which is similar to the results obtained in Fig 2A described in ((Viladomiu, Bassaganya-Riera et al. 2017) (also described in detail in point by point response 2.b). The manuscript has been updated with the above addressed points, please refer to Line247 - Line253.

iii. How is migration implemented for cells and bacteria agents?

Response: 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 (<https://github.com/NIMML/ENISI-MSM/src/compartment/Compartment.cpp>) file.

The manuscript has been updated with the above addressed points, please refer to Line294 - Line307.

b. ODE: What's the COPASI setup for the solver? How is the solver in sync with the ABM?

Response: The COPASI setup for the solver uses 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 manuscript has been updated with the above addressed points, please refer to Line266 - Line274.

c. PDE: What package and numerical scheme is used to solve the PDEs? What's the setting?

Response: ENISI MSM is a multiscale agent-based modeling platform for computational immunology which was building on our previous works, ENISI MSM that integrated COPASI, the ODE solver, ENISI, an agent based simulator (Mei, Abedi et al. 2015).

The ENISI MSM PDE solver uses a simple numerical scheme to solve the PDEs (<https://github.com/NIMML/ENISI-MSM/tree/master/src/diffuser>) and process distributed value layer (<https://github.com/NIMML/ENISI-MSM/blob/master/src/grid/ValueLayer.h>). The ValueLayer stores the value for a grid space and provides methods to change the values of individual grid cells. The Diffuser is used to diffuse the values of the ValueLayer using diffusion (d) and degradation (delta) constants as described in (Mei, Abedi et al. 2015). The diffusion constant

determines the migration of values of a grid cell to its neighboring grid cells. As implemented in ValueLayer library, the diffusion of cytokines follows the equation shown below also described in Mei et al, 2015. Here,  $v_n$  is the value of the grid cell itself at step  $n$ . The values of  $c_{\Delta}$  and  $c_d$  are degradation and diffusion constant respectively.

$$v_n = v_{n-1} + c_{\Delta} * [\sum (c_d \text{ neighbor} * v_{n-1 \text{ neighbor}}) - 6.0 * v_{n-1}]$$

0.3  
1.2  
0.3  
1.2  
-6.0  
1.2  
0.3  
1.2  
0.3

The PDE solver uses the above number scheme  $c_d \text{ neighbor}$  for the diffusion process. The step size  $c_{\Delta}$  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 the identical with the spatial discretization for the agents.

We updated the manuscript details to solve PDEs and the setting, please refer to L275-L293.

2. The authors listed the values of parameters in Table S1.

a. However, it's not clear what their units are (the baseline column seems to include characters such as "l<sup>2</sup>", "#", or "d". are these units? Please clarify).

Response: We thank the reviewer for pointing this out. Those were the units for the parameters in the COPASI ODE models. The parameters described in Table S1 are probability values (ranging from 0 to 1) and hence do not have any units. The characters have been removed and the column 3 of Table\_S1 has been updated.

b. Also, the sources of the parameter values are not very clear, except for the vague statement "expert judgement" (Saltelli, Tarantola et al. 2000 is cited, but this is an article on SA and does not contain parameters).

Response: The values of the parameters for the model presented here are obtained via 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 (Viladomiu, Bassaganya-Riera et al. 2017) published by NIMML (described here below in the last paragraph detail).

We want to clarify the practice of using expert opinion is known in the SA field and hence we cited Saltelli, Tarantola et al. 2000 as it supports the statement. As discussed in (Thorne, Bailey et al. 2007), 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. 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 experimental 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 Fig 1A of Viladomiu, Bassaganya-Riera et al. 2017. 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 (Fig 2a,d) 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 Fig. 2A, 2E of Viladomiu, Bassaganya-Riera et al. 2017. 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 (Fig 2b,e and Fig 2c,f) 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 updated the manuscript accordingly, please refer to L340-L350 and L654-671.

c. Please in the table explain what mechanism each parameter corresponds to. Some can be inferred from the name, but it's not very clear.

Response: We thank the reviewer for this valuable suggestion. We added column 2 in Table S1 that describes the detailed mechanism that each parameter corresponds to. We updated the manuscript accordingly, please refer to L236-L238. For e.g. 3 rows of the Table S1 are shown below –

Table S1

Name of parameters

Description

p\_epiinfbactdamage

Epithelial cell damage due to infectious bacteria

p\_epith1damage

Epithelial cell damage due to Th1 cells

p\_epith17damage

Epithelial cell damage due to Th17 cells

d. Some parameters are not included in the table. For example, the diffusivity of the cytokines are not listed.

We thank the reviewer for this valuable observation. We listed the diffusivity of the cytokines and updated the Table\_S1.

3. In Table 1 and Table 2, there is a T cell class named "Tr", which is not explained in the text. Please clarify.

Response: We thank the reviewer for pointing this out. The Tr cells are the type 1 regulatory (Tr1) T cells that are regulatory subset of T cells, whose expansion is dependent on environmental IL-10 (produced by Mreg). These are different than iTreg which are T cells differentiated from naïve T cell in presence of tolerogenic dendritic cells and TGF- $\beta$  cytokine. We clarified this point and updated the manuscript, please refer L208-211.

4. The authors used a Gaussian emulator as surrogate model for the hybrid model. In line 582, the authors mentioned that performance is evaluated using diagnostic plots in Figure S4. Please clarify what the "Observed" data refers to. Are these the same simulations from the training set which the emulator fitted to, or are these new simulations done? If these are the training set results, the authors need to run simulations and emulation on a new testing set and evaluate the performance; if it's already done, please clarify how its done (range of parameters, number of simulations, etc.)

Response: We thank the reviewer for your careful reading and bringing up the issues in the description of the original plot. Below please find our response to your comments.

First, the "observed" data, i.e. the 'x' axis in the first half of lower panel in Figure S4 (shown here below as Fig 1a) (please note in the revised manuscript the Figure S4 is now updated and referred to as by Fig S5.), refers to the observed output values of the simulations obtained after running the hybrid computer model, whereas the 'y' axis refers to the predicted values obtained from the cross validated model. These 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 second half of lower panel refers to the standard residual plot wherein the 'x' axis represents the observed values obtained from the simulation and the 'y' axis refers to the residual error ( $\{\text{error (predicted values - observed values) / standard deviation (error)}\}$ ) obtained.

In fact, the models used for plotting are the cross-validated ones and are not fit using

the entire dataset. Cross-Validation (section 7.10 of *The Elements of Statistical Learning* (Trevor, Robert et al. 2009), is a legitimate approach for model assessment and it is especially suitable in our case because the simulation data is expensive to obtain (each simulation takes ~9-10 minutes to run, thus 267 parameter sets with 20 replicates = 5,340 simulations. The entire simulation dataset took us about 2 months to obtain.

Nevertheless, we would like to show that using separate testing and training dataset for model assessment we obtained similar conclusions as those using the cross validated model.

We randomly split the observed output simulation dataset for one of the datasets (Fig S4 "a"), *Helicobacter pylori* in Lamina propria into training (80%) and testing (20%) sets and built the Gaussian emulator using the *mlegp* package. As observed in the Fig 1b and Fig 1c below, we plotted the predicted (values predicted using *mlegp*) vs. the observed simulation data values for both the training set (top panel of Fig 1b) and testing sets (top panel of Fig 1c). In the top panels of Fig 1b, the black circles denote the cross validated prediction points for the training dataset. Similarly, the top panel of Fig 1c, the black open circles are obtained after plotting the predictions for testing dataset, made using the model trained on the 80% of the randomly split dataset, vs. the observed values (known) for the 20% of the randomly split used as testing dataset here. Additionally, we calculated the standardized residuals for each of the 80% and 20% randomly split datasets and plotted the standardized residual plots in the lower panels of Fig. 1b and Fig 1c respectively.

As observed in the bottom panels of the Fig 1a, 1b and 1c the amount of standard residuals obtained for the cross validated model (Fig S4 a) from the paper and also the one mentioned in previous paragraph), the training dataset (80% randomly split dataset), and testing dataset (20% randomly split dataset) respectively, were similar. Thus, here we demonstrated that the results obtained from the cross-validated model built using *mlegp* (from Figure S4 a) and as shown here in Fig 1a) were similar to the results obtained using the cross-validation technique by randomly splitting the data into 80% and 20% (shown here in Fig 1b and Fig 1c).

Fig 1a. Original plot from Fig S4 a). The plot shows the predicted vs. observed simulation values for the Cross Validated (CV) model (top panel) and residual error plot for the CV model (bottom panel).

Fig 1b. The plot shows the predicted vs. observed simulation values for the randomly split 80% of the dataset (top panel) and residual error plot for the randomly split 80% of the dataset(bottom panel).

Fig 1c. The plot shows the predicted vs. observed simulation values for the randomly split 20% of the dataset (top panel) and residual error plot for the randomly split 20% of the dataset(bottom panel).

We clarified that the observed data refers to observed simulation values and recreated the Figure S4 (now updated to Fig S5 with updated legends. Please refer to L1045-L1058 in the manuscript.

Reviewer #2: The authors present results from a multi-scale hybrid model of host immune responses to H pylori exposure in the gut. The paper addresses outstanding questions in this complex system and overall the results are interesting. Some comments/questions to be addressed are outlined below.

1. A key component of the introduction ("double edge sword, p 1 line 4") as well as in the discussion (p28 line 672 "dual role as pathogen and beneficial organism") mentions the conflicting roles of H pylori infection - however the results do not clearly connect to help answer this dichotomy. More detailed analysis/discussion of the results should be provided to clarify the conclusion or the focus of the intro/discussion should be adjusted to relate more closely to the results currently presented.

Response: We thank the reviewer for this valuable suggestion. This study addresses the dichotomy in the introduction but mostly focuses on investigating the dynamics that promote the tolerance to the bacterium in the gastrointestinal mucosa and its systemic immunoregulatory effects. We view the dichotomy represented by the beneficial effects of regulatory responses (immune tolerance to the bacterium) in lesion development versus the detrimental actions of effector responses. Since, the majority of H. pylori-colonized individuals, approximately 85%, do not present any detrimental effect, we wanted to contribute towards the further investigation of the dynamics of immunoregulatory mechanisms underlying H. pylori infection using computational modeling. We emphasized the need for investigation of the immunoregulatory role and the adjusted the focus of the introduction and discussion to relate more closely to the results highlighting regulatory immune cells here. We updated the manuscript accordingly, please refer to L4-L7, L52-L54, L62-L65 in introduction and L850-L854 in discussion.

2. Section 3.4 and p 29 line 694 discuss the involvement of regulatory macrophages and tolerogenic DCs on the colonization of H pylori. These conclusions appear to be drawn based on correlation between responses in H pylori and macrophage/DC populations upon epithelial cell proliferation adjustment (Fig 5). A causal connection between the macrophages/DCs and H pylori is not made (or is not clear to me from the text). If such a connection is embedded in the mechanisms included in Table 1 it should be outlined in the results section where the conclusion is made otherwise simulations targeting the macrophage/DC populations would be needed to confirm this hypothesis.

Response: We thank the reviewer for pointing this out. We want to clarify that computational modeling based studies are capable of providing predictive modeling derived insights, however, any definitive causal connection should be validated in an experimental or clinical setting. In this study, based on the results obtained from the metamodel based global SA, the epithelial cell proliferation parameter was shown to an impact on the H. pylori population.

Following these findings which highlighted the importance of epithelial cell proliferation, the biological hypothesis derived from this prediction is that the epithelial cell proliferation parameter 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 a varying range of values (0.9-0.1) 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 computational model, as we varied the epithelial cell proliferation parameter. We analyzed the outputs from the hybrid computer model and observed 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. Overall, these cell populations varied due to the variation in the epithelial cell proliferation parameter.

We want to clarify that such connection was not embedded in the mechanisms included in Table 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 caused a change in regulatory and tolerogenic dendritic cell population. 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 a mechanism that involves the regulatory macrophages and tolerogenic cells. This is in line with our own conclusions drawn from a previous paper (Viladomiu, Bassaganya-Riera et al. 2017) 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.

We updated the manuscript with the detailed clarification, please refer to L788-826 and L908-L915.

3. Clarity is needed on some parts of the methods description:

3.1 P6, line 131: what are the units of the grid dimensions given. Are these the dimensions of a single grid cell or the entire grid? How are the 4 compartments separated on the grid?

Response : We thank the reviewer for pointing this out. These are the dimensions of the entire grid. An individual grid cell is 1nm x 1nm. The 4 compartments are separated by border implementation such that the dimensions of the 4 compartments are lumen (2nm), epithelium (1nm), lamina propria (5nm) and gastric lymph node (2nm). The following compartments are adjacent to each other:

- Lumen - epithelium
- Epithelium - lamina propria
- Lamina propria – gastric lymph node

We updated the manuscript with detailed model description, please refer to L222-L232. We also added a figure describing the grid in the Additional file Fig S2.

3.2 P6 line 149: what data were the ODEs calibrated to? Is there a reference?

Response: The CD4+ ODE model was calibrated using the experimental data provided in the Table S1 of the reference - Carbo, Hontecillas et al. 2013. The Particle Swarm algorithm implemented in COPASI was used to determine unknown model parameter values and fully calibrate the model. The intracellular macrophage ODE model was calibrated using a combination of sourced and new data generated from in vitro macrophage differentiation studies, compiled into a dataset provided within S2 file of Leber, Bassaganya-Riera et al. 2016.

We accordingly updated the manuscript, please refer to L155-L166.

3.3 P6 line 150, and p22 line 524: ABM parameters were calibrated to "qualitatively resemble" the patterns observed in in vivo model. What patterns? What is considered to be qualitatively similar enough? Do the simulations reproduce the dynamics as well and the endpoint experimental observations? Inclusion of experimental data alongside the simulations in figure 2 or a description of the key dynamics (e.g. fold-changes, peak values etc.) would go a long way in communicating confidence in the model parameters.

Response: We thank the reviewer for their valuable suggestion. The values of the parameters are obtained based on the qualitative comparison of the model outputs with the experimental results obtained from the mouse model of *H. pylori* infection. The simulations reproduced similar dynamics as described below - 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 (lacking PPARg in myeloid cells) was observed as shown in Fig 1A of (Viladomiu, Bassaganya-Riera et al. 2017). The decrease in bacterial burden led to a significant and sustained lower



colonization levels when compared to WT and CD4Cre (lacking PPAR $\gamma$  in T cells). Similar to the results observed in the mouse model, we observed a significant decrease (Fig 2a,d) 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 Fig. 2A, 2E of (Viladomiu, Bassaganya-Riera et al. 2017). 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 (Fig 2b,e and Fig 2c,f) in the number of resident macrophages as well as monocyte derived macrophages in the simulated WT groups in comparison to the simulated LysMcre group. As shown below, the peak of resident macrophages in lamina propria (refer Fig 2b of this paper) was observed at ~16-21 days which was similar to the peak observed in the CD64+F480hi macrophages at day 21, in Fig 2a described in (Viladomiu, Bassaganya-Riera et al. 2017). We included the experimental data alongside the simulation and revised the Fig 2 as shown below.

We accordingly updated the manuscript, please refer to L344-L350, L654-L671 and updated legend for Fig 2, L647-L649.

3.4 P11 line 246: the authors state that they perform global SA of the hybrid computer model. I believe they mean the metamodel here?

Response: We thank the reviewer for pointing this out. Although, a metamodel was built using the hybrid computer model, overall the global SA that included two stages –i) screening the influential inputs using PRCC (which was performed on the outputs from hybrid computer model simulations) and building a metamodel (using the outputs from the hybrid computer model) followed by calculating the Sobol' indices. Hence, we stated that we performed the global SA of the hybrid computer model.

3.5 P 21 line 480 and 484: parameter values were 'reduced' to emulate biological KOs. By how much were the parameters reduced?

Response: We thank the reviewer for pointing this out. We added new columns in Table S1 with the values of the parameters used to emulate the biological KOs. A complete set of parameter for each of the biological KOs are included as separate columns in Table S1.

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 baseline (WT) value respectively (refer Table S1). As described in (Carbo, Hontecillas et al. 2013), to simulate the LysMCre experimental conditions, the probabilities of i) a monocyte transitioning to a regulatory macrophage ( $p_{MonotoMreg}$ ) and ii) immature dendritic cells switching to tolerogenic dendritic cells ( $p_{iDCtotDC}$ ) were reduced approximately to 60% and 30% of the baseline (WT) value, respectively (refer Table S1).

We updated the manuscript with the above listed values, please refer to L602-615.

3.6 The in vivo model is mentioned several times before it is clarified to be a mouse model.

Response: We thank the reviewer for their valuable suggestion. We updated the manuscript and clarified that the in vivo model is a mouse model.

Comment from the Editor: Further, our series Guest Editor, Paul Macklin has had a quick look at the manuscript from a reproducibility point-of-view and suggests that you include somewhere (e.g., in supporting info) the specific examples for this paper, including detailed instructions on how to create the specific examples presented. Note that our curators also asked for detailed instructions on how to require detailed

instructions for usability - not just code.

In addition, please register any new software application in the SciCrunch.org database to receive a RRID (Research Resource Identification Initiative ID) number, and include this in your manuscript. This will facilitate tracking, reproducibility and re-use of your tool.

Response: The RRID (Research Resource Identification Initiative ID) number as assigned by the SciCrunch.org database is SCR\_016918. We included this in the manuscript, please refer to L180-L182.

Detailed instructions for the usability are described below and also included in Additional file S1. Detailed instructions on how to create the specific examples presented here are also included. We accordingly updated the manuscript, please refer to L178-180.

#### Additional file S1

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) 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 -  
`i.git clone --recursive https://github.com/NIMML/ENISI-Dependencies`
4. Change the path to the ENISI-Dependencies folder: `cd ENISI-Dependencies`
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 –  
`i.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_MSM_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 - PRCC\_activeinactiveinputs-added.xlsx.

Stage 2 Initialization

•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 file.

•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.

References

Mei, Y., V. Abedi, A. Carbo, X. Zhang, P. Lu, C. Philipson, R. Hontecillas, S. Hoops, N. Liles and J. Bassaganya-Riera (2015). "Multiscale modeling of mucosal immune responses." BMC Bioinformatics 16 Suppl 12: S2.

Trevor, H., T. Robert and F. JH (2009). The elements of statistical learning: data

	<p>mining, inference, and prediction, New York, NY: Springer.</p> <p>Viladomiu, M., J. Bassaganya-Riera, N. Tubau-Juni, B. Kronsteiner, A. Leber, C. W. Philipson, V. Zoccoli-Rodriguez and R. Hontecillas (2017). "Cooperation of Gastric Mononuclear Phagocytes with Helicobacter pylori during Colonization." J Immunol 198(8): 3195-3204.</p>
<b>Additional Information:</b>	
<b>Question</b>	<b>Response</b>
Are you submitting this manuscript to a special series or article collection?	No
<p><b>Experimental design and statistics</b></p> <p>Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>. Information essential to interpreting the data presented should be made available in the figure legends.</p> <p>Have you included all the information requested in your manuscript?</p>	Yes
<p><b>Resources</b></p> <p>A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite <a href="#">Research Resource Identifiers</a> (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>?</p>	Yes
<p><b>Availability of data and materials</b></p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in <a href="#">publicly available repositories</a> (where available and ethically appropriate), referencing such data using</p>	Yes

a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript.

Have you have met the above requirement as detailed in our [Minimum Standards Reporting Checklist](#)?

[Click here to view linked References](#)

## High-Resolution Computational Modeling of Immune Responses in the Gut

Meghna Verma<sup>1,2</sup>, Josep Bassaganya-Riera<sup>1</sup>, Andrew Leber<sup>1</sup>, Nuria Tubau-Juni<sup>1</sup>,  
Stefan Hoops<sup>1</sup>, Vida Abedi<sup>1</sup>, Xi Chen<sup>3</sup>, Raquel Hontecillas<sup>1,\*</sup>

<sup>1</sup>Nutritional Immunology and Molecular Medicine Laboratory, Biocomplexity Institute of Virginia Tech, Blacksburg, VA 24060, USA.

<sup>2</sup>Graduate Program in Translational Biology, Medicine and Health, Virginia Tech, Blacksburg, VA, 24061, USA.

<sup>3</sup>Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA.

### \* Correspondence:

Dr. Raquel Hontecillas

Email: [rmagarzo@vt.edu](mailto:rmagarzo@vt.edu)

**Keywords:** agent-based model, dendritic cells, ordinary differential equation, Gaussian process, *Helicobacter pylori*, high-performance computing, immune system, macrophages, metamodel, sensitivity analysis.

Email address for all authors:

Meghna Verma: [meghna89@vt.edu](mailto:meghna89@vt.edu)

Josep Bassaganya-Riera: [jbassaga@vt.edu](mailto:jbassaga@vt.edu)

Andrew Leber: [ajleber@vt.edu](mailto:ajleber@vt.edu)

Nuria Tubau-Juni: [nuriaj@vt.edu](mailto:nuriaj@vt.edu)

Stefan Hoops: [shoops@vt.edu](mailto:shoops@vt.edu)

Vida Abedi: [vidaabedi@gmail.com](mailto:vidaabedi@gmail.com)

Xi Chen: [xchen6@vt.edu](mailto:xchen6@vt.edu)

Raquel Hontecillas: [rmagarzo@vt.edu](mailto:rmagarzo@vt.edu)

1  
2  
3  
4  
5  
6 **1 Abstract**

7  
8 2 Background: *Helicobacter pylori* causes gastric cancer in 1-2% of cases, but it  
9  
10 3 exerts beneficial health effects including protection against allergies and  
11  
12 4 gastroesophageal diseases. [The majority of \*H. pylori\*-colonized individuals, an](#)  
13  
14 5 [estimated 85%, do not present any detrimental effects. The mechanisms that](#)  
15  
16 6 [promote host tolerance to the bacterium in the gastrointestinal mucosa and](#)  
17  
18 7 [systemic regulatory effects requires further study](#) hence we investigated the  
19  
20 8 dynamics of immunoregulatory mechanisms triggered by *H. pylori* infection using  
21  
22 9 a high-performance computing driven **EN**teric **I**mmunity **S**imulator (ENISI)  
23  
24 10 multiscale model. The immune responses were simulated in a high-resolution  
25  
26 11 model integrating agent-based model, ordinary and partial differential equations.  
27  
28 12 Results: The outputs were analyzed using two sequential stages wherein the first  
29  
30 13 stage used a partial rank correlation coefficient regression-based and the second  
31  
32 14 employed a metamodel-based global sensitivity analysis. The influential  
33  
34 15 parameters screened from the first stage were selected to be varied for the  
35  
36 16 second stage. The outputs from both stages were combined as a 'training  
37  
38 17 dataset' to build a spatiotemporal metamodel. The Sobol' indices measured the  
39  
40 18 time-varying impact of input parameters during the initiation, peak and chronic  
41  
42 19 phases of infection. The data analytics methods identified epithelial cell  
43  
44 20 proliferation and epithelial cell death as key parameters that control infection  
45  
46 21 outcomes. *In-silico* validation showed [that](#) colonization [with](#) *H. pylori* decreased  
47  
48 22 with a decrease in epithelial cell proliferation [which was linked to](#) regulatory  
49  
50 23 macrophages and tolerogenic dendritic cells.

51  
52  
53 24 Conclusion: The hybrid model of *H. pylori* infection identified epithelial cell  
54  
55  
56  
57  
58  
59 25 proliferation as a key factor for successful colonization of the gastric niche and  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

26 highlighted the role of tolerogenic dendritic cells and regulatory macrophages in  
27 both modulating the host responses and shaping infection outcomes.

28 **1. Background**

29 Computational modeling of the immune response dynamics can provide  
30 novel insights and facilitate the systems level understanding of the interactions  
31 at the gastric mucosa during infection. Ordinary differential equation (ODE-  
32 based methods are deterministic and based on the average response of cells  
33 over time. Dynamical models are used in immunology for system-level analyses  
34 of CD4+ T cell differentiation (Carbo, Bassaganya-Riera et al. 2013),  
35 macrophage differentiation (Leber, Bassaganya-Riera et al. 2016), immune  
36 responses elicited by *Clostridium difficile* infection (Leber, Viladomiu et al. 2015),  
37 co-infections (Verma, Erwin et al. 2017), and in cancer and immunotherapy  
38 (Qomlaqi, Bahrami et al. 2017). However, ODE-based models lack the spatial  
39 aspects and the features to study the [organ and](#) immune cell [topology](#) over time.  
40 Agent-based models (ABM) employ a bottom-up approach that focuses on the  
41 spatial and temporal aspects of individual immune cells, unlike the ODE-based  
42 methods. This rule-based method includes agents that act as local entities which  
43 interact locally with other agents, move in space, and follow set of rules  
44 representing their role in a given system and contribute towards generating an  
45 emergent behavior. Since, the immune system is a complex dynamical system  
46 (Vodovotz, Xia et al. 2017) wherein the components *i.e.*, the immune cells move  
47 in space and time [changing their location](#), ABMs are useful tools that can be  
48 employed to understand biological mechanisms and the hidden insights.

49 *Helicobacter pylori* is a gram-negative bacterium that has persistently  
50 colonized the human stomach since early evolution (Kusters, van Vliet et al.



1  
2  
3  
4 51 2006) (Mane, Dominguez-Bello et al. 2010) and is currently found in over 50%  
5  
6 52 (Cover and Blaser 2009) of the global population. H. pylori has co-evolved with  
7  
8  
9 53 humans for thousands of years, such that an estimated 85% of the H. pylori-  
10  
11 54 colonized individuals, do not present any detrimental effects. Thus, the vast  
12  
13 55 majority of carriers (*i.e.* up to 75%) remain asymptomatic, while only 15%  
14  
15 56 develop ulcers, and less than 3% develop cancer. Further, growing and  
16  
17 57 sometimes contradictory evidence from recent experimental, clinical studies and  
18  
19 58 epidemiological studies suggest that *H. pylori* might provide protection against  
20  
21 59 obesity-related inflammation and type 2 diabetes (Bassaganya-Riera,  
22  
23 60 Dominguez-Bello et al. 2012), esophageal, cardiac pathologies, childhood  
24  
25 61 asthma and allergies (Oertli, Sundquist et al. 2012) and autoimmune diseases.  
26  
27 62 In this context, it is crucial to understand the mechanisms that promote host  
28  
29 63 tolerance to the bacterium in the gastrointestinal mucosa and its systemic  
30  
31 64 regulatory effects since these have been linked to the beneficial commensal  
32  
33 65 aspects of H. pylori-human host interaction. Computational models provide a  
34  
35 66 cost-effective and predictive way to study the complex and dynamic immune  
36  
37 67 system interactions and form a non-intuitive novel hypothesis. Solving the  
38  
39 68 complex puzzle of immunoregulatory mechanisms that include large  
40  
41 69 spatiotemporal scales ranging from cellular, intracellular, tissue and organ level  
42  
43 70 scales is a major unsolved challenge that requires applying computational  
44  
45 71 modeling and data analytics.

50  
51 72 An advanced hybrid model used to study the mucosal immune response  
52  
53 73 during gut inflammation highlighted the mechanisms by which effector CD4+ T  
54  
55 74 cell responses, contributed to tissue damage in the gut mucosa following  
56  
57 75 immune dysregulation (Mei, Abedi et al. 2015). Other hybrid models with the  
58  
59 76 integration of ABM, ODE, and PDE technologies, were developed to understand

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

77 the dynamics of tumor development (Gong, Milberg et al. 2017) and tumor  
78 growth models (Wang, Birch et al. 2009). These combined techniques have been  
79 used to develop multi-organ models in various situations, including the study of  
80 granuloma formation (Marino, El-Kebir et al. 2011) and pressure-driven ulcer  
81 formation in post spinal cord injury patients (Solovyev, Mi et al. 2013). The  
82 summary of different agent-based simulators with immunology related  
83 applications are discussed and summarized in (Bassaganya-Riera 2015,  
84 Cappuccio, Tieri et al. 2016). The comparison between different multiscale  
85 modeling tools and agent-based immune simulators, are discussed in (An, Mi et  
86 al. 2009, Mei, Abedi et al. 2015).

87 In this study, we utilize a high-resolution **EN**teric Immunity **SI**mulator (ENISI)-  
88 based model of the stomach for simulating the mucosal immune responses to *H*  
89 *pylori* infection. The advanced hybrid multiscale modeling platform ENISI  
90 multiscale model (MSM) is capable of scaling up to  $10^{12}$  agents (Abedi,  
91 Hontecillas et al. 2015). The host immune responses initiated during *H. pylori*  
92 infection and the underlying immunoregulatory mechanisms are captured using  
93 the ENISI multiscale hybrid model. The underlying intracellular mechanisms that  
94 control cytokine production, signaling and differentiation of macrophages and T  
95 cells are modeled by using ODEs, the diffusion of cytokine values is modeled  
96 using PDEs and the location and interactions among the immune cells, bacteria  
97 and epithelial cells are modeled by using ABMs. The hybrid model thereby  
98 represents a high-performance computing (HPC)-driven large-scale simulation  
99 of the massively interacting cells and molecules in the immune system,  
100 integrating the multiple modeling technologies from molecules to systems across  
101 multiple spatiotemporal scales.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

102 To understand the dynamics and emergent immunological patterns  
103 described by this hybrid model, we employed sensitivity analysis (SA), an  
104 important part of the model analysis used to explore the influence of varying  
105 model parameters on the simulation outputs. The influence of the effects of  
106 changes in parameter values on the model output explains the model dynamics  
107 that underlay the outputs (Ligmann-Zielinska, Kramer et al. 2014, Ten Broeke,  
108 Van Voorn et al. 2016). Furthermore, SA examines the robustness of the model  
109 output at a different range of parameter values that correspond to a range of  
110 different assumptions. We employed global SA and conducted a two-stage  
111 spatiotemporal global SA approach. First, we used a regression-based method  
112 such as the partial rank correlation coefficient (PRCC) and screened the  
113 important input parameters that were shown to have the most influence on the  
114 output cell populations obtained from the hybrid model. Second, the screened  
115 input parameters from the first stage were varied to build a second stage  
116 parameter design matrix, and the computer simulations were again run using the  
117 hybrid ENISI model. The outputs from both analytics stages were combined and  
118 used as a ‘training dataset’ to build a spatiotemporal Gaussian process based  
119 metamodel. Finally, variance-based decomposition global SA was used to  
120 compute the Sobol’ indices and the most influential parameters over the course  
121 of infection were identified. The data analytics methods conducted on the hybrid  
122 model identified the epithelial cell parameters such as epithelial cell proliferation  
123 as the most influential ones, required for the successful colonization of *H. pylori*  
124 in the gastric microenvironment.

## 2. Methods

### 2.1 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 30nm x 10 nm. An individual grid cell for our simulation is 1nm x 1nm, however, this is a configurable run parameter and can be changed without modifying the model. An individual grid cell 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 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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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 (Carbo, Hontecillas et al. 2013) and macrophages (Leber, Bassaganya-Riera et al. 2016) 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 \(Carbo, Hontecillas et al. 2013\). 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 \(Carbo, Hontecillas et al. 2013\). 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 \(Leber, Bassaganya-Riera et al. 2016\). The parameter values are specified within the previously published manuscripts - CD4+ T cell ODE model \(Carbo, Hontecillas et al. 2013\) and macrophages \(Leber, Bassaganya-Riera et al. 2016\).](#) The parameters of the calibrated ODEs were kept unchanged, and the 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\_ (Viladomiu, Bassaganya-Riera et al. 2017), [also described in detail in section 3.1.](#)

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 grid cell to its neighboring grid cell similar to as described](#) in (Mei, Abedi et al. 2015). The

1  
2  
3  
4 175 features of ABM, ODE, and PDE were combined to create a multiscale modeling  
5  
6 176 environment which spanned across different orders of spatiotemporal scales.

7  
8  
9 177 The code for the hybrid model is freely accessible and can be downloaded  
10  
11 178 at <https://github.com/NIMML/ENISI-MSM>. [The detailed instructions for the](#)  
12  
13 179 [usability, instructions on 'how to run a simulation' and codes for creating specific](#)  
14  
15 180 [examples presented here are presented in Additional file S1. The SciCrunch.org](#)  
16  
17 181 [database assigned research identification initiative ID \(RRID\) for ENISI-MSM is](#)  
18  
19 182 [RRID:SCR\\_016918](#). The design of the implementation of the code structure is  
20  
21 183 depicted in the Additional file **Fig S1**. The hybrid model is implemented in C++  
22  
23  
24 184 and utilized the Repast HPC library ([https://repast.github.io/repast\\_hpc.html](https://repast.github.io/repast_hpc.html))  
25  
26 185 (Collier and North 2011). For the ODEs, we utilized COPASI (Hoops, Sahle et  
27  
28 186 al. 2006), an ODE-based modeling tool used in computational biology. The rules  
29  
30 187 in the model that described the interaction of *H. pylori* with the gastric mucosa  
31  
32 188 and the immune responses resulting from the infection are derived from the  
33  
34 189 findings in our previously published studies (Carbo, Bassaganya-Riera et al.  
35  
36 190 2013, Leber, Bassaganya-Riera et al. 2016). Specifically, this hybrid model  
37  
38 191 reproduced the immune responses generated by the interaction *H. pylori* and the  
39  
40 192 resident macrophages as shown in [the mouse model of H. pylori infection](#)  
41  
42 193 (Viladomiu, Bassaganya-Riera et al. 2017). The rules for each cell type in the *H.*  
43  
44 194 *pylori* infection are summarized in **Table 1**. A pictorial representation of the rules  
45  
46 195 is depicted in *Fig 1*. These cell types represented as agents, act according to the  
47  
48 196 rules (as in **Table 1**) that are updated at discrete simulation cycle.  
49  
50  
51  
52  
53  
54  
55  
56  
57

58 199 **Fig 1. *Helicobacter pylori* infection schematic diagram of the hybrid ABM**  
59  
60 200 **ODE model**

1  
2  
3  
4 201 *The model comprises four compartments, i) the lumen that contains *H. pylori* and*  
5  
6 202 *bacteria, ii) epithelium that contains epithelial cells and dendritic cells, iii) lamina*  
7  
8 203 *propria that contains variety of immune cells including the infiltrating effector*  
9  
10 204 *(eDCs) and tolerogenic (tDCs) dendritic cells, monocytes, regulatory*  
11  
12 205 *macrophages (both resident and monocyte-derived macrophages), T helper*  
13  
14 206 *cells and naïve CD4+ T cells (nT), Th1, iTreg, Th17, [Tr cells](#). and iv) gastric*  
15  
16 207 *lymph node compartment that contains eDCs, tDCs, Th1, Th17, iTreg and nT.*  
17  
18 208 [The Tr cells in the lamina propria are the type 1 regulatory \(Tr1\) T cells with](#)  
19  
20 209 [regulatory function whose expansion is largely dependent on environmental IL-](#)  
21  
22 210 [10. These are different than iTreg which are T cells differentiated from naïve T](#)  
23  
24 211 [cell in presence of tolerogenic dendritic cells and TGF- \$\beta\$  cytokine](#) *The two*  
25  
26 212 *calibrated ODEs for T cells and regulatory macrophages are integrated as the*  
27  
28 213 *ODE components in the hybrid model. The cellular agents are simulated in a*  
29  
30 214 *two-dimensional grid space with their behavior defined by a set of rules during a*  
31  
32 215 *course of *H. pylori* infection.*

### 216 217 [\*\*Model description\*\*](#)

218 [ENISI MSM is a multiscale agent-based modeling platform for computational](#)  
219 [immunology which was built on our previous works, ENISI-MSM \(Mei, Abedi et](#)  
220 [al. 2015\) that integrated COPASI, the ODE solver, ENISI, an agent based](#)  
221 [simulator.](#)

### 222 [\*\*Spatial discretization\*\*](#)

223 [The model has a spatial discretization such that the dimension of the entire \(two](#)  
224 [dimensional\) grid is 30nm x 10 nm. An individual grid cell is 1nm x 1nm, however,](#)  
225 [this is a configurable run parameter and can be changed without modifying the](#)  
226 [model. The four functionally and anatomically distinct sized compartments are](#)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

227 separated by border implementation such that the dimensions of the four  
228 compartments are lumen (2nm), epithelium (1nm), lamina propria (5nm) and  
229 gastric lymph node (2nm). The following compartments are adjacent to each  
230 other: lumen – epithelium, epithelium - lamina propria and lamina propria –  
231 gastric lymph node. A figure describing the spatial discretization is shown in the  
232 Additional file **Fig S2**.

233 The parameters that define the initial concentration of the agents and the  
234 diffusivity of cytokines are obtained from a properties file (*model.props* in the  
235 Howtorunasimulation folder in the GitHub repository). All the value of the  
236 parameters as listed in *Table S1*. The detailed mechanism that each parameter  
237 corresponds to is described in the second column, *parameter description*, of  
238 *Table S1*. We demonstrate below how we obtain a count of thousands resident  
239 macrophages. For e.g., if the initial concentration of resident macrophages in the  
240 lamina propria is 30, the total number of these resident macrophages can be  
241 calculated by the equation described below -

$$\frac{n(\text{resident macrophages})}{(\text{resident macrophages})} = \frac{\text{size}_{\text{compartment}}(\text{lamina propria}) \times \text{concentration}_{\text{initial}}}{(\text{resident macrophages})}$$
$$n(\text{resident macrophages}) = (30 \times 5) \times 30 = 4500.$$

### **Time Step size**

248 The time step size is 1 tick ~ 1 day which was obtained during the process of  
249 qualitatively comparing the output to the results from the mouse model of *H.*  
250 *pylori* infection. For e.g., the peak of resident macrophages in lamina propria  
251 (refer *Fig 2b, d*) is observed at ~21 days which is similar to the results obtained



1  
2  
3  
4 252 [in Fig 2A described in \(Viladomiu, Bassaganya-Riera et al. 2017\) \(also described](#)  
5  
6 253 [in detail in section 3.1\).](#)

7  
8 254  
9  
10  
11 255 **Updating**

12  
13 256 Each agent has an 'act' function within the code\_ that describes the rules  
14  
15 257 implemented for each of the agent groups. At every simulation cycle, each agent  
16  
17 258 inspects its [location and](#) updates its state. [If the agents were T cells and](#)  
18  
19 259 [macrophages, they obtained the cytokine concentration from the ValueLayers,](#)  
20  
21 260 [sent that information to COPASI that calculated the differentiation subtype of the](#)  
22  
23 261 [agent and cytokines to be secreted that into the environment \(Mei, Abedi et al.](#)  
24  
25 262 [2015\).](#) The input to the ODEs were the cytokine values at the agent's location.  
26  
27 263 Thus, the intracellular ODE models were utilized to determine and update the  
28  
29 264 state. Each agent proliferated, died, changed its state and moved across the  
30  
31 265 compartment, following the set of rules defined for them.

32  
33 266 [The COPASI setup for the solver used the LSODA \(Livermore Solver for](#)  
34  
35 267 [Ordinary Differential Equations\) differential equation solver. The default values](#)  
36  
37 268 [for the setup such as the - relative tolerance \(1e-6\), absolute tolerance \(1e-12\)](#)  
38  
39 269 [and maximum internal steps of 10000 were maintained. The ENISI MSM sends](#)  
40  
41 270 [the current concentrations of the cytokines](#)  
42  
43 271 [to COPASI. COPASI uses those values to integrate the deterministic](#)  
44  
45 272 [model for one tick, i.e., 1 day. The resulting time series of cytokine](#)  
46  
47 273 [concentrations are used to update the cytokine value in the ABM/PDE](#)  
48  
49 274 [system. COPASI simulates different model for each relevant cell type.](#)

50  
51 275 [The ENISI MSM PDE solver uses a simple numerical scheme to solve the PDEs](#)  
52  
53 276 [\(https://github.com/NIMML/ENISI-MSM/tree/master/src/diffuser\) and process](#)  
54  
55 277 [distributed value layer \(https://github.com/NIMML/ENISI-](#)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

278 [MSM/blob/master/src/grid/ValueLayer.h](#)). The ValueLayer stores the value for a  
279 [grid space and provides methods to change the values of individual grid cells.](#)  
280 [The Diffuser is used to diffuse the values of the ValueLayer using diffusion \(d\)](#)  
281 [and degradation \(delta\) constants as described in \(Mei, Abedi et al. 2015\). The](#)  
282 [diffusion constant determines the migration of values of a grid cell to its](#)  
283 [neighboring grid cells. As implemented in ValueLayer library, the diffusion of](#)  
284 [cytokines follows the equation shown below also described in Mei et al, 2015.](#)  
285 [Here,  \$v\_n\$  is the value of the grid cell itself at step n. The values of  \$c\_{delta}\$  and  \$c\_d\$  are](#)  
286 [degradation and diffusion constant respectively.](#)

$$v_n = v_{n-1} + c_{delta} * [\sum (c_d^{neighbor} * v_{n-1}^{neighbor}) - 6.0 * v_{n-1}]$$

0.3	1.2	0.3
1.2	-6.0	1.2
0.3	1.2	0.3

288  
289 [The PDE solver uses the above number scheme  \$c\_d^{neighbor}\$  for the diffusion](#)  
290 [process. The step size  \$c\_{delta}\$  is automatically adjusted at the beginning of the](#)  
291 [simulation based on the degradation and diffusion constants to avoid underflow](#)  
292 [errors, i.e., multiple PDE steps are in general executed per tick. The grid size is](#)  
293 [the identical with the spatial discretization for the agents.](#)

294 **Movement**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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

308  
309 The hybrid model simulations were run on an Ivy Bridge-EX E7-4890 v2 2.80  
310 GHz (3.40 GHz Turbo) quad processor nodes. The code was parallelized such  
311 that the simulation time on a single node with four parallel tasks, varied between  
312 9-10 minutes. This runtime was based on the model parameters at the initiation  
313 stage, which included the number of immune cell, bacteria, epithelial cells,  
314 number of time steps, and size of the two-dimensional grid. To facilitate the  
315 investigation of the mechanisms underlying host responses during *H. pylori*  
316 infection, anatomical and functional compartments were spatially linked such that  
317 the agents had both unidirectional and bidirectional movement. All the agents  
318 worked in a synchronous format wherein the two agent populations  
319 (macrophages and T cells) made function calls to their respective ODE models  
320 (Leber, Bassaganya-Riera et al. 2016) (Carbo, Hontecillas et al. 2013). These

agents used the varying cytokine concentration (*i.e.* environment variable) in

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

their grid spaces as inputs to the ODE model, and these models were run using

COPASI (Hoops, Sahle et al. 2006). **Table 2** shows information on the agents

and the states that they can acquire.

**Table 2. List of all the agents and the states they can acquire.**

*All the agents can acquire at least 1 and at the most 5 states. The names chosen for the acquired states are closely related to their functional properties based on the underlying “rules”.*

## 2.2 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

1  
2  
3  
4 337 that did not show a lot of variation were not included). A range of values  
5  
6 338 (maximum and minimum) was specified for each of the parameters (refer to  
7  
8  
9 339 Additional File **Table S1**) by expert judgment, summarized by bounded intervals.  
10  
11 340 The practice of using expert judgment is known in the SA field as supported in  
12  
13 341 (Saltelli, Tarantola et al. 2000). As discussed in (Thorne, Bailey et al. 2007), one  
14  
15 342 of the challenges encountered using ABM is the process of determining the  
16  
17 343 parameter values, for e.g. this may include the lack of the availability of  
18  
19 344 experimental techniques to measure such parameters. The values of the  
20  
21 345 parameters for the model presented here are obtained via the best guess based  
22  
23 346 on the qualitative comparison of the computer model outputs with that of the  
24  
25 347 experimental results obtained from the mouse model of *H. pylori* infection  
26  
27 348 (Viladomiu, Bassaganya-Riera et al. 2017) (as described in detail in Section 3.1).  
28  
29 349 Since, the source of the parameters is not known we estimated the values to fit  
30  
31 350 the data obtained from the mouse model of infection.  
32  
33  
34

35  
36 351 The values of these parameters were normalized within the range of 0 and 1  
37  
38 352 for SA purposes. We employed a two-stage metamodeling methodology to  
39  
40 353 determine the influence of each input parameter to the model output, in a high  
41  
42 354 dimensional screening setting inspired by (Moon, Dean et al. 2012). The step-  
43  
44 355 wise procedure is described in the Additional file, *Fig S3*. All the files for global  
45  
46 356 SA are freely accessible and can be downloaded at  
47  
48  
49 357 <https://github.com/NIMML/Sensitivity-Analysis>.  
50

51 358 The two-stage global SA is described in detail in the below section. To  
52  
53 359 summarize, for the first stage the input parameter matrix was designed using the  
54  
55 360 method described in Moon, Dean et al. 2012 and simulations were run using the  
56  
57  
58 361 hybrid computer model. The simulation output from the first stage was analyzed  
59  
60 362 using PRCC as it was computationally efficient, and the active inputs (significant  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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, describe a detailed step by step explanation of the procedure.

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 (Ten Broeke, Van Voorn et al. 2016). 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  $n_1 >$

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

389  $5 \cdot d^{0.5} = 2.5d$  as in (Moon, Dean et al. 2012). To construct a  $n_1 \times (n_1 - 1)$   
390 preliminary input parameter design matrix,  $X^*$ , needed to be constructed ((Moon,  
391 Dean et al. 2012)). The input parameter design matrix for first stage sampling  
392 was drawn from  $X^*$ .

393 The algorithm for the first stage design generated a design matrix  $X^{(1)}$  that  
394 satisfied the below three listed properties as in (Moon, Dean et al. 2012)

- 395 i) The columns of  $X^*$  were uncorrelated thereby facilitating the independent  
396 assessments of the effects due to the input parameters.
- 397 ii) The maximum and minimum value in each input parameter column were  
398 ensured to be 0 and 1 respectively, thereby preventing any input values  
399 with larger values to have a larger influence on the response, induced by  
400 the design.
- 401 iii) The designs defined by  $X^*$  had “space-filling” properties such that all the  
402 regions of the input space were exhaustively explored.

404 First stage sampling plan:

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

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

1  
2  
3  
4 414  
5  
6 415 First stage analysis  
7  
8  
9

10 416 We analyzed the outputs from first stage analysis and screened the active  
11  
12 417 inputs from using PRCC. To measure the effect of input parameter on output, we  
13  
14 418 performed both PRCC and the spearman rank correlation coefficient (SRCC)  
15  
16 419 analysis. PRCC and SRCC were chosen because they were computationally  
17  
18  
19 420 efficient (accounting for the low computational budget). A correlation analysis  
20  
21 421 provides a measure of the strength of linear association between input and  
22  
23 422 output variable (Marino, Hogue et al. 2008). A correlation coefficient between  $x_j$   
24  
25 423 and  $y$  is calculated as follows:  
26  
27  
28

29  
30 424 
$$r_{x_j y} = \frac{Cov(x_j, y)}{\sqrt{Var(x_j)Var(y)}} = \frac{\sum_{i=1}^N (x_{ij} - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_{ij} - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$
  
31  
32  
33  
34

35 425 
$$j = 1, 2, \dots, k.$$
  
36  
37  
38

39 426 where  $Cov(x_j, y)$  stands for the covariance between  $x_j$  and  $y$ , and  $Var(x_j)$  and  
40  
41 427  $Var(y)$  are the variance of  $x_j$  and  $y$  respectively.  
42  
43  
44

45 428 PRCC is performed when i) a non-linear but monotonic relation exists  
46  
47 429 between the input and outputs, and ii) when little or no correlation exists between  
48  
49 430 the input variables (which is guaranteed by the property (i) of our input parameter  
50  
51 431 matrix,  $X^{(1)}$  described above). As described in Marino, Hogue et al. 2008, the  
52  
53 432 PRCC between rank transformed  $x_j$  and  $y$  is the CC between the two residuals  
54  
55  
56 433  $(x_j - \widehat{x}_j)$  and  $(y_j - \widehat{y}_j)$  where  $\widehat{x}_j$  and  $\widehat{y}_j$  are rank transformed and follow the linear  
57  
58 434 regression models as follows:  
59  
60  
61  
62  
63  
64  
65



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

$$\hat{x}_j = c_o + \sum_{\substack{p=j \\ p \neq j}}^k c_p x_p \text{ and } \hat{y}_j = c_o + \sum_{\substack{p=j \\ p \neq j}}^k c_p x_p .$$

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 parameters and assessed if the PRCC coefficients were significantly different than zero (Marino, Hogue et al. 2008). We run the PRCC analysis for 13 output cell populations (*Fig 3* 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 [Fig S4](#)). 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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

458 described in (Moon, Dean et al. 2012) for the second stage and the number of  
459 design points amounted to,  $n_2 = 100\% * 5 * a$  where 'a' stands for the number of  
460 active inputs from the first stage. This resulted into  $n_2 = 23 * 5 = 115$  parameters  
461 combinations for the second stage input parameter design matrix. Since outputs  
462 from both stages are to be combined for second stage analysis, per (Moon, Dean  
463 et al. 2012), the design for the second stage was chosen to build on top of  $X^{(1)}$ .  
464 The sampling phase design algorithm ensured that the columns satisfied the  
465 properties (i) (uncorrelated design points) and (ii) (between values 0 and 1) as  
466 listed in the previous section. We constructed the 115 x 38 (115 parameter  
467 setting and 38 parameters) design matrix for the second stage that incorporated  
468 the 23 active inputs obtained from the PRCC screening in the first stage output  
469 analysis. After combining the design points from both the stages, the parameter  
470 design matrix  $X$  with space filling properties contained 267 (152 from the first  
471 stage and 115 from the second stage) design points.

472 Second stage analysis

473 We run the computer code for the hybrid model with the second stage input  
474 parameter design matrix (with 115 ( $n_2$ ) design points), for 20 (r) replicates, which  
475 amounted to 115 x 20 (2300) runs. The outputs from the first stage (152 x 20  
476 runs) and second stage (115 x 20 runs) were combined to provide the training  
477 data to build a spatio-temporal metamodel. For the second stage analyses, we  
478 utilized a metamodeling-based approach. Metamodels are surrogate models that  
479 can be used as a substitute for the simulation model (Saltelli, Ratto et al. 2008).  
480 The use of metamodels reduces the computational budget, cost of analysis, and  
481 are useful options in cases when the simulation model is expensive to run (in our

case 9-10 minutes for 1 design point) (Saltelli, Ratto et al. 2008). 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 (Rasmussen and  
 Williams 2006, Santner, Williams et al. 2013). 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 (Rasmussen and Williams 2006). A GP is a  
 stochastic process for which any finite set of y-variables has a joint multivariate  
 Gaussian distribution (Thiele, Kurth et al. 2014) (Rasmussen and Williams 2006).  
 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:

$$y_j(w) = Y(w) + \varepsilon_j(w) = \beta_0 + M(w) + \varepsilon_j(w), \quad (1)$$

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  $X \in \chi \subset \mathbb{R}^d$  and the time  $t \in$   
 $T \subset \mathbb{R}_+$ ; 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 (Ankenman, Nelson et al. 2010).

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 (Lamoureux, Mechbal et al. 2014). 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-

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

wherein,  $\exp(-\sum_{r=1}^d \theta_r (X_{ar} - X_{br})^2)$  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  $\exp(-\sum_{r=1}^d \gamma_r (t_{ar} - t_{br})^2)$  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  $\bar{y} = (\bar{y}(w_1), \bar{y}(w_2), \dots, \bar{y}(w_k))^T$ , where in  $\bar{y}(w_i)$  is the resulting estimate of performance measure obtained at design point  $w_i$  and  $\bar{\varepsilon}(w_i)$  is the sampling variability inherent in a stochastic simulation (Ankenman, Nelson, & Staum, 2010). The equations associated with  $\bar{y}(w_i)$  and  $\bar{\varepsilon}(w_i)$  are described below in equation (3):

$$\bar{y}(w_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} y_j(w_i) = Y(w_i) + \bar{\varepsilon}(w_i) \quad \text{and} \quad \bar{\varepsilon}(w_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} \varepsilon_j(w_i), \quad i = 1, 2, \dots, k. \quad (3)$$

Similar as in (Ankenman, Nelson, & Staum, 2010), shown below in equation (4), let  $\Sigma_M$  be the  $k \times k$  covariance matrix across all design points and let  $\Sigma_M(w_o, \cdot)$  be the  $k \times 1$  vector,  $(\text{Cov}[M(w_o, w_1)], \text{Cov}[M(w_o, w_2)], \dots, \text{Cov}[M(w_o, w_k)])^T$  that contains spatial covariance between the  $k$  design points and a given prediction point  $w_o$ . Also, let  $\Sigma_\varepsilon$  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 (Ankenman, Nelson et al. 2010), the best linear predictor  $Y(w_o)$  that has the minimum mean squared error (MSE) among all linear predictors at a given point  $w_o = (X_o^T, t_o)^T$  can be given by equation (4):

$$\hat{Y}(w_o) = \widehat{\beta}_o + \Sigma_M(w_o, \cdot)^T [\Sigma_M + \Sigma_\varepsilon]^{-1} (\bar{y} - 1_k \widehat{\beta}_o), \quad (4)$$

where,  $1_k$  is the  $k \times 1$  vector of ones and  $\widehat{\beta}_o$  is estimated to be 1. The corresponding optimal MSE as in (Ankenman, Nelson et al. 2010) is given by equation (5):

$$\text{MSE}(\hat{Y}(w_o)) = \Sigma_M X_o, w_o - \Sigma_M(w_o, \cdot)^T [\Sigma_M + \Sigma_\varepsilon]^{-1} \Sigma_M(w_o, \cdot) \quad (5).$$

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  $(Y(w_o), \bar{y}^T)^T$  follows a multivariate normal distribution, for e.g., see (Ankenman, Nelson et al. 2010) and (Chen and Kim 2014). The function implemented in the *mlegp* package in R (Dancik and Dorman 2008) is used for the estimation of the parameters. Once the parameters are estimated the prediction then follows equations (4) and (5).

### Sensitivity index calculation

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575

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 (Thiele, Kurth et al. 2014).

We independently generated 10,000 x 38 sampling matrices, such that the parameter combinations are generated via Latin Hypercube sampling and as described in (Saltelli, Annoni et al. 2010). Simulations were performed using the GP spatio-temporal model as described in the previous section, and the Sobol' indices were computed as described in (Sobol 1993) (Saltelli, Annoni et al. 2010). The Sobol' method quantitatively measured the contribution of each input parameter by computing the first order and total order index (Saltelli, Annoni et al. 2010). For output  $Y$ , input parameter matrix  $X_i$  where,  $i$  is the input parameters of the model, the Sobol' indices are computed as follows:

$$SI_1^{X_i} = \frac{V[E(Y|X_i)]}{V(Y)},$$

and

$$SI_{tot}^{X_i} = \frac{V[E(Y|X_{\sim i}])}{V(Y)}.$$

The Sobol' first order sensitivity index  $SI_1^{X_i}$  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 (Saltelli, Annoni et al. 2010) (Sobol', Tarantola et al. 2007) whereas, the total

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

576 order indices were computed using Sobol-Jansen as in (Jansen 1999, Saltelli,  
577 Annoni et al. 2010).

578

579

### 3. Results

#### 3.1 Hybrid model simulations produce similar immune [response](#) dynamics observed in previously published experimental data

583

584

We first aimed to simulate the findings observed in previous gut models (Viladomiu, Bassaganya-Riera et al. 2017) to ensure that we obtained similar response dynamics from the hybrid ENISl model of *H. pylori* infection. As in (Viladomiu, Bassaganya-Riera et al. 2017), 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 (PPAR $\gamma$ ) macrophage-specific knockout (KO) model. PPAR $\gamma$  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 PPARg 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 (Viladomiu, Bassaganya-Riera et al. 2017), monocyte-derived (infiltrating) and

591

592

593

594

595

596

597

598

1  
2  
3  
4 599 macrophage populations with regulatory (M2, or alternatively activated) and pro-  
5  
6 600 inflammatory function (M1 or classically activated) (see *Table 2*).

7  
8  
9 601 We simulated an *in-silico H. pylori* infection by creating four groups, i) a control -  
10  
11 602 WT (representing a wild-type group), ii) CD4Cre (T cell specific PPAR $\gamma$  [KO-lacks](#)  
12  
13 603 [PPAR \$\gamma\$  gene in all CD4 T cells](#)), iii) LysMCre (Myeloid cell specific PPAR $\gamma$  [KO-](#)  
14  
15 604 [lacks PPAR \$\gamma\$  gene in all macrophages](#)) and clodronate group (simulating the  
16  
17 605 removal of macrophages by chemical depletion via clodronate treatment). To  
18  
19 606 simulate the CD4Cre group, the probabilities of a naive T cell transitioning to an  
20  
21 607 iTreg cell ([p\\_nTtoiTreg](#)) and Th17 cell differentiating to iTreg ([p\\_Th17toiTreg](#))  
22  
23 608 were reduced [to 5% and 10% of the control value, respectively \(refer to Table](#)  
24  
25 609 [S1\)](#). As described in (Carbo, Hontecillas et al. 2013), to simulate the LysMCre  
26  
27 610 experimental conditions, the probabilities of i) a monocyte transitioning to a  
28  
29 611 regulatory macrophage ([p\\_Mregdiff](#)) and ii) immature dendritic cells switching to  
30  
31 612 tolerogenic dendritic cells ([p\\_iDCtotDC](#)) were reduced [approximately to 60% and](#)  
32  
33 613 [30% of the control value, respectively \(refer to Table S1\)](#). [A complete set of](#)  
34  
35 614 [parameter for each of the biological KOs are included as separate columns in](#)  
36  
37 615 [Table S1](#). Lastly, the removal of macrophages by clodronate were simulated by  
38  
39 616 decreasing the initial numbers of the macrophage population including the  
40  
41 617 resident macrophages. The rationale to include the clodronate group  
42  
43 618 (macrophage removal) was to evaluate if depletion of phagocytic cells  
44  
45 619 (terminology with respect to model, *i.e.*, monocytes, resident, monocyte-derived  
46  
47 620 macrophages and inflammatory macrophages) would affect *H. pylori*  
48  
49 621 colonization levels, as we have previously reported in an *in vivo* model  
50  
51 622 (Viladomiu, Bassaganya-Riera et al. 2017). Further, [to](#) simulate the myeloid cell  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



1  
2  
3  
4 623 PPAR $\gamma$  KO system, the initial population of resident macrophages were also  
5  
6 624 reduced.

7  
8  
9 625 All the groups were initialized with equal loads of *H. pylori* agents. Ten  
10  
11 626 replicates of the simulations were performed for each of the input parameter  
12  
13 627 settings specific to each group. The outputs were averaged, and standard error  
14  
15 628 of the means were plotted as ribbons (shaded regions) across the graphs. After  
16  
17  
18 629 running the ten replicates of the time series *in-silico* simulation, the hybrid model  
19  
20 630 showed significantly ( $p < 0.05$ ) higher levels of *H. pylori* in the WT and CD4Cre  
21  
22 631 groups as compared to LysMCre KO and macrophage-depleted groups (*Fig. 2,*  
23  
24 632 panel a and d).

25  
26 633  
27  
28 634  
29  
30  
31 635 **Fig 2. Time course simulations representing the immune response during**  
32  
33 636 ***Helicobacter pylori* infection.**

34  
35 637 *The upper half of the plot in both the panels shows the dynamics of the*  
36  
37 638 *population cells*  
38  
39  
40 639 *over time representing the number of cells (y-axis) versus time (x-axis) in a WT*  
41  
42 640 *(black), CD4Cre (green), clodronate (red) and LysMCre (blue) simulated in-silico*  
43  
44 641 *groups during *H. pylori* infection. The cell populations include - a) *H. pylori*; b)*  
45  
46 642 *the resident macrophages and, c) monocyte-derived macrophages in the lamina*  
47  
48  
49 643 *propria compartment. The figures in the lower half (d-f) of both the panels, show*  
50  
51 644 *the results for statistical comparison between the groups using ANOVA with the*  
52  
53 645 *post-hoc analysis. The letters 'a', 'ab' and 'b' represent statistically significant*  
54  
55 646 *differences ( $P < 0.05$ ) between the groups obtained after running the Tukey's*  
56  
57  
58 647 *Honestly Significant Difference. A side by side comparison with the bacterial load*  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 648 [and macrophage population as observed in the mouse model of \*H. pylori\*](#)  
5  
6 649 [infection are also included.](#)  
7  
8  
9 650

10  
11 651 In addition to the increase in *H. pylori*, WT and CD4Cre *in-silico* experimental  
12  
13 652 groups had [a](#) higher resident as well as [monocyte](#)-derived regulatory  
14  
15 653 macrophages as compared to clodronate (macrophage depleted) and LysMCre  
16  
17 654 groups ([Fig. 2b-c, e-f](#)). [The results in the mouse model indicated that between](#)  
18  
19  
20 655 [weeks 2 and 3 post-infection a decrease in bacterial burden in the stomach of](#)  
21  
22 656 [LysMcre mice was observed as shown in Fig 1A of Viladomiu, Bassaganya-Riera](#)  
23  
24 657 [et al. 2017. The decrease in bacterial burden led to a significant and sustained](#)  
25  
26 658 [lower colonization levels when compared to WT and CD4Cre. Similar to the](#)  
27  
28  
29 659 [results observed in the mouse model, we observed a significant decrease \(Fig](#)  
30  
31 660 [2a, d\) in the bacterial burden in the simulated LysMcre group as compared to](#)  
32  
33 661 [the simulated WT and CD4cre groups. Furthermore, the results from the mouse](#)  
34  
35 662 [model indicated that a significant increase in numbers of F4/80hiCD11b+ CD64+](#)  
36  
37 663 [CX3CR1+ cells \(here referred to as resident macrophages in this paper\), was](#)  
38  
39 664 [observed in WT mice in comparison with LysMcre mice as shown in Fig. 2A, 2E](#)  
40  
41 665 [of Viladomiu, Bassaganya-Riera et al. 2017. These cells accumulated in the](#)  
42  
43 666 [stomach mucosa starting on day 14 post-infection in the WT mice but not in the](#)  
44  
45 667 [LysMcre mice. We observed a similar increase \(Fig 2b,e and Fig 2c,f\) in the](#)  
46  
47 668 [number of resident macrophages as well as monocyte derived macrophages in](#)  
48  
49 669 [the simulated WT groups in comparison to the simulated LysMcre group. We](#)  
50  
51 670 [estimated the parameter values to fit the data obtained from the mouse model of](#)  
52  
53 671 [\*H. pylori\* infection.](#) Thus, the observations were qualitatively similar to the  
54  
55  
56  
57 672 findings in (Viladomiu, Bassaganya-Riera et al. 2017), where the stomach of WT  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

673 mice was enriched in a population of F4/80+CD11b+CD64+ myeloid cells,  
674 compared to LysMCre mice.

675 Overall, with the results in *Fig 2*, we showed the ability of the hybrid model to  
676 replicate the experimental results in (Viladomiu, Bassaganya-Riera et al. 2017),  
677 and this preliminary data was used as a base calibration setting for SA and other  
678 *in-silico* findings.

### 679 **3.2 Partial correlation coefficient analysis screened the influential** 680 **parameters**

681  
682 To reduce the computational complexity of varying an input parameter space  
683 of 38 parameters, we divided the SA process in two stages. For first-stage  
684 analysis, we utilized the PRCC regression-based SA method to screen the  
685 influential inputs and used it for the second stage design of the experiments (refer  
686 Methods 2.2). Using PRCC, we determined the impact of the input parameters  
687 on the output cell populations in the model. The parameters with significant  
688 correlation with *H. pylori* in the gastric lamina propria compartment and resident  
689 macrophages are shown in *Fig 3*, along with their PRCC values. The bars in  
690 blue, highlight the parameters that are significantly different than 0, at  $P < 0.05$   
691 compared to grey bars which are not significant. It is important to note that at this  
692 stage the analysis using PRCC was non-temporal.

693 The SA from first stage results showed that the epithelial damage due to  
694 infectious bacteria (*epiinfbctdam*) with a coefficient value of (~0.2), was positively  
695 correlated with the colonization of *H. pylori* in the lamina propria compartment,  
696 indicating the important role of epithelial cell damage during the course of  
697 infection, similar to our findings obtained in (Alam, Deng et al. 2015). 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 (*epiinfbtctdam* = (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).

**Fig 3. 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 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, refer Methods 2.2), were selected to be varied for the second stage design. All the selected inputs are shown in Additional Fig S4. In all, we obtained 23 active inputs from the first stage.

**3.3 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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

723 combined to be used as a training dataset. The combined output was utilized to  
724 build a Gaussian process based spatiotemporal metamodel (refer Methods 2.2),  
725 using *mlegp* package in R (Dancik and Dorman 2008).

726  
727 The outputs from the training dataset were sub-divided into 6 datasets,  
728 corresponding to six time periods (Days 1-14, 15-21, 22-30, 31-42, 43-90, 91-  
729 201) and averaged across these periods. The sub-division of output across the  
730 time periods, aided the temporal analysis over the initiation (Day 1-14), peak of  
731 infection (Days 15-30) and chronic phase (post Day 31) stages as in (Viladomiu,  
732 Bassaganya-Riera et al. 2017). We then fit a Gaussian process model (with  
733 nugget) and evaluated the performance of the fitting of the metamodel for *H.*  
734 *pylori*, resident macrophages, and monocyte-derived macrophages in *lamina*  
735 *propria* compartment, and tolerogenic DC in the gastric lymph node, using the  
736 diagnostic plots (see figures in Additional file, Fig S5). After fitting the models,  
737 we performed variance based global SA by computing the Sobol' total order and  
738 first order sensitivity index (refer Methods 2.2). The estimates of the Sobol' total  
739 order indices for the input parameters calculated over the six time periods are  
740 shown in Fig 4 (a-d).

741  
742  
**Fig 4. Heat-maps of Sobol' total order index for the input parameters across  
743 different output populations.**

744  
745 *The values in the heat-map indicate the Sobol' total order sensitivity index*  
746 *obtained from the metamodel, for the 38 input parameters with respect to the cell*  
747 *populations. The values with darker color indicate a stronger influence on the cell*  
748 *population as compared to the ones with lighter shade that indicate non-*

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

749 influential parameters for the cell populations - a) *H. pylori*, b) monocyte-derived  
750 macrophages, c) resident macrophages, in the lamina propria compartment and  
751 d) tolerogenic DCs, in the gastric lymph node compartment. The indexes are  
752 calculated over six time points ranging across the three stages of infection,  
753 including initiation (Day 1-14), peak (Days 15-42) and recovery stages (Days 43-  
754 201).

As shown in Fig 4a, 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. Further, Fig 4b 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 (Viladomiu, Bassaganya-Riera et al. 2017), 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 (Viladomiu, Bassaganya-Riera et al. 2017) wherein we observed that these subsets of macrophages [expand in the gastric](#) stomach lamina propria during the course of *H. pylori* infection.

1  
2  
3  
4 775 Finally, for the tolerogenic DCs in *Fig 4d*, we observed that the epithelial cell  
5  
6 776 death (*Epicelldeath*) seemed to have an impact. Another parameter that stands  
7  
8  
9 777 for the probability of naive T cell transitioning to iTreg cell (*nTtoiTreg*) was shown  
10  
11 778 to have an impact on the tolerogenic dendritic cells. Tolerogenic dendritic cells  
12  
13 779 are involved in the rule that transitions the naive T cells to iTreg cells in the gastric  
14  
15 780 lymph node, and the stronger impact of the *nTtoiTreg* during the initiation and  
16  
17 781 peak stages of the infection highlights the role of the tolerogenic dendritic cells  
18  
19  
20 782 during the course of infection.

21  
22 783 The global SA data suggested that the main contributors of the chronic  
23  
24 784 colonization of *H. pylori* in the lamina propria are the epithelial cells, specifically  
25  
26 785 the epithelial cell proliferation parameter.

### 30 786 **3.4 Effect of different ranges of epithelial cell proliferation**

31  
32  
33 787 An interesting prediction derived from the metamodel based global SA is that  
34  
35 788 epithelial cell proliferation is one of the parameters that has a strong impact on  
36  
37 789 the size of *H. pylori* population. The biological hypothesis derived from this  
38  
39 790 prediction is that the epithelial cell proliferation is responsible for the higher  
40  
41 791 colonization of *H. pylori*. Prior to conducting any experimental studies, we wanted  
42  
43 792 to explore the hypothesis using our hybrid computer model *in silico* and study  
44  
45 793 the model outputs obtained after we changed the epithelial cell proliferation  
46  
47 794 parameter. Thus, we varied the epithelial cell proliferation parameter across  
48  
49 795 different ranges (0.1-0.9, with 0.6 being the value for baseline conditions) and  
50  
51 796 ran the simulations using the hybrid model and studied its effect on the different  
52  
53 797 output cell population (obtained after running the simulations). These outputs  
54  
55 798 were the ones obtained after running the simulation using the hybrid computer  
56  
57 799 model, as we varied the epithelial cell proliferation parameter. We analyzed the  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

800 [outputs from the hybrid computer model and interestingly, observed that](#) upon  
801 decreasing the *Epiprolifer* from a range of values 0.9-0.1, [the output cell](#)  
802 [populations with regulatory function, namely regulatory macrophages and](#)  
803 [tolerogenic dendritic cells were found to vary.](#) We observed a decreasing effect  
804 (Fig 5a-d) on *H. pylori*, monocyte-derived macrophages, resident macrophages  
805 in the lamina propria compartment and tolerogenic dendritic cells in gastric lymph  
806 node. [Overall, these cell populations varied due to the variation in the epithelial](#)  
807 [cell proliferation parameter.](#)

808 [For clarification, such connection was not embedded in the mechanisms](#)  
809 [included in Table 1 but it represents an emergent behavior from the simulations](#)  
810 [predicting the involvement of regulatory and tolerogenic dendritic cells in the](#)  
811 [mechanisms of immunoregulation during \*H. pylori\* infection. Finally, the](#)  
812 [simulations targeting the epithelial cell proliferation resulted in changes in](#)  
813 [regulatory and tolerogenic dendritic cell populations. This shows that the](#)  
814 [simulations indirectly targeted the regulatory and tolerogenic dendritic cell](#)  
815 [population. Thus, we hypothesize that epithelial cell proliferation might be](#)  
816 [responsible for the higher colonization of \*H. pylori\* through an immunoregulatory](#)  
817 [mechanism that involves regulatory macrophages and tolerogenic cells. This is](#)  
818 [in line with our own conclusions drawn from a previous paper \(Viladomiu,](#)  
819 [Bassaganya-Riera et al. 2017\) where we show that the presence of cells with](#)  
820 [regulatory phenotype favor higher levels of \*H. pylori\* colonization. The results](#)  
821 [from the sensitivity analysis presented in this paper suggest that epithelial](#)  
822 [proliferation might be a crucial part of the mechanisms by which these regulatory](#)  
823 [responses are induced and that there is a link between these parameters. The](#)  
824 [exact biological process however cannot be inferred from the current model and](#)  
825 [it will be investigated in follow-up in vivo studies.](#)



1  
2  
3  
4 826  
5  
6 827  
7  
8 828  
9  
10  
11 829  
12  
13 830  
14  
15 831  
16  
17 832  
18  
19 833  
20  
21 834  
22  
23 835  
24  
25 836  
26  
27 837  
28  
29 838  
30  
31 839  
32  
33 840  
34  
35 841  
36  
37 842  
38  
39 843  
40  
41 844  
42  
43 845  
44  
45 846  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

**Fig 5. *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 ( $p_{Epi}$  proliferation) 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 *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.*

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

847

#### 848 4. Discussion

849 *H. pylori* is the dominant indigenous bacterium of the gastric microbiota. In the  
850 majority of individuals, *H. pylori* colonizes the stomach without causing adverse  
851 effects, with little to no activation of inflammatory pathways. However, certain  
852 members of the population lose immune tolerance to the bacterium thereby  
853 contributing to the development of chronic gastric diseases. The immunological  
854 mechanisms underlying its ability to persist in a harsh acidic gastric environment and  
855 its dual role as a pathogen and beneficial organism remain unknown. A subset of  
856 macrophages helps create a regulatory microenvironment that promotes the chronic  
857 colonization of *H. pylori* (Viladomiu, Bassaganya-Riera et al. 2017). However, the  
858 immune regulatory mechanisms are incompletely understood. Computational  
859 models of the immune system featuring immune responses are powerful tools for  
860 testing the different ‘what-if’ scenarios. Multiscale models of the immune response  
861 are attractive in terms of modeling the responses at different spatiotemporal scales  
862 (Heiner and Gilbert 2013).

863 In this study, we developed a HPC-driven hybrid, high-resolution, multiscale  
864 model to simulate the complex immunoregulatory mechanisms during *H. pylori*  
865 infection. The hybrid model was integrated with two intracellular ODEs capturing the  
866 dynamics of CD4+ T cells and regulatory macrophages. The inputs to the hybrid  
867 model are the set of parameters whose variation governs the immune system  
868 dynamics during infection. The obtained outputs were emergent patterns of different  
869 cell types, cytokines, and bacterial levels for instance the levels of *H. pylori*, and that  
870 qualitatively matched the patterns observed in an *in vivo* infection model (Carbo,  
871 Bassaganya-Riera et al. 2013, Viladomiu, Bassaganya-Riera et al. 2017). We

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

872 presented an *in-silico* framework that evaluated the global SA of the hybrid model  
873 and studied how the variation in the biological parameters affected the simulation  
874 outputs. The two-stage global SA indicated that epithelial cell parameters,  
875 specifically, the proliferation of epithelial cells affected the colonization of *H. pylori* in  
876 the gastric mucosa. These results were validated *in silico*, and highlighted the  
877 involvement of regulatory macrophages and tolerogenic DC in facilitating *H. pylori*  
878 colonization of the gastric mucosa. Previous studies highlighted *H. pylori* inhabits the  
879 apical surfaces of the epithelial cells and maintains a persistent infection (Alzahrani,  
880 Lina et al. 2014).

881 Further, Mimuro et al. demonstrated that *H. pylori* promotes epithelial gastric cell  
882 survival by attenuating apoptosis. These events showed how *H. pylori* regulated the  
883 gastric niche and utilized epithelial cells to facilitate its persistence within the  
884 stomach (Mimuro, Suzuki et al. 2007) (Wroblewski and Peek 2007). Thus, the  
885 findings in the current study are in line with the literature that suggests epithelial cell  
886 proliferation favor the colonization of *H. pylori* in the stomach.

887 Our group also showed another mechanism used by *H. pylori* to create a gut  
888 microenvironment that involved the induction of IL-10-driven regulatory mechanism  
889 mediated by CD11b<sup>+</sup>F4/80<sup>hi</sup>CD64<sup>+</sup>CX<sub>3</sub>CR1<sup>+</sup> mononuclear phagocytes, which  
890 facilitated [bacterial](#) colonization (Viladomiu, Bassaganya-Riera et al. 2017).  
891 Additionally, in this paper, we reported that regulatory macrophages were involved  
892 in the process of colonization with *H. pylori* when we varied the epithelial cell  
893 proliferation parameter *in-silico*. Zhang et al., demonstrated that *H. pylori* directed  
894 active tolerogenic programming of DCs that favored chronic bacterial colonization,  
895 by altering the balance of Th17/Treg cells (Zhang, Liu et al. 2010). Rizzuti, Ang et  
896 al., demonstrated *H. pylori*-mediated IL-10 release caused the activation of signal  
897 transducer and activator of transcription 3 (STAT3) in DC. This activation of STAT3

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

898 via IL-10 release was shown to induce the production of tolerogenic DC phenotype  
899 (Rizzuti, Ang et al. 2015). The findings from this paper also indicated the involvement  
900 of tolerogenic DCs in affecting the mucosal levels of *H. pylori*. Therefore, the  
901 literature combined with the results from this study, collectively suggest that during  
902 *H. pylori* infection, the epithelial cell favors the colonization of *H. pylori* by creating a  
903 regulatory microenvironment. This process is mediated by the regulatory  
904 macrophages and tolerogenic programming of DC. Based on the results from this  
905 paper and findings from the literature, this leads us to propose that the induction of  
906 IL-10 by the regulatory macrophages is potentially involved in directing the  
907 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.](#)

915  
916 At its current stage, the hybrid ENISI model reproduces the overall immune  
917 system dynamics observed during an *H. pylori* infection. The parameters of  
918 calibrated ODEs were kept unchanged, whereas the ABM parameters were  
919 calibrated by qualitatively matching the patterns of the output simulations as  
920 observed in an *in vivo* model of *H. pylori* infection (Viladomiu, Bassaganya-Riera et  
921 al. 2017). For ABM, its calibration and validation remain the major key issues,  
922 discussed elsewhere (Ten Broeke, Van Voorn et al. 2016) (Windrum, Fagiolo et al.  
923 2007) (Fagiolo, Moneta et al. 2007). Further, developing targeted methods of SA

1  
2  
3  
4 924 have been identified as an important challenge in the field (Crooks, Castle et al.  
5  
6 925 2008, Filatova, Verburg et al. 2013, Ten Broeke, Van Voorn et al. 2016). In this  
7  
8 926 paper, we highlighted the use of SA methods with a two-stage global SA framework  
9  
10 927 comprised of first, screening the input parameters (using PRCC) and second,  
11  
12 928 building of a surrogate model (using GP) of the hybrid model, to understand the  
13  
14 929 emergent behavior of the represented system. It is important to note that each SA  
15  
16 930 method known, has its own merits and produces useful information however none  
17  
18 931 provide a complete picture of the emergent model behavior (Ten Broeke, Van Voorn  
19  
20 932 et al. 2016). First, we employed PRCC methods as the initial step in our two staged  
21  
22 933 SA that aided the screening of active inputs and reduced the parameter space. The  
23  
24 934 choice of PRCC was advantageous and justified by the low computational cost and  
25  
26 935 low complexity in the computation of the coefficients. Another advantage of the  
27  
28 936 regression-based PRCC method is that the complex output from our hybrid model  
29  
30 937 was condensed into a descriptive relationship that can be described by statistical  
31  
32 938 measures such as  $R^2$  (Ten Broeke, Van Voorn et al. 2016). As described in (Ten  
33  
34 939 Broeke, Van Voorn et al. 2016) the results from PRCC are good descriptors of the  
35  
36 940 outputs produced if the regression function constitutes a good fit to the output (Ten  
37  
38 941 Broeke, Van Voorn et al. 2016). However, if the function does not yield a good fit,  
39  
40 942 the regression-based SA are proven to be useful in screening the influential  
41  
42 943 parameters for further analysis (Ten Broeke, Van Voorn et al. 2016), as described in  
43  
44 944 our analysis.

50  
51 945 Further, the interaction effects between the parameters are not considered in  
52  
53 946 regression-based methods, and hence it was followed by the use of variance-based  
54  
55 947 methods in later stage analysis. Second, we employed metamodeling-based  
56  
57 948 approach and Sobol' method as they provided information on the interaction  
58  
59 949 between the input variable and the use of metamodels allowed to compute the  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

950 sensitivity indices. One of the advantages of the Sobol' method is that it is model-  
951 free and no fitting functions are used to decompose the output variance (Saltelli,  
952 Ratto et al. 2008). It considers the averaged effect of parameters over the whole  
953 parameter space but fails to explore the different patterns within the space (Ten  
954 Broeke, Van Voorn et al. 2016). Further, the method is not suitable for quantification  
955 of output variability if the output distributions deviate from a normal distribution (Ten  
956 Broeke, Van Voorn et al. 2016). The detailed comparison of different SA methods  
957 used for the global SA of ABMs are described in detail in (Ten Broeke, Van Voorn et  
958 al. 2016). Thus, we performed both the PRCC and computation of Sobol' indices  
959 approaches to evaluate the influence of the input parameter variation and identified  
960 the parameters involved in the successful colonization of the gastric niche by *H.*  
961 *pylori*.

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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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

### 5. Potential Implications

987 The computational model of the gut contains high-resolution information  
988 processing representations of immune responses that are generalizable for other  
989 infectious and autoimmune diseases. Complex diseases such as autoimmune  
990 disorders, infectious diseases, and cancer all require integration of the multiscale  
991 level data, information and knowledge, ranging from genes, proteins, cells, tissue to  
992 organ level. The ENISI model of the gut presented here can be generalized to other  
993 diseases by implementing the agents and rules specific to that disease, plus  
994 recalibrating the model based on data that are specific to the new indication. Since  
995 ABMs have modular architectures, an addition of new agent-types and modification  
996 of rules can be done without restructuring the entire simulation setup (An, Mi et al.  
997 2009). The use of ABM in such hybrid models not only facilitates the implementation  
998 of already known mechanisms but also helps validate and predict any unforeseen  
999 new mechanisms using data analytics methods such as global SA to analyze  
1000 emerging behaviors at the systems level. The finer details regarding intracellular and  
1001 intercellular interactions that contribute towards the nonlinear and complex behavior

1  
2  
3  
4 1002 of the gut can also be studied by integrating the intracellular ODE models as  
5  
6 1003 implemented here.  
7  
8  
9 1004

10  
11 1005 **Tables**

Name of Agent	Agent Type	Rules
<i>Helicobacter pylori</i>	<i>H. pylori</i>	<ul style="list-style-type: none"> <li>- Moves across the epithelial cell border if near damaged epithelial layer</li> <li>- Proliferates in the lumen and lamina propria</li> <li>- Dies (removed from the simulation) in lamina propria and in the lumen due to the damage of epithelial cells by Th1 or Th17 cells</li> </ul>
Macrophages	Monocyte	<ul style="list-style-type: none"> <li>- Proliferates in presence <u>the</u> of effector dendritic cells or damaged epithelial cells</li> <li>- Proliferates in the lamina propria</li> <li>- Differentiates to regulatory macrophage in based on the output from the Macrophage ODE</li> <li>- Differentiates to inflammatory macrophages in presence of IFN-<math>\gamma</math></li> <li>- Dies naturally (removed from the model)</li> </ul>
	Resident	<ul style="list-style-type: none"> <li>- Proliferates in the presence of <i>H. pylori</i></li> <li>- Secretes IL10</li> <li>- Dies naturally</li> <li>- Dies due to Th1 and Tr cells</li> </ul>
	Regulatory	<ul style="list-style-type: none"> <li>- Proliferates and removes bacteria</li> <li>- Dies</li> <li>- Secretes IL10</li> </ul>
	Inflammatory	<ul style="list-style-type: none"> <li>-Proliferates in the presence of damaged epithelial cell</li> <li>-Dies naturally</li> </ul>
Dendritics	Immature	<ul style="list-style-type: none"> <li>-Moves from lamina propria to epithelium compartment and from the epithelium to the lamina propria</li> <li>- Differentiates to tolerogenic dendritic cell in the presence of tolerogenic bacteria, both in epithelium and lamina propria</li> <li>- Differentiates to effector dendritic cell in the presence of <i>H. pylori</i></li> <li>- Proliferates in lamina propria and gastric lymph node</li> <li>- Dies naturally</li> </ul>
	Effector	<ul style="list-style-type: none"> <li>- Moves from lamina propria to gastric lymph node</li> <li>- Moves form epithelium to lamina propria</li> <li>- Secretes IL6 and IL12</li> <li>- Dies naturally</li> </ul>
	Tolerogenic	<ul style="list-style-type: none"> <li>- Moves from lamina propria to gastric lymph node</li> <li>- Moves from epithelium to lamina propria</li> <li>- Secretes TGF-<math>\beta</math></li> <li>- Dies naturally</li> </ul>



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

T cells	Naïve	In the presence of effector dendritic cells: <ul style="list-style-type: none"> <li>- Differentiates to Th1 in the presence of IFN-<math>\gamma</math> or IL12</li> <li>- Differentiates to Th17 in the presences of IL6 or TGF-<math>\beta</math></li> </ul> In the presence of tolerogenic dendritic cells: <ul style="list-style-type: none"> <li>- Differentiates to iTreg in the presence of TGF-<math>\beta</math></li> <li>- Differentiates to Tr in the presences of IL10</li> <li>- Dies naturally</li> </ul>
	Th1	<ul style="list-style-type: none"> <li>- Secretes IFN-<math>\gamma</math></li> <li>- Moves from gastric lymph node to lamina propria</li> <li>- Proliferates in lamina propria and gastric lymph node</li> <li>- Dies naturally</li> </ul>
	Th17	<ul style="list-style-type: none"> <li>- Secretes IL17</li> <li>- In <u>the</u> presence of tolerogenic dendritic cell, transition to iTreg cells</li> <li>- Moves from gastric lymph node to lamina propria</li> <li>- Proliferates in lamina propria and gastric lymph node</li> <li>- Dies naturally</li> </ul>
	iTreg	<ul style="list-style-type: none"> <li>- Secretes IL10</li> <li>- In the presence of tolerogenic dendritic cell, transition to iTreg cells</li> <li>- Moves from gastric lymph node to lamina propria</li> <li>- Proliferates in lamina propria and gastric lymph node</li> <li>- Dies naturally</li> </ul>
	Tr	<ul style="list-style-type: none"> <li>- Secretes IL10</li> <li>- Dies naturally</li> <li>- Proliferates in the lamina propria</li> </ul>
Epithelial	Healthy	<ul style="list-style-type: none"> <li>-Damaged due to infectious bacteria</li> <li>-Damaged due to Th1 and Th17 cells</li> <li>-Proliferates</li> <li>-Secretes IL6 and IL12</li> <li>-Dies naturally</li> </ul>
	Damaged	<ul style="list-style-type: none"> <li>-<u>Transitions</u> to healthy state in the presence of IL10</li> <li>-Dies naturally</li> </ul>
Bacteria	Infectious	<ul style="list-style-type: none"> <li>- Dies due to Th1 or Th17 or inflammatory macrophages or damaged epithelial cells</li> <li>- Dies naturally</li> <li>- Proliferates in the lamina propria</li> </ul>
	Tolerogenic	<ul style="list-style-type: none"> <li>- Moves from lumen to the epithelium in the presence of damaged epithelial cells</li> <li>- Becomes infectious if moves in the lamina propria compartment</li> <li>- Proliferates in lumen and lamina propria</li> <li>- Dies naturally</li> </ul>

**Table 1. A list of rules for all the agent types implemented in the hybrid model**

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

**Additional Files**

[File S1](#)

Fig S1

Table S1

Fig S2

Fig S3

Fig S4

[Fig S5](#)

[File S1 – The detailed instruction to Install ENISI MSM \(Step I\), Run a simulation \(Step II\) and Conduct Sensitivity Analysis \(Step III\) are described.](#)

**Fig S1. Design implementation of the hybrid multiscale model used to simulate *Helicobacter pylori* infection**

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 S1** Table describing the input parameters used in the sensitivity analysis and their ranges used.

[Fig S2. A pictorial representation of the spatial discretization of the 2D grid.](#)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1033  
1034 **Fig S3. Flowchart for the two-staged global sensitivity analysis.**

1035  
1036 **Fig S4. The active and inactive inputs selected from the stage 1 analysis**

1037 *The rows represent the input parameters and columns represent the output cell*  
1038 *populations. The green boxes highlight the ‘active’ input parameters (row) that are*  
1039 *shown to have a significant influence (calculated based on the results obtained from*  
1040 *partial correlation coefficient analysis), on an output cell (columns) under*  
1041 *consideration.*

1042  
1043 **Fig S5. Diagnostic and residual plots obtained for the Gaussian processes**  
1044 **fitted metamodels**

1045 The upper panel represents the diagnostic Q-Q plots where the open circles  
1046 represent the cross-validated predictions; solid black lines represent observed  
1047 response. The “observed simulations” data in the first half of the lower panel, refer  
1048 to the observed output values of the simulations obtained after running the hybrid  
1049 computer model, whereas the y axis refers to the predicted simulation values  
1050 obtained from the Cross-validated model. Each point represents 1 output point  
1051 obtained as an output from the simulation. The second half of the lower panel, refers  
1052 to the standard residual plot wherein the x-axis represents the observed simulation  
1053 values obtained from the simulation and the y-axis refers to the residual error ((error  
1054 (predicted values – observed values) / standard deviation (error))) obtained. The  
1055 diagnostic plots denote the black circles which are the cross-validated prediction.  
1056 Cross-validation is in the sense that for predictions made at design point x, all  
1057 observations at design point x are removed from the training set. The lower panel  
1058 *represents the residual plots for the cell populations –(a) Helicobacter pylori; (b)*

1  
2  
3  
4 1059 *Resident macrophages; (c) Monocyte-derived macrophages in the Lamina propria*  
5  
6 1060 *and (d) Tolerogenic dendritic cells in the Gastric lymph node compartment.*  
7  
8  
9

10 1061  
11  
12  
13 1062

14  
15 1063 **Availability of source code and requirements**  
16

- 17  
18 1064
  - Project Name: ENISI MSM

19  
20 1065
  - Project homepage: <https://github.com/NIMML/ENISI-MSM>

21  
22 1066
  - Programming language: C++, R

23  
24

25  
26 1067 **Availability of supporting data and materials**  
27

28  
29 1068 The data sets and files supporting the results of this article are available in the ENISI-MSM  
30 1069 GitHub repository, <https://github.com/NIMML/ENISI-MSM>.  
31

32  
33 1070 **Declarations**  
34

35 1071 **List of abbreviations**  
36

- 37 1072 ABM – Agent based model  
38  
39 1073 DC – Dendritic cells  
40  
41 1074 ENISI MSM – Enteric Immunity Simulator Multi-scale Modeling  
42  
43 1075 GLN – gastric lymph node  
44  
45 1076 GP - Gaussian process  
46  
47 1077 *H. pylori* – *Helicobacter pylori*  
48  
49  
50 1078 HPC – High performance computing  
51  
52 1079 LP – Lamina propria  
53  
54 1080 ODE – Ordinary Differential Equation  
55  
56 1081 PDE – Partial Differential Equation  
57  
58  
59 1082 SA – Sensitivity analysis  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 1083 PRCC - Partial rank correlation coefficient

5  
6 1084 **Consent for publication**

7  
8  
9 1085 Not applicable.

10  
11 1086 **Competing interests**

12  
13 1087 The author(s) declare that they have no competing interests.

14  
15 1088 **Authors' contributions**

16  
17 1089 MV, RH and JBR formulated the model, implemented, performed the simulations,  
18  
19  
20 1090 analyzed model-generate outputs, made the figures and wrote the manuscript. MV,  
21  
22 1091 AL, JBR, RH, and SH formulated the model. SH, AL and VA implemented the code  
23  
24 1092 architecture and benchmarked the parallel version of the hybrid model. XC and MV  
25  
26 1093 wrote the codes for global sensitivity analysis and generated the design matrices.

27  
28  
29 1094 [NTJ generated macrophage and \*H. pylori\* experimental data.](#) JBR, VA, and RH  
30  
31 1095 supervised the project. JBR and RH edited the manuscript. JBR, AL, NTJ, SH, VA,  
32  
33 1096 XC and RH participated in discussions on the model and results. All authors provided  
34  
35 1097 critical feedback on the project.

36  
37  
38 1098 **Acknowledgements**

39  
40 1099 This work was supported by the Defense Threat Reduction Agency (DTRA) grant HDTRA1-18-  
41  
42 1100 1-0008 to JBR and RH and funds from the Nutritional Immunology and Molecular Medicine  
43  
44 1101 Laboratory ([www.nimml.org](http://www.nimml.org)).

45  
46  
47 1102 .

48  
49 1103  
50  
51 1104 **References**

52  
53 1105 Abedi, V., R. Hontecillas, S. Hoops, N. Liles, A. Carbo, P. Lu, C. Philipson and J. ABassaganya-  
54 1106 Riera (2015). ENISI multiscale modeling of mucosal immune responses driven by high  
55 1107 performance computing. 2015 IEEE International Conference on Bioinformatics and Biomedicine  
56 1108 (BIBM).

57  
58 1109 Alam, M., X. Deng, C. Philipson, J. Bassaganya-Riera, K. Bisset, A. Carbo, S. Eubank, R.  
59 1110 Hontecillas, S. Hoops, Y. Mei, V. Abedi and M. Marathe (2015). "Sensitivity Analysis of an ENteric  
60

1  
2  
3  
4 1111 Immunity Simulator (ENISI)-Based Model of Immune Responses to Helicobacter pylori Infection."  
5 1112 PLoS One **10**(9): e0136139.  
6 1113 Alzahrani, S., T. T. Lina, J. Gonzalez, I. V. Pinchuk, E. J. Beswick and V. E. Reyes (2014). "Effect  
7 1114 of Helicobacter pylori on gastric epithelial cells." World Journal of Gastroenterology : WJG **20**(36):  
8 1115 12767-12780.  
9 1116 An, G., Q. Mi, J. Dutta-Moscato and Y. Vodovotz (2009). "Agent-based models in translational  
10 1117 systems biology." Wiley Interdiscip Rev Syst Biol Med **1**(2): 159-171.  
11 1118 Ankenman, B., B. L. Nelson and J. Staum (2010). "Stochastic kriging for simulation  
12 1119 metamodeling." Operations research **58**(2): 371-382.  
13 1120 Bassaganya-Riera, J. (2015). Computational Immunology: Models and Tools, Academic Press.  
14 1121 Bassaganya-Riera, J., M. G. Dominguez-Bello, B. Kronsteiner, A. Carbo, P. Lu, M. Viladomiu, M.  
15 1122 Pedragosa, X. Zhang, B. W. Sobral, S. P. Mane, S. K. Mohapatra, W. T. Horne, A. J. Guri, M.  
16 1123 Groeschl, G. Lopez-Velasco and R. Hontecillas (2012). "Helicobacter pylori colonization  
17 1124 ameliorates glucose homeostasis in mice through a PPAR  $\gamma$ -dependent mechanism." PLoS One  
18 1125 **7**(11): e50069.  
19 1126 Cappuccio, A., P. Tieri and F. Castiglione (2016). "Multiscale modelling in immunology: a review."  
20 1127 Brief Bioinform **17**(3): 408-418.  
21 1128 Carbo, A., J. Bassaganya-Riera, M. Pedragosa, M. Viladomiu, M. Marathe, S. Eubank, K.  
22 1129 Wendelsdorf, K. Bisset, S. Hoops, X. Deng, M. Alam, B. Kronsteiner, Y. Mei and R. Hontecillas  
23 1130 (2013). "Predictive computational modeling of the mucosal immune responses during  
24 1131 Helicobacter pylori infection." PLoS One **8**(9): e73365.  
25 1132 Carbo, A., R. Hontecillas, B. Kronsteiner, M. Viladomiu, M. Pedragosa, P. Lu, C. W. Philipson, S.  
26 1133 Hoops, M. Marathe, S. Eubank, K. Bisset, K. Wendelsdorf, A. Jarrah, Y. Mei and J. Bassaganya-  
27 1134 Riera (2013). "Systems modeling of molecular mechanisms controlling cytokine-driven CD4+ T  
28 1135 cell differentiation and phenotype plasticity." PLoS Comput Biol **9**(4): e1003027.  
29 1136 Chen, X. and K.-K. Kim (2014). "Stochastic kriging with biased sample estimates." ACM Trans.  
30 1137 Model. Comput. Simul. **24**(2): 1-23.  
31 1138 Collier, N. and M. North (2011). "Repast HPC: A platform for large-scale agentbased modeling."  
32 1139 Large-Scale Computing Techniques for Complex System Simulations: 81-110.  
33 1140 Cover, T. L. and M. J. Blaser (2009). "Helicobacter pylori in health and disease." Gastroenterology  
34 1141 **136**(6): 1863-1873.  
35 1142 Crooks, A., C. Castle and M. Batty (2008). "Key challenges in agent-based modelling for geo-  
36 1143 spatial simulation." Computers, Environment and Urban Systems **32**(6): 417-430.  
37 1144 Dancik, G. M. and K. S. Dorman (2008). "mleqp: statistical analysis for computer models of  
38 1145 biological systems using R." Bioinformatics **24**(17): 1966-1967.  
39 1146 Fagiolo, G., A. Moneta and P. Windrum (2007). "A critical guide to empirical validation of agent-  
40 1147 based models in economics: Methodologies, procedures, and open problems." Computational  
41 1148 Economics **30**(3): 195-226.  
42 1149 Filatova, T., P. H. Verburg, D. C. Parker and C. A. Stannard (2013). "Spatial agent-based models  
43 1150 for socio-ecological systems: Challenges and prospects." Environmental modelling & software  
44 1151 **45**: 1-7.  
45 1152 Gong, C., O. Milberg, B. Wang, P. Vicini, R. Narwal, L. Roskos and A. S. Popel (2017). "A  
46 1153 computational multiscale agent-based model for simulating spatio-temporal tumour immune  
47 1154 response to PD1 and PDL1 inhibition." J R Soc Interface **14**(134).  
48 1155 Heiner, M. and D. Gilbert (2013). "BioModel engineering for multiscale Systems Biology."  
49 1156 Progress in Biophysics and Molecular Biology **111**(2): 119-128.  
50 1157 Hoops, S., S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes and  
51 1158 U. Kummer (2006). "COPASI—a complex pathway simulator." Bioinformatics **22**(24): 3067-3074.  
52 1159 Jansen, M. J. (1999). "Analysis of variance designs for model output." Computer Physics  
53 1160 Communications **117**(1-2): 35-43.  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Kusters, J. G., A. H. van Vliet and E. J. Kuipers (2006). "Pathogenesis of *Helicobacter pylori* infection." Clin Microbiol Rev **19**(3): 449-490.

Lamoureux, B., N. Mechbal and J.-R. Massé (2014). "A combined sensitivity analysis and kriging surrogate modeling for early validation of health indicators." Reliability Engineering & System Safety **130**: 12-26.

Leber, A., J. Bassaganya-Riera, N. Tubau-Juni, V. Zoccoli-Rodriguez, M. Viladomiu, V. Abedi, P. Lu and R. Hontecillas (2016). "Modeling the Role of Lanthionine Synthetase C-Like 2 (LANCL2) in the Modulation of Immune Responses to *Helicobacter pylori* Infection." PLoS One **11**(12): e0167440.

Leber, A., M. Viladomiu, R. Hontecillas, V. Abedi, C. Philipson, S. Hoops, B. Howard and J. Bassaganya-Riera (2015). "Systems Modeling of Interactions between Mucosal Immunity and the Gut Microbiome during *Clostridium difficile* Infection." PLoS One **10**(7): e0134849.

Ligmann-Zielinska, A., D. B. Kramer, K. Spence Cheruvelil and P. A. Soranno (2014). "Using uncertainty and sensitivity analyses in socioecological agent-based models to improve their analytical performance and policy relevance." PLoS One **9**(10): e109779.

Mane, S., M. Dominguez-Bello, M. Blaser, B. Sobral, R. Hontecillas, J. Skoneczka, S. Mohapatra, O. Crasta, C. Evans and T. Modise (2010). "Host-interactive genes in Amerindian *Helicobacter pylori* diverge from their Old World homologs and mediate inflammatory responses." Journal of bacteriology **192**(12): 3078-3092.

Marino, S., M. El-Kebir and D. Kirschner (2011). "A hybrid multi-compartment model of granuloma formation and T cell priming in tuberculosis." J Theor Biol **280**(1): 50-62.

Marino, S., I. B. Hogue, C. J. Ray and D. E. Kirschner (2008). "A methodology for performing global uncertainty and sensitivity analysis in systems biology." J Theor Biol **254**(1): 178-196.

Mei, Y., V. Abedi, A. Carbo, X. Zhang, P. Lu, C. Philipson, R. Hontecillas, S. Hoops, N. Liles and J. Bassaganya-Riera (2015). "Multiscale modeling of mucosal immune responses." BMC Bioinformatics **16 Suppl 12**: S2.

Mimuro, H., T. Suzuki, S. Nagai, G. Rieder, M. Suzuki, T. Nagai, Y. Fujita, K. Nagamatsu, N. Ishijima, S. Koyasu, R. Haas and C. Sasakawa (2007). "*Helicobacter pylori* dampens gut epithelial self-renewal by inhibiting apoptosis, a bacterial strategy to enhance colonization of the stomach." Cell Host Microbe **2**(4): 250-263.

Moon, H., A. M. Dean and T. J. Santner (2012). "Two-stage sensitivity-based group screening in computer experiments." Technometrics **54**(4): 376-387.

Oertli, M., M. Sundquist, I. Hitzler, D. B. Engler, I. C. Arnold, S. Reuter, J. Maxeiner, M. Hansson, C. Taube, M. Quiding-Jarbrink and A. Muller (2012). "DC-derived IL-18 drives Treg differentiation, murine *Helicobacter pylori*-specific immune tolerance, and asthma protection." J Clin Invest **122**(3): 1082-1096.

Qomlaqi, M., F. Bahrami, M. Ajami and J. Hajati (2017). "An extended mathematical model of tumor growth and its interaction with the immune system, to be used for developing an optimized immunotherapy treatment protocol." Math Biosci **292**: 1-9.

Rasmussen, C. E. and C. K. Williams (2006). "Gaussian processes for machine learning. 2006." The MIT Press, Cambridge, MA, USA **38**: 715-719.

Rizzuti, D., M. Ang, C. Sokollik, T. Wu, M. Abdullah, L. Greenfield, R. Fattouh, C. Reardon, M. Tang, J. Diao, C. Schindler, M. Cattral and N. L. Jones (2015). "*Helicobacter pylori* inhibits dendritic cell maturation via interleukin-10-mediated activation of the signal transducer and activator of transcription 3 pathway." J Innate Immun **7**(2): 199-211.

Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto and S. Tarantola (2010). "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index." Computer Physics Communications **181**(2): 259-270.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana and S. Tarantola (2008). Global sensitivity analysis: the primer, John Wiley & Sons.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Saltelli, A., S. Tarantola and F. Campolongo (2000). "Sensitivity analysis as an ingredient of modeling." Statistical Science **15**(4): 377-395.

Santner, T. J., B. J. Williams and W. I. Notz (2013). The design and analysis of computer experiments, Springer Science & Business Media.

Sobol, I. M. (1993). "Sensitivity estimates for nonlinear mathematical models." Mathematical modelling and computational experiments **1**(4): 407-414.

Sobol', I. M., S. Tarantola, D. Gatelli, S. S. Kucherenko and W. Mauntz (2007). "Estimating the approximation error when fixing unessential factors in global sensitivity analysis." Reliability Engineering & System Safety **92**(7): 957-960.

Solovyev, A., Q. Mi, Y. T. Tzen, D. Brienza and Y. Vodovotz (2013). "Hybrid equation/agent-based model of ischemia-induced hyperemia and pressure ulcer formation predicts greater propensity to ulcerate in subjects with spinal cord injury." PLoS Comput Biol **9**(5): e1003070.

Ten Broeke, G., G. Van Voorn and A. Ligtenberg (2016). "Which sensitivity analysis method should I use for my agent-based model?" Journal of Artificial Societies and Social Simulation **19**(1): 5.

Thiele, J. C., W. Kurth and V. Grimm (2014). "Facilitating parameter estimation and sensitivity analysis of agent-based models: A cookbook using NetLogo and R." Journal of Artificial Societies and Social Simulation **17**(3): 11.

Thorne, B. C., A. M. Bailey and S. M. Peirce (2007). "Combining experiments with multi-cell agent-based modeling to study biological tissue patterning." Briefings in Bioinformatics **8**(4): 245-257.

Verma, M., S. Erwin, V. Abedi, R. Hontecillas, S. Hoops, A. Leber, J. Bassaganya-Riera and S. M. Ciupe (2017). "Modeling the Mechanisms by Which HIV-Associated Immunosuppression Influences HPV Persistence at the Oral Mucosa." PLoS One **12**(1): e0168133.

Viladomiu, M., J. Bassaganya-Riera, N. Tubau-Juni, B. Kronsteiner, A. Leber, C. W. Philipson, V. Zoccoli-Rodriguez and R. Hontecillas (2017). "Cooperation of Gastric Mononuclear Phagocytes with Helicobacter pylori during Colonization." J Immunol **198**(8): 3195-3204.

Vodovotz, Y., A. Xia, E. L. Read, J. Bassaganya-Riera, D. A. Hafler, E. Sontag, J. Wang, J. S. Tsang, J. D. Day, S. Kleinstein, A. J. Butte, M. C. Altman, R. Hammond and S. C. Sealfon (2017). "Solving Immunology?" Trends in immunology **38**(2): 116-127.

Wang, Z., C. M. Birch, J. Sagotsky and T. S. Deisboeck (2009). "Cross-scale, cross-pathway evaluation using an agent-based non-small cell lung cancer model." Bioinformatics **25**(18): 2389-2396.

Windrum, P., G. Fagiolo and A. Moneta (2007). "Empirical validation of agent-based models: Alternatives and prospects." Journal of Artificial Societies and Social Simulation **10**(2): 8.

Wroblewski, L. E. and R. M. Peek (2007). "Orchestration of Dysregulated Epithelial Turnover by a Manipulative Pathogen." Cell Host & Microbe **2**(4): 209-211.

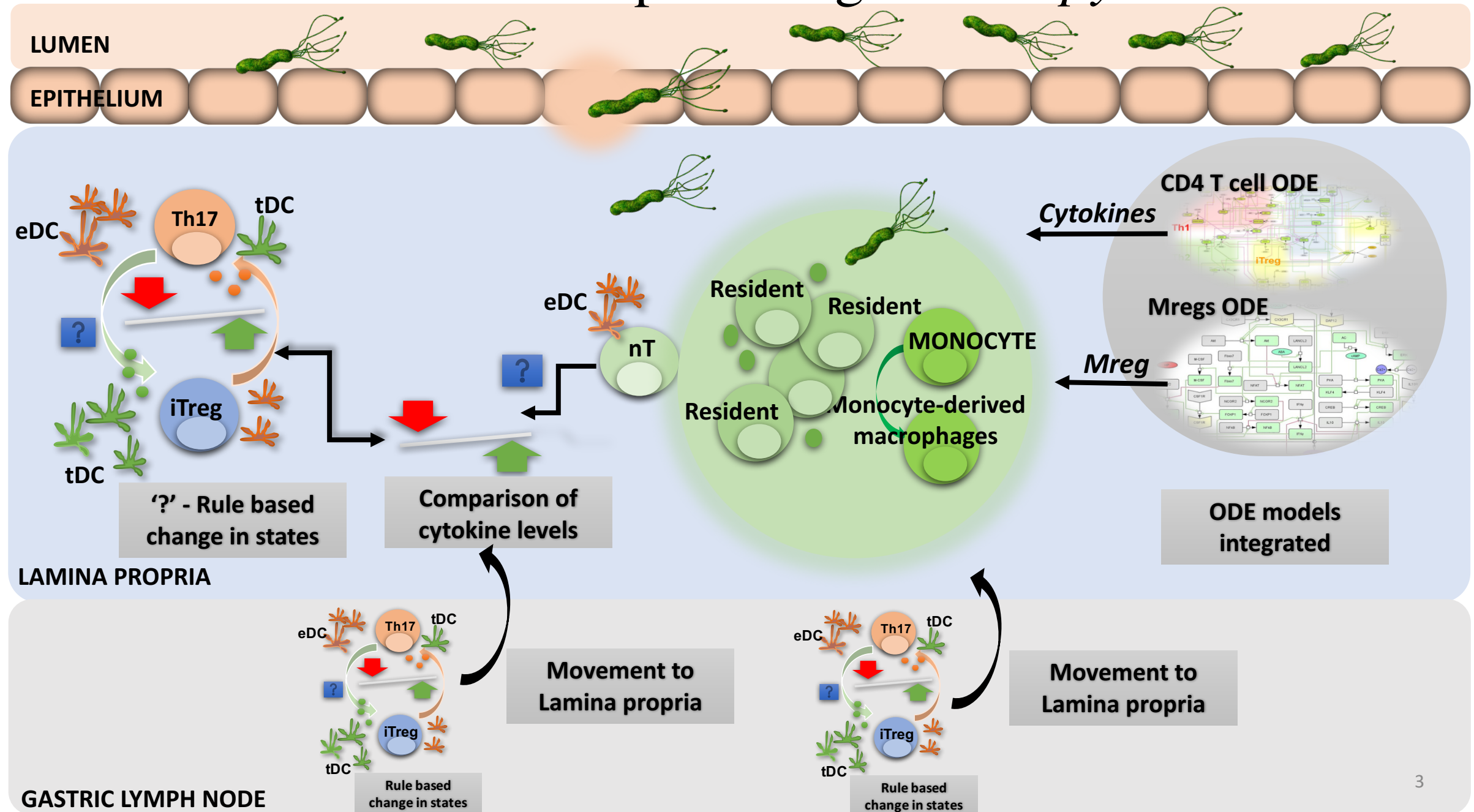
Zhang, M., M. Liu, J. Luther and J. Y. Kao (2010). "Helicobacter pylori directs tolerogenic programming of dendritic cells." Gut Microbes **1**(5): 325-329.



Name of Agent	Agent Type	Rules
<i>Helicobacter pylori</i>	<i>H. pylori</i>	<ul style="list-style-type: none"> <li>- Moves across the epithelial cell border if near damaged epithelial layer</li> <li>- Proliferates in the lumen and lamina propria</li> <li>- Dies (removed from the simulation) in lamina propria and in the lumen due to the damage of epithelial cells by Th1 or Th17 cells</li> </ul>
Macrophages	Monocyte	<ul style="list-style-type: none"> <li>- Proliferates in presence the of effector dendritic cells or damaged epithelial cells</li> <li>- Proliferates in the lamina propria</li> <li>- Differentiates to regulatory macrophage in based on the output from the Macrophage ODE</li> <li>- Differentiates to inflammatory macrophages in presence of IFN-<math>\gamma</math></li> <li>- Dies naturally (removed from the model)</li> </ul>
	Resident	<ul style="list-style-type: none"> <li>- Proliferates in the presence of <i>H. pylori</i></li> <li>- Secretes IL10</li> <li>- Dies naturally</li> <li>- Dies due to Th1 and Tr cells</li> </ul>
	Regulatory	<ul style="list-style-type: none"> <li>- Proliferates and removes bacteria</li> <li>- Dies</li> <li>- Secretes IL10</li> </ul>
	Inflammatory	<ul style="list-style-type: none"> <li>-Proliferates in the presence of damaged epithelial cell</li> <li>-Dies naturally</li> </ul>
Dendritics	Immature	<ul style="list-style-type: none"> <li>-Moves from lamina propria to epithelium compartment and from the epithelium to the lamina propria</li> <li>- Differentiates to tolerogenic dendritic cell in the presence of tolerogenic bacteria, both in epithelium and lamina propria</li> <li>- Differentiates to effector dendritic cell in the presence of <i>H. pylori</i></li> <li>- Proliferates in lamina propria and gastric lymph node</li> <li>- Dies naturally</li> </ul>
	Effector	<ul style="list-style-type: none"> <li>- Moves from lamina propria to gastric lymph node</li> <li>- Moves form epithelium to lamina propria</li> <li>- Secretes IL6 and IL12</li> <li>- Dies naturally</li> </ul>
	Tolerogenic	<ul style="list-style-type: none"> <li>- Moves from lamina propria to gastric lymph node</li> <li>- Moves from epithelium to lamina propria</li> <li>- Secretes TGF-<math>\beta</math></li> <li>- Dies naturally</li> </ul>
T cells	Naïve	<p>In the presence of effector dendritic cells:</p> <ul style="list-style-type: none"> <li>- Differentiates to Th1 in the presence of IFN-<math>\gamma</math> or IL12</li> <li>- Differentiates to Th17 in the presences of IL6 or TGF-<math>\beta</math></li> </ul> <p>In the presence of tolerogenic dendritic cells:</p> <ul style="list-style-type: none"> <li>- Differentiates to iTreg in the presence of TGF-<math>\beta</math></li> </ul>

		<ul style="list-style-type: none"> <li>- Differentiates to Tr in the presences of IL10</li> <li>- Dies naturally</li> </ul>
	Th1	<ul style="list-style-type: none"> <li>- Secretes IFN-<math>\gamma</math></li> <li>- Moves from gastric lymph node to lamina propria</li> <li>- Proliferates in lamina propria and gastric lymph node</li> <li>- Dies naturally</li> </ul>
	Th17	<ul style="list-style-type: none"> <li>- Secretes IL17</li> <li>- <u>i</u>n the presence of tolerogenic dendritic cell, transition to iTreg cells</li> <li>- Moves from gastric lymph node to lamina propria</li> <li>- Proliferates in lamina propria and gastric lymph node</li> <li>- Dies naturally</li> </ul>
	iTreg	<ul style="list-style-type: none"> <li>- Secretes IL10</li> <li>- In the presence of tolerogenic dendritic cell, transition to iTreg cells</li> <li>- Moves from gastric lymph node to lamina propria</li> <li>- Proliferates in lamina propria and gastric lymph node</li> <li>- Dies naturally</li> </ul>
	Tr	<ul style="list-style-type: none"> <li>- Secretes IL10</li> <li>- Dies naturally</li> <li>- Proliferates in the lamina propria</li> </ul>
Epithelial	Healthy	<ul style="list-style-type: none"> <li>- Damaged due to infectious bacteria</li> <li>- Damaged due to Th1 and Th17 cells</li> <li>- Proliferates</li> <li>- Secretes IL6 and IL12</li> <li>- Dies naturally</li> </ul>
	Damaged	<ul style="list-style-type: none"> <li>- <u>Transitions</u> to healthy state in the presence of IL10</li> <li>- Dies naturally</li> </ul>
Bacteria	Infectious	<ul style="list-style-type: none"> <li>- Dies due to Th1 or Th17 or inflammatory macrophages or damaged epithelial cells</li> <li>- Dies naturally</li> <li>- Proliferates in the lamina propria</li> </ul>
	Tolerogenic	<ul style="list-style-type: none"> <li>- Moves from lumen to the epithelium in the presence of damaged epithelial cells</li> <li>- Becomes infectious if moves in the lamina propria compartment</li> <li>- Proliferates in lumen and lamina propria</li> <li>- Dies naturally</li> </ul>

# Immune responses against *H. pylori*



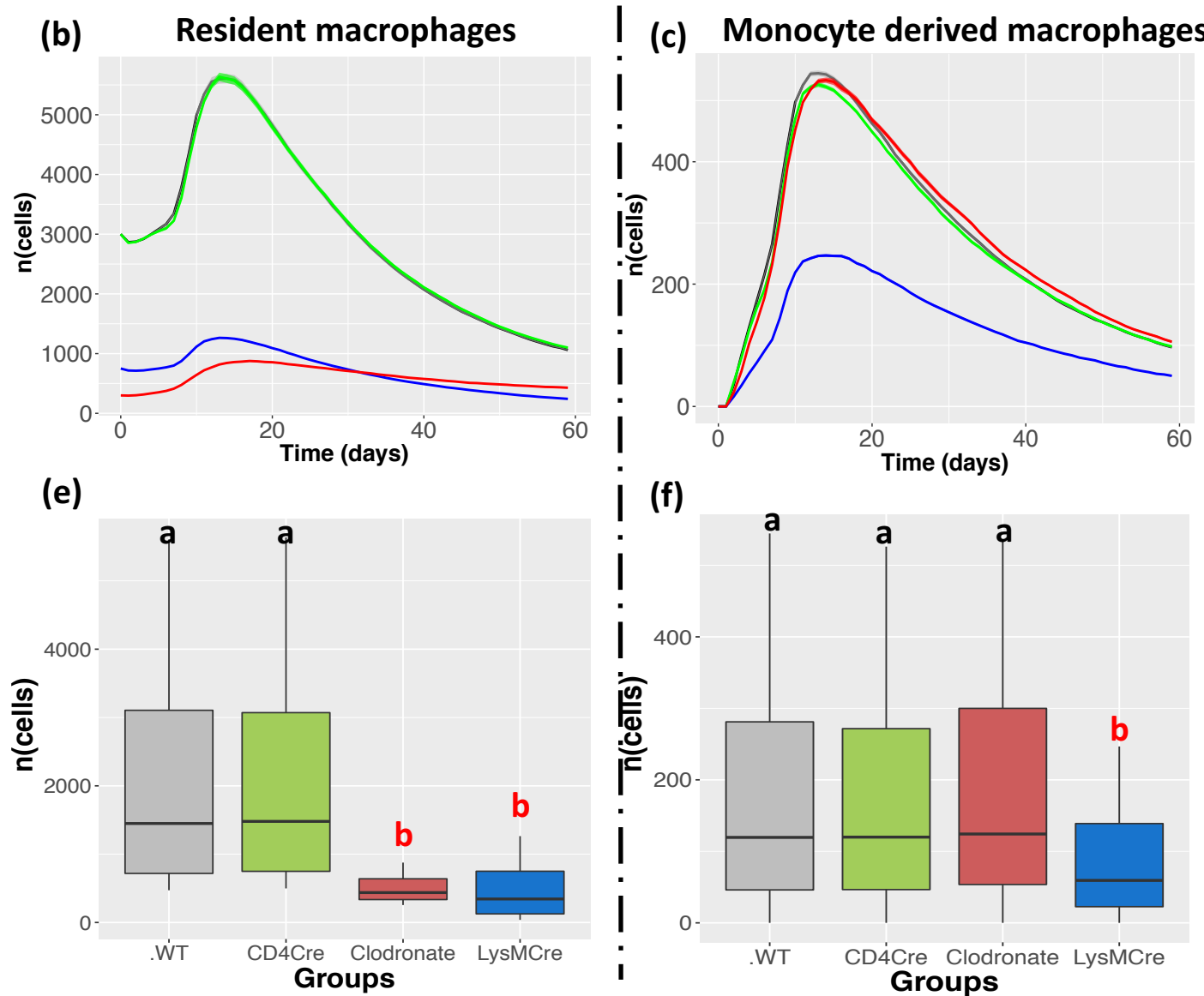
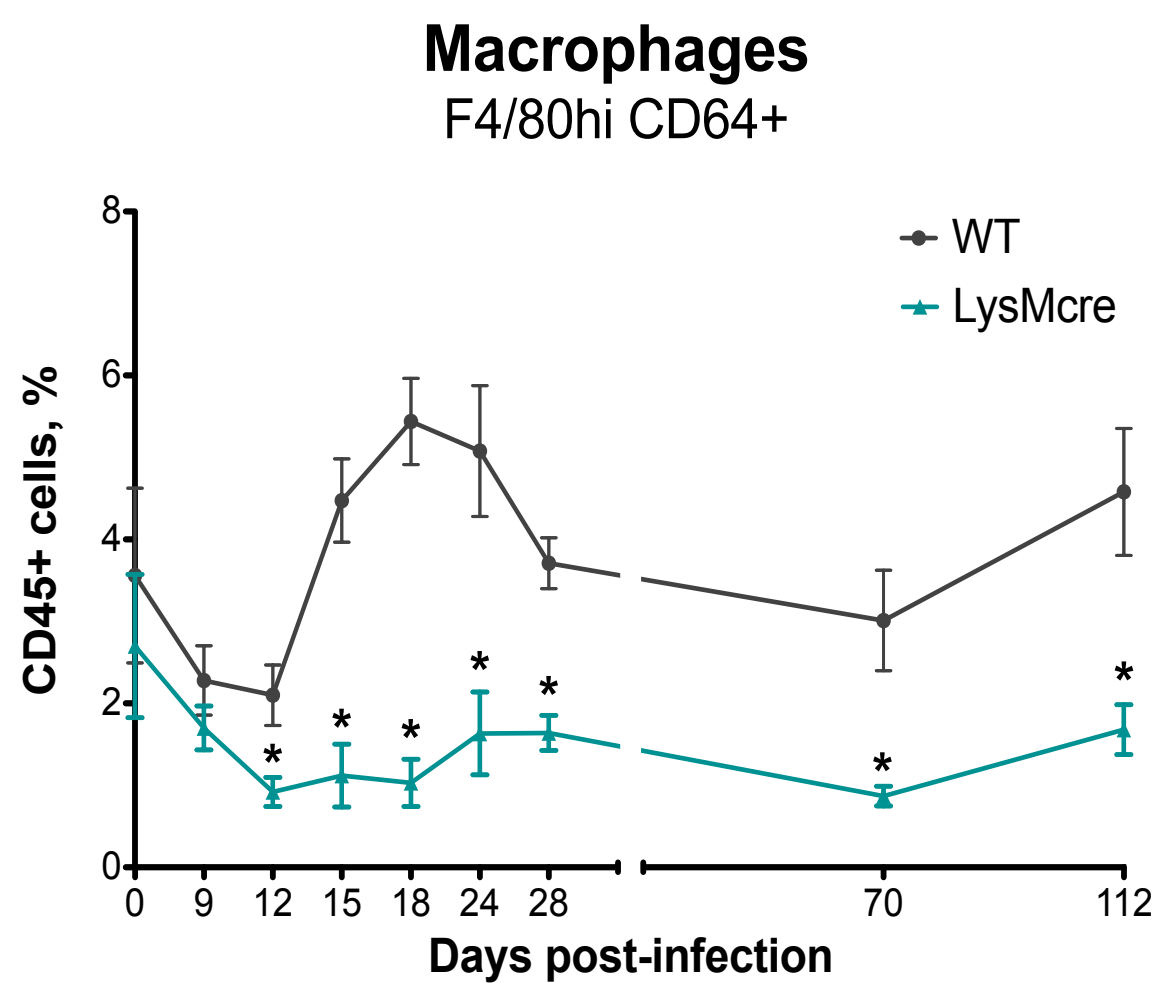
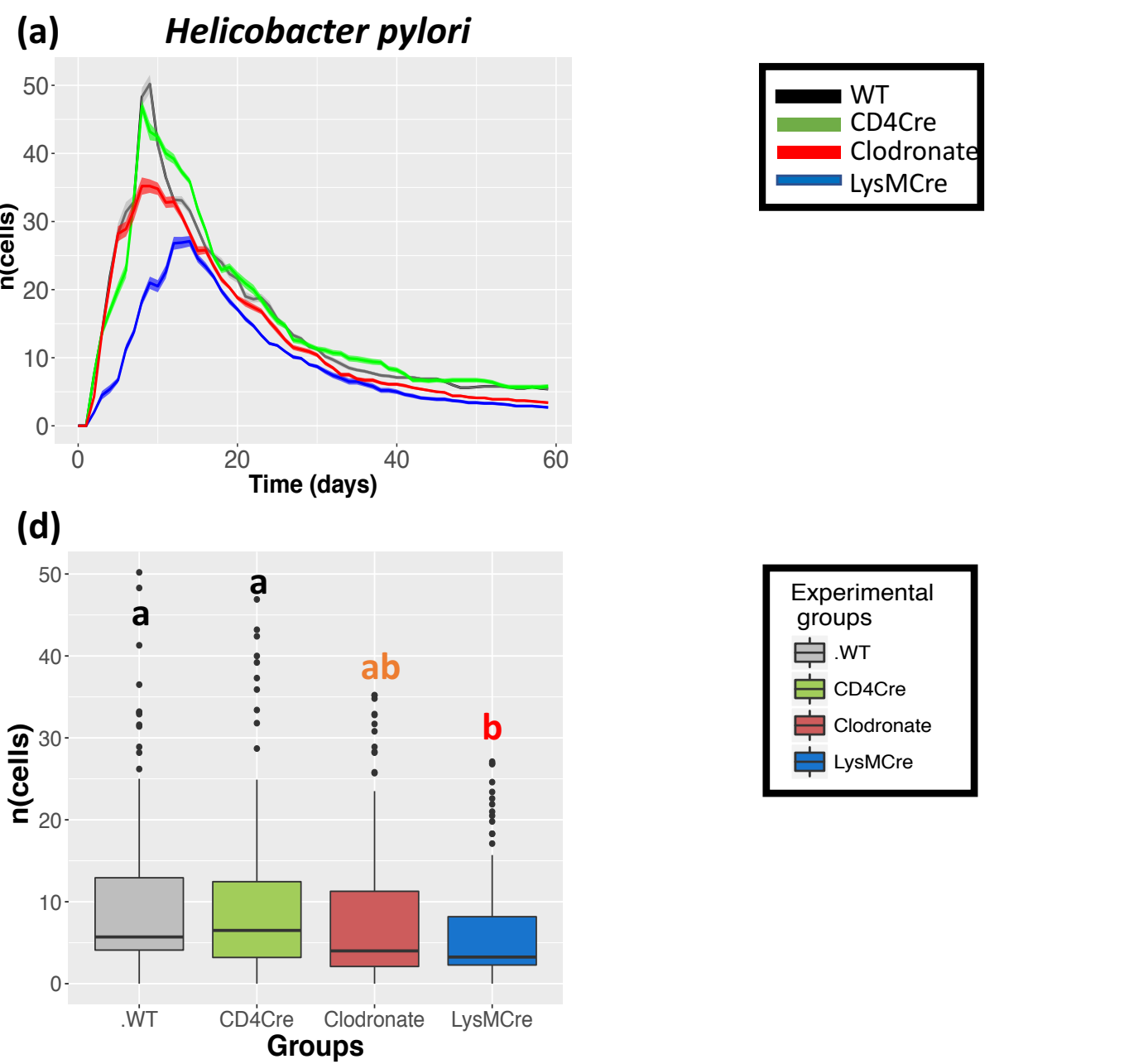
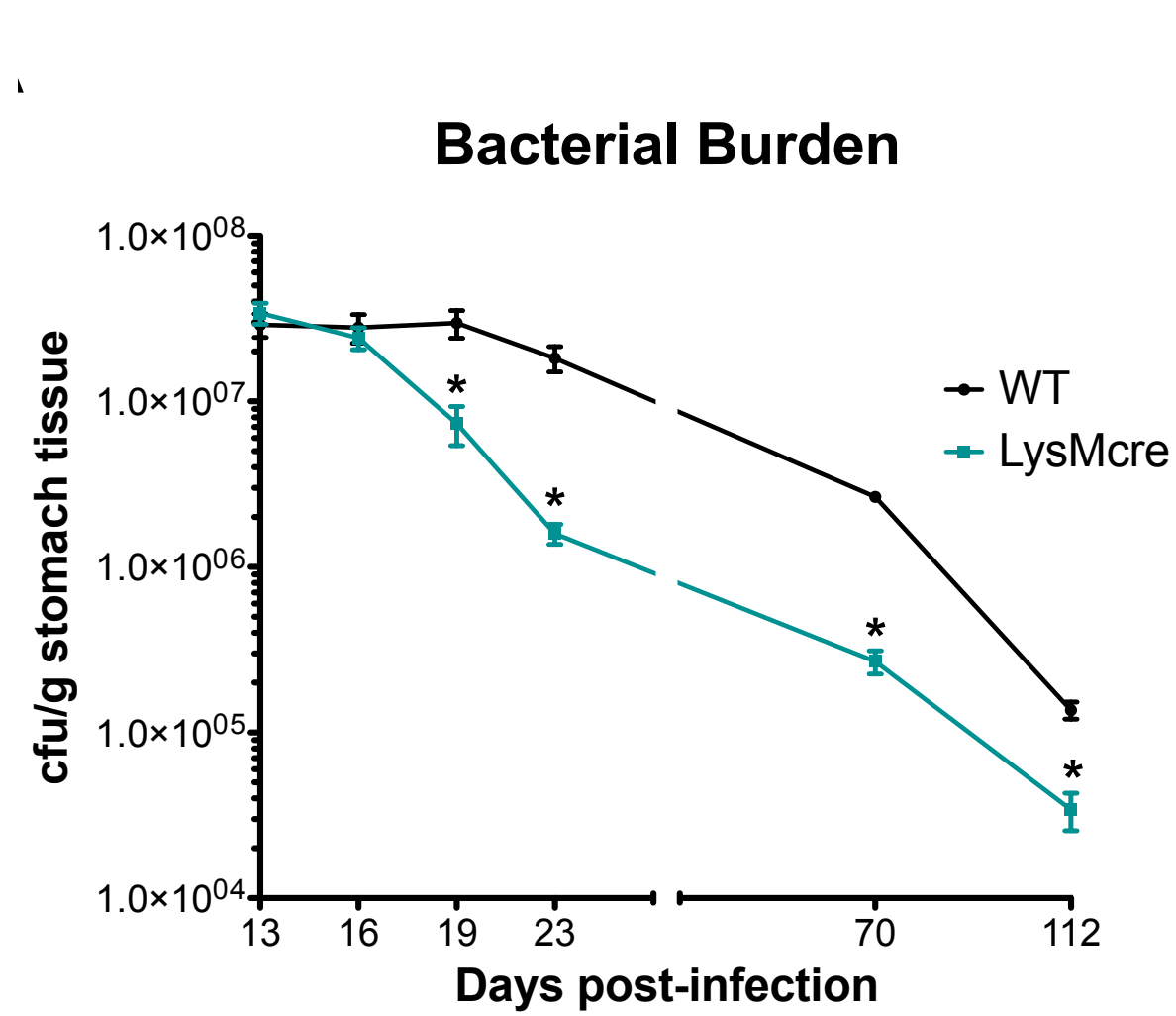
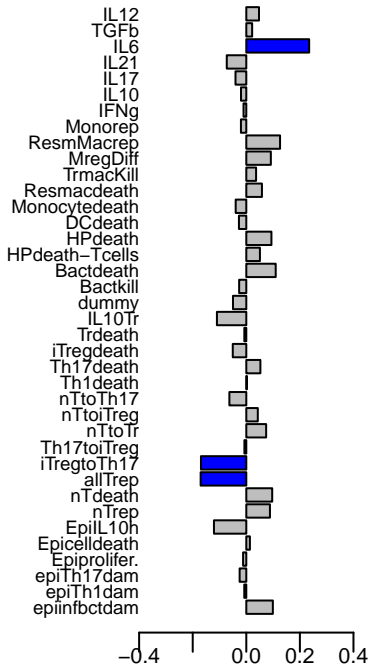


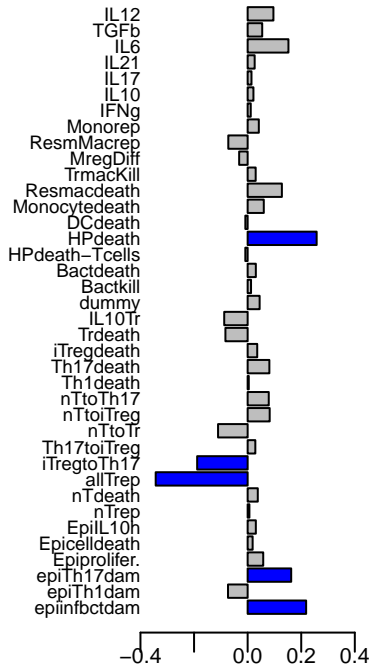
Figure 3

**Helicobacter pylori in Lamina propria**

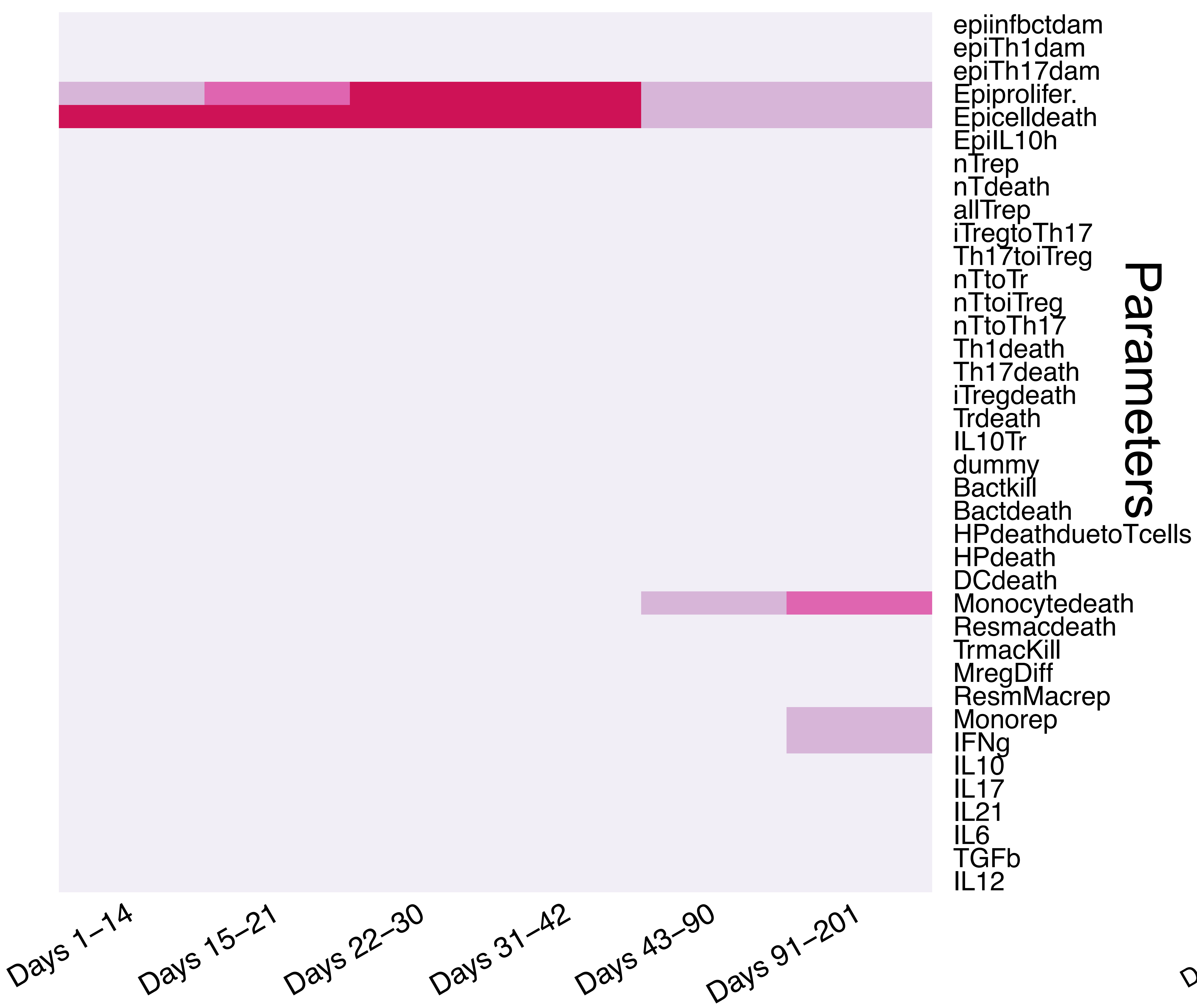


[Click here to access/download; Figures; Figure3.pdf](#)

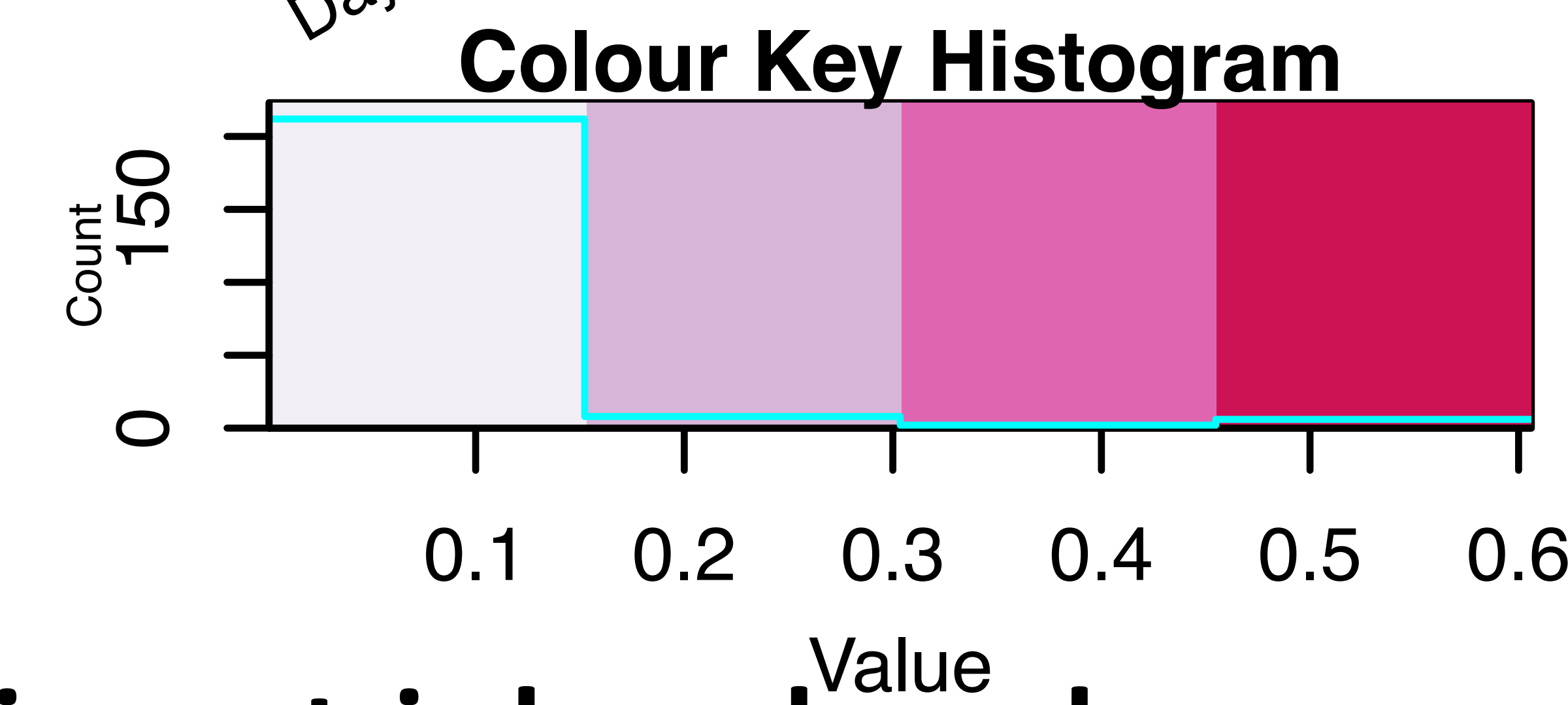
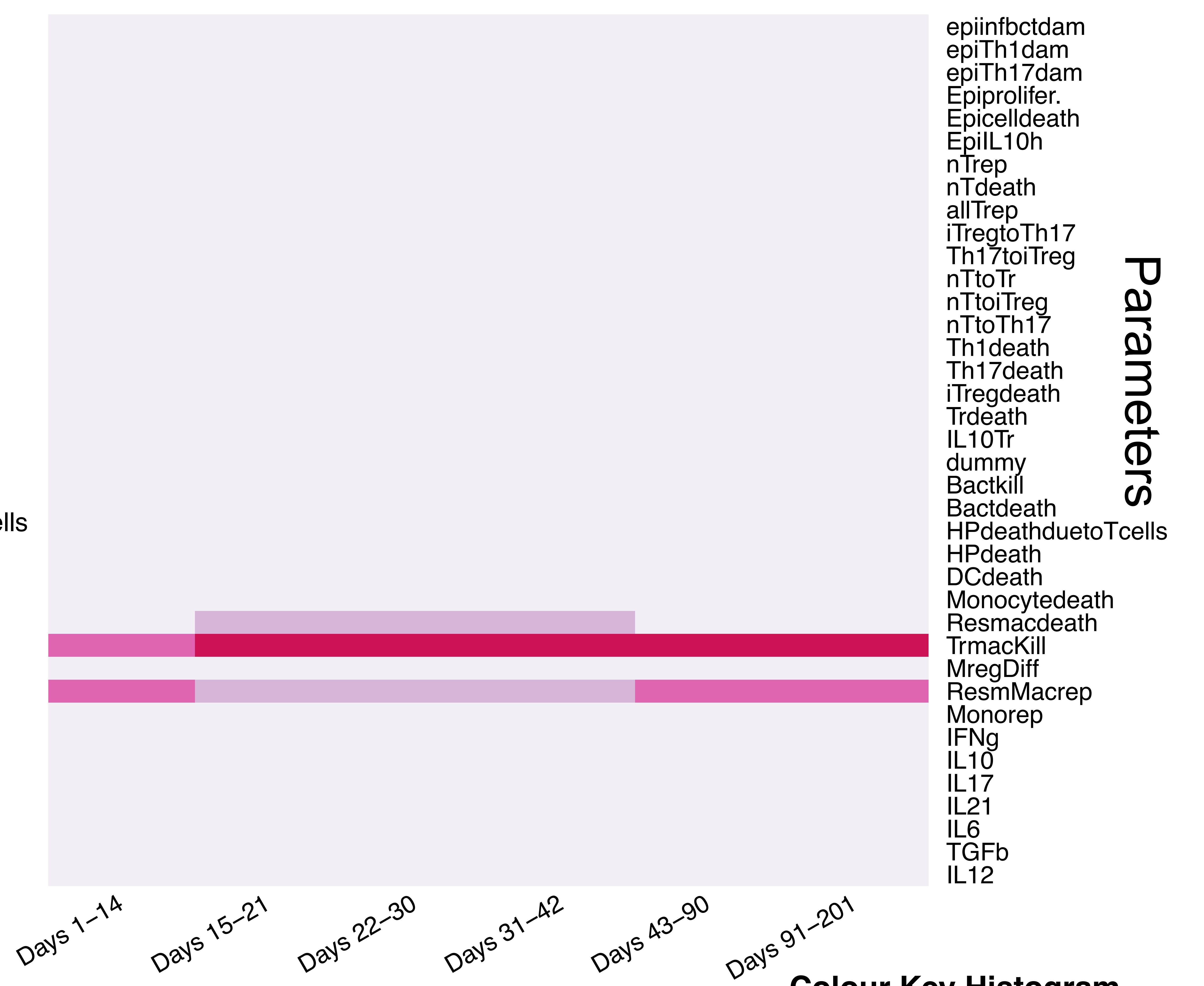
**Resident macrophages in Lamina propria**



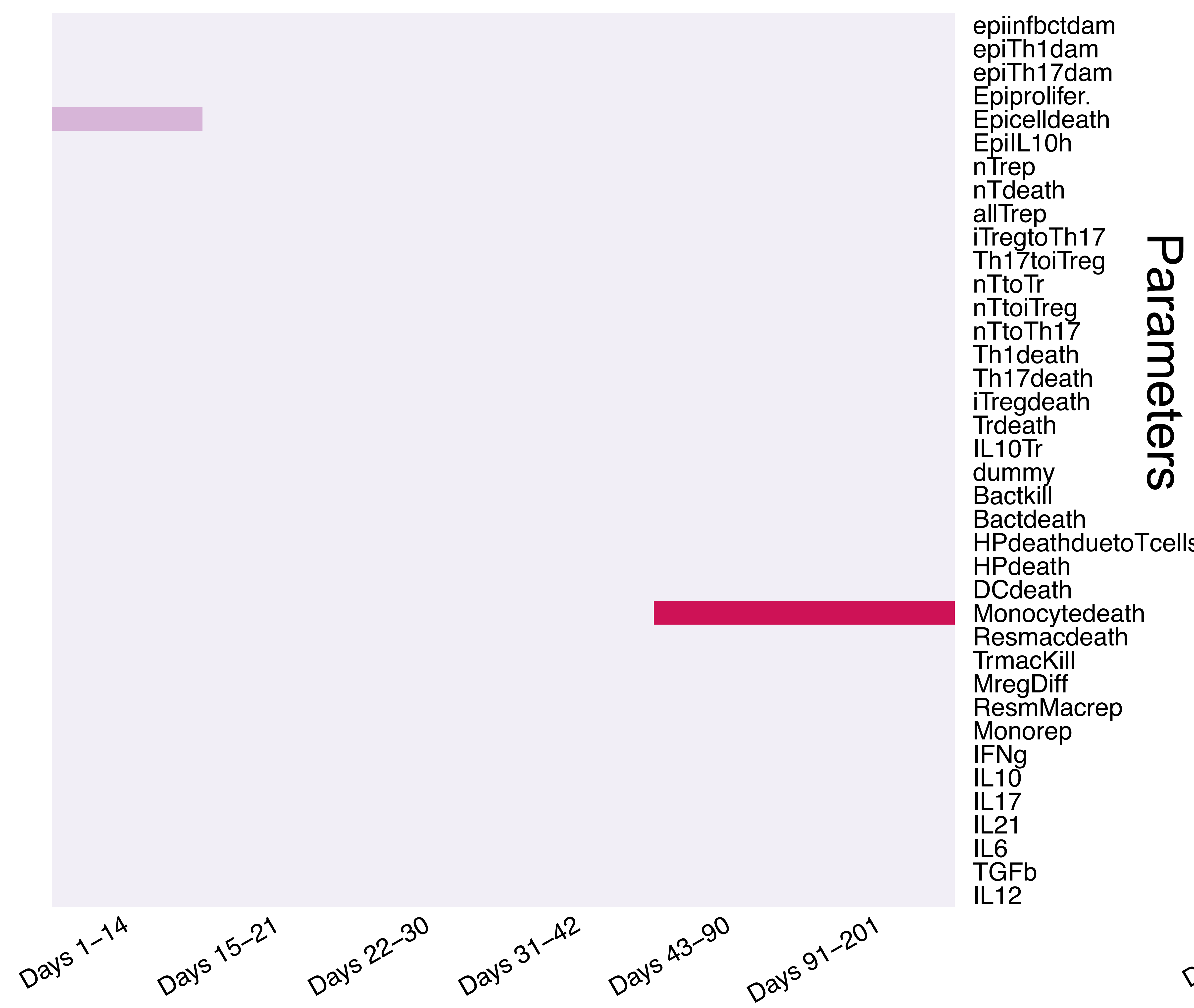
**(a) *Helicobacter pylori* in lamina propria**



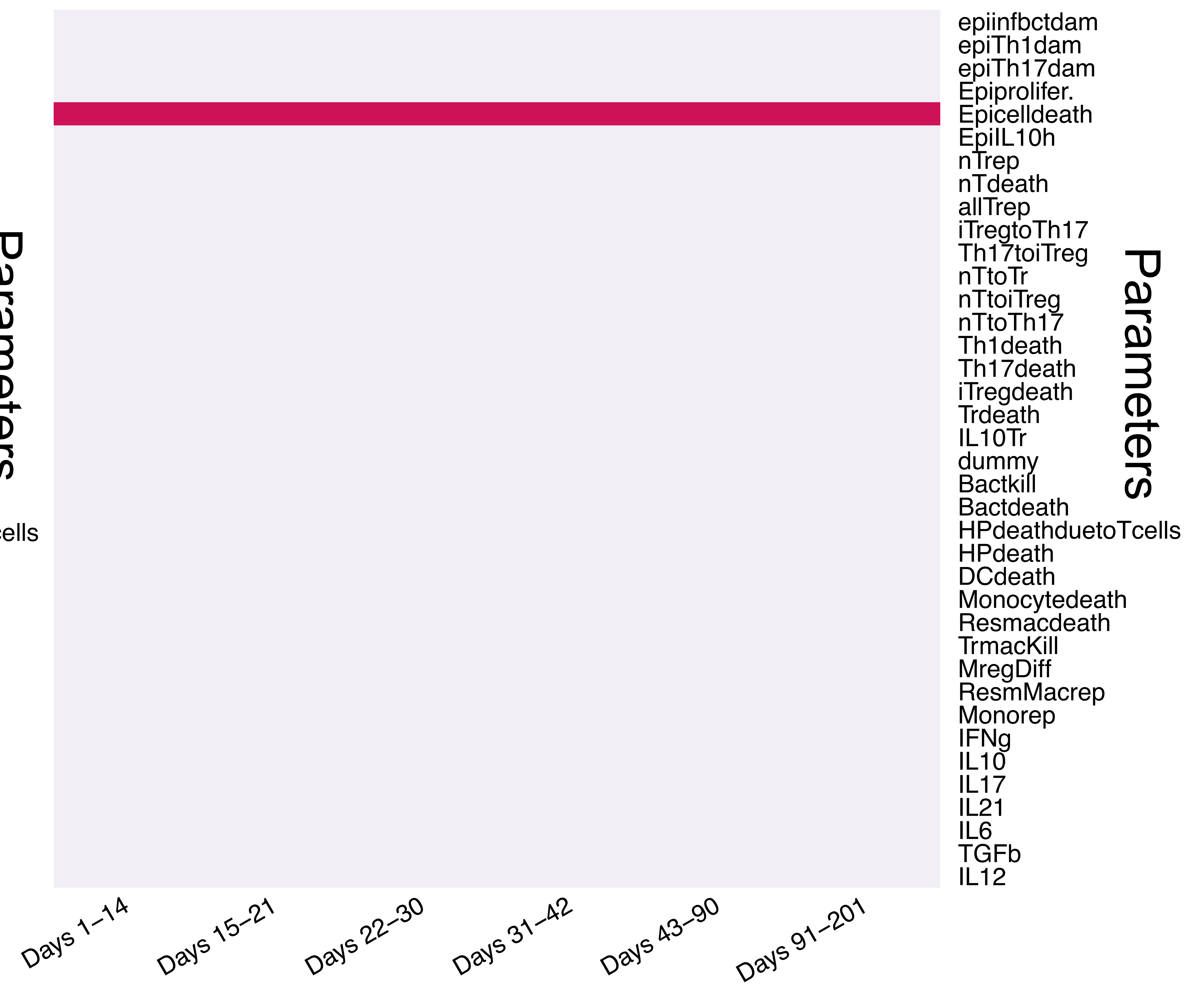
**(b) Resident macrophages in lamina propria**

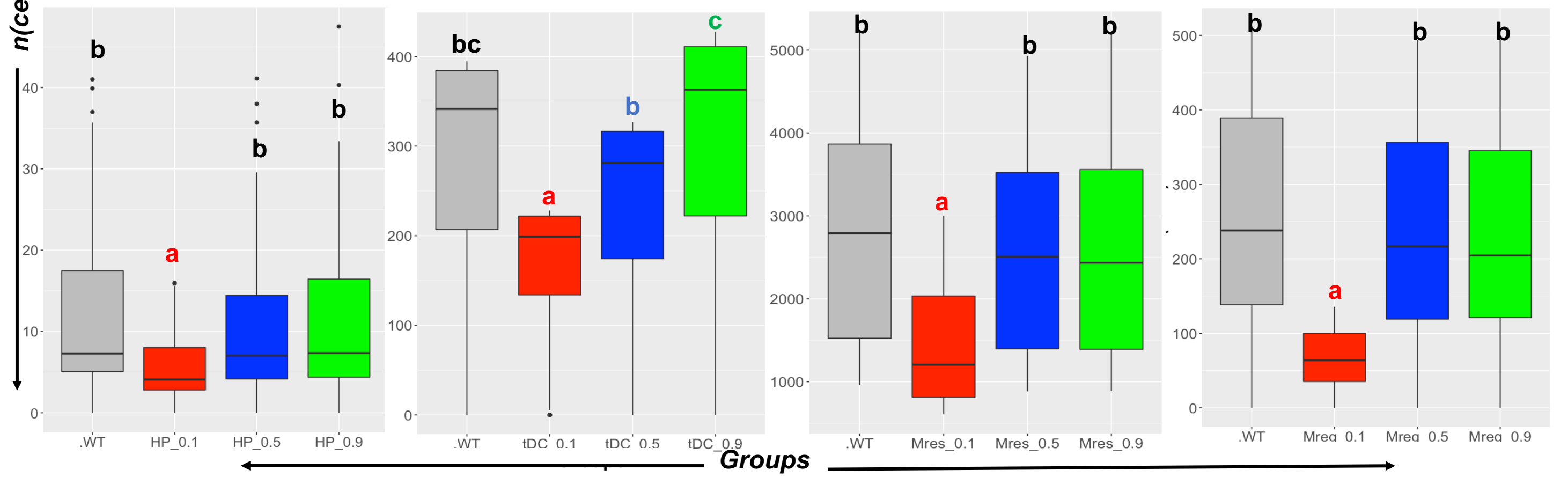
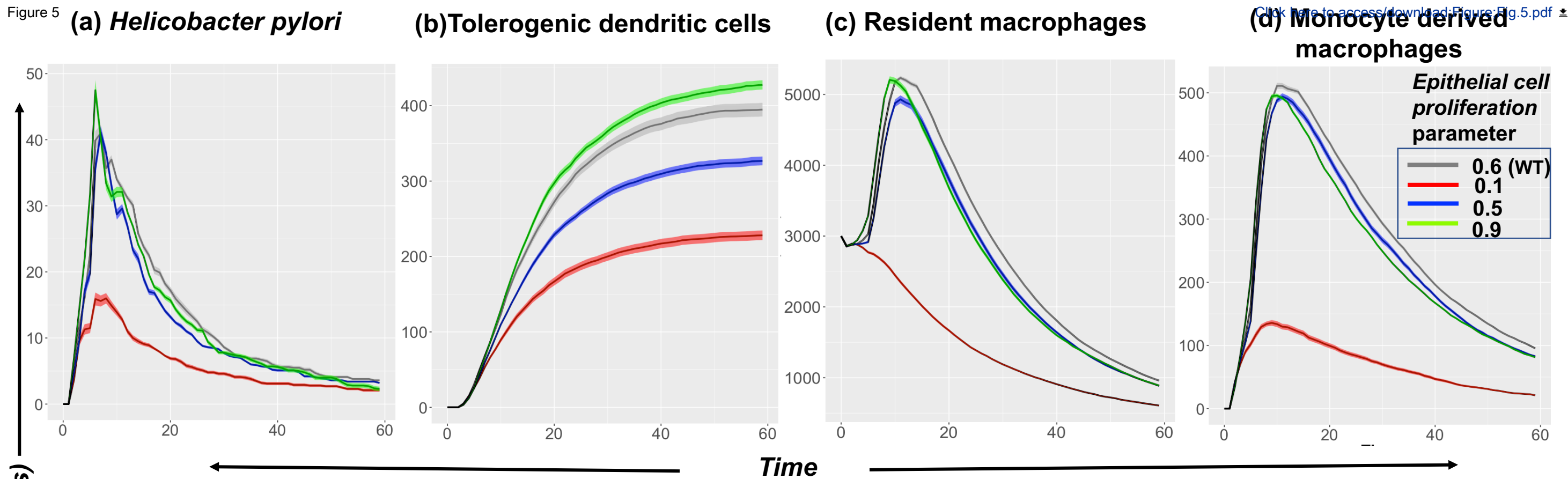



**(c) Monocyte derived macrophages in lamina propria**



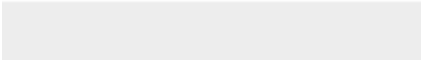

**(d) Tolerogenic dendritic cells in gastric lymph node**



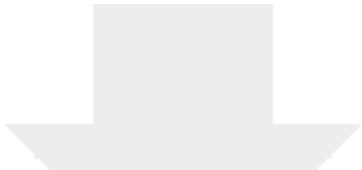





Click here to access/download  
**Supplementary Material**  
FileS1.pdf









Click here to access/download  
**Supplementary Material**  
FigS1.png



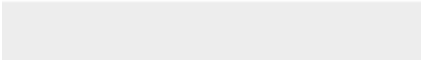



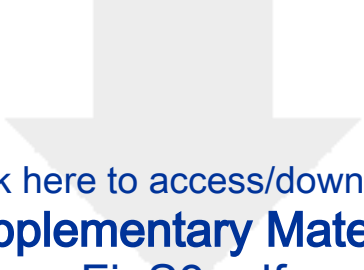
Click here to access/download  
**Supplementary Material**  
Table\_S1\_revised.docx



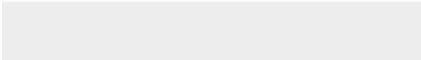



Click here to access/download  
**Supplementary Material**  
FigS2.pdf





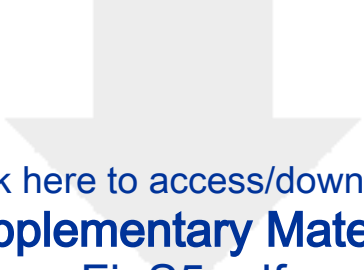
Click here to access/download  
**Supplementary Material**  
FigS3.pdf



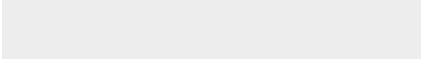



Click here to access/download  
**Supplementary Material**  
FigS4.docx





Click here to access/download  
**Supplementary Material**  
FigS5.pdf



### Point by point response to the Reviewer reports

We would like to thank the reviewers and editors involved for dedicating valuable time to review our manuscript entitled “*High-Resolution Computational Modeling of Immune Responses in the Gut*”. We are extremely grateful to both reviewers and the editors for their time and attention to our submission. The review was very helpful to us, and the recommended changes have allowed us to provide better documentation and clarity to the present work. The review process has been helpful in the improvement of our submission. We have considered the comments that were made and have prepared the following point-by-point response.

**Reviewer #1:** In this manuscript, the authors constructed a multi-scale systems biology model of *Helicobacter pylori* infection to study the interaction between bacterial infection and the immune system. Some modifications could be considered to improve the quality of this manuscript:

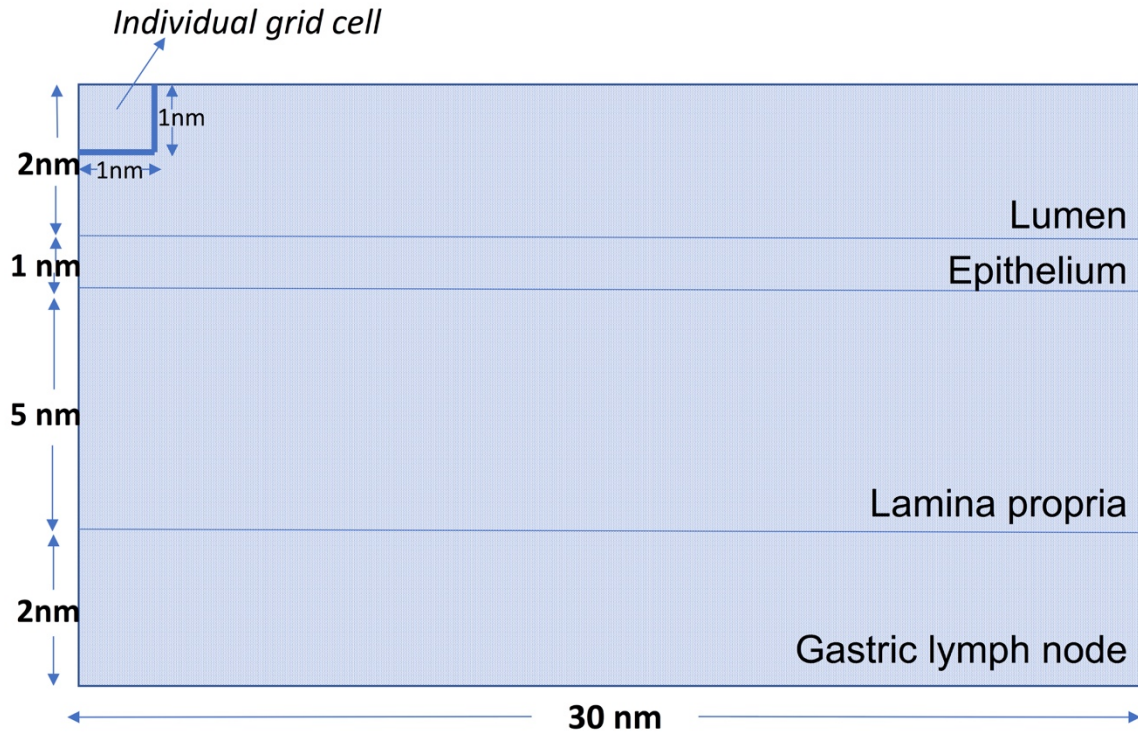
1. The model needs to be more clearly described in the text. Some details might be available from the code; nevertheless, it would be helpful for readers to understand if the authors can include more information regarding the model. For example:

*We thank the reviewer for their valuable suggestion. We agree with the reviewer’s comment and accordingly we updated the manuscript with the response described below in detail.*

a. Agent-based model:

i. What is the spatial discretization? The authors mentioned it's a 30\*10 2D grid cell, but resident macrophages are in thousands. So multiple cells are allowed in the same grid location? How many?

**Response:** *i) The model has a spatial discretization such that the dimension of the entire (two-dimensional) grid is 30nm x 10 nm). An individual grid cell is 1nm x 1nm, however, this is a configurable run\_parameter and can be changed without modifying the model. An individual grid cell 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 resulting time series of cytokine concentrations will be used to update the cytokine value in the ABM/PDE system and COPASI would simulate a different model for each of the relevant cell type within that individual grid cell. Below is a figure describing the grid, also added in the Additional file Fig S2.*



The entire grid is divided within into 4 functionally and anatomically distinct sized compartments such that the dimensions of the 4 compartments are lumen (2nm), epithelium (1nm), lamina propria (5nm) and gastric lymph node (2nm).

The following compartments are adjacent to each other:

- Lumen - epithelium
- Epithelium - lamina propria
- Lamina propria – 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 agents were randomly placed separated by the four compartments 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 that there is no limit of agents within each individual spatial grid cell. 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. We demonstrate below how we obtain a count of thousands of 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 (1) described below -

$$n(\text{resident macrophages}) = \text{size}_{\text{compartment}}(\text{lamina propria}) \times \text{concentration}_{\text{initial}} (\text{resident macrophages}) \quad (1).$$

$$n(\text{resident macrophages}) = (30 \times 5) \times 30 = 4500.$$



The manuscript has been updated with the above addressed points, please refer to Line129 - Line147, and L217-245.

ii. What is the time step size?

**Response:** The time step size is 1 tick ~ 1 day which was obtained during the process of fitting the output to the results from the mouse model of *H. pylori* infection. For e.g. the peak of resident macrophages in lamina propria (refer Fig 2b,d) is observed at ~21 days which is similar to the results obtained in Fig 2A described in ((Viladomiu, Bassaganya-Riera et al. 2017) (also described in detail in point by point response 2.b).

The manuscript has been updated with the above addressed points, please refer to Line247 - Line253.

iii. How is migration implemented for cells and bacteria agents?

**Response:** 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 (<https://github.com/NIMML/ENISI-MSM/src/compartment/Compartment.cpp>) file.

The manuscript has been updated with the above addressed points, please refer to Line294 - Line307.

b. ODE: What's the COPASI setup for the solver? How is the solver in sync with the ABM?

**Response:** The COPASI setup for the solver uses 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 manuscript has been updated with the above addressed points, please refer to Line266 - Line274.

c. PDE: What package and numerical scheme is used to solve the PDEs? What's the setting?

**Response:** ENISI MSM is a multiscale agent-based modeling platform for computational immunology which was building on our previous works, ENISI MSM that integrated COPASI, the ODE solver, ENISI, an agent based simulator (Mei, Abedi et al. 2015).

The ENISI MSM PDE solver uses a simple numerical scheme to solve the PDEs (<https://github.com/NIMML/ENISI-MSM/tree/master/src/diffuser>) and process distributed value

layer (<https://github.com/NIMML/ENISI-MSM/blob/master/src/grid/ValueLayer.h>). The ValueLayer stores the value for a grid space and provides methods to change the values of individual grid cells. The Diffuser is used to diffuse the values of the ValueLayer using diffusion ( $d$ ) and degradation ( $\delta$ ) constants as described in (Mei, Abedi et al. 2015). The diffusion constant determines the migration of values of a grid cell to its neighboring grid cells. As implemented in ValueLayer library, the diffusion of cytokines follows the equation shown below also described in Mei et al, 2015. Here,  $v_n$  is the value of the grid cell itself at step  $n$ . The values of  $c_{\delta}$  and  $c_d$  are degradation and diffusion constant respectively.

$$v_n = v_{n-1} + c_{\delta} * [\sum (c_d^{neighbor} * v_{n-1}^{neighbor}) - 6.0 * v_{n-1}]$$

0.3	1.2	0.3
1.2	-6.0	1.2
0.3	1.2	0.3

The PDE solver uses the above number scheme  $c_d^{neighbor}$  for the diffusion process. The step size  $c_{\delta}$  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 the identical with the spatial discretization for the agents.

We updated the manuscript details to solve PDEs and the setting, please refer to L275-L293.

2. The authors listed the values of parameters in Table S1.

a. However, it's not clear what their units are (the baseline column seems to include characters such as "l<sup>2</sup>", "#", or "d". are these units? Please clarify).

**Response:** We thank the reviewer for pointing this out. Those were the units for the parameters in the COPASI ODE models. The parameters described in Table S1 are probability values (ranging from 0 to 1) and hence do not have any units. The characters have been removed and the column 3 of Table\_S1 has been updated.

b. Also, the sources of the parameter values are not very clear, except for the vague statement "expert judgement" (Saltelli, Tarantola et al. 2000 is cited, but this is an article on SA and does not contain parameters).

**Response:** The values of the parameters for the model presented here are obtained via 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 (Viladomiu, Bassaganya-Riera et al. 2017) published by NIMML (described here below in the last paragraph detail).

We want to clarify the practice of using expert opinion is known in the SA field and hence we cited Saltelli, Tarantola et al. 2000 as it supports the statement. As discussed in (Thorne, Bailey et al. 2007), 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. 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 experimental 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 Fig 1A of Viladomiu, Bassaganya-Riera et al. 2017. 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 (Fig 2a,d) 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 Fig. 2A, 2E of Viladomiu, Bassaganya-Riera et al. 2017. 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 (Fig 2b,e and Fig 2c,f) 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 updated the manuscript accordingly, please refer to L340-L350 and L654-671.

c. Please in the table explain what mechanism each parameter corresponds to. Some can be inferred from the name, but it's not very clear.

**Response:** We thank the reviewer for this valuable suggestion. We added column 2 in Table S1 that describes the detailed mechanism that each parameter corresponds to. We updated the manuscript accordingly, please refer to L236-L238.

For e.g. 3 rows of the Table S1 are shown below –

**Table S1**

Name of parameters	Description
p_epiinfbactdamage	Epithelial cell damage due to infectious bacteria
p_epith1damage	Epithelial cell damage due to Th1 cells
p_epith17damage	Epithelial cell damage due to Th17 cells

d. Some parameters are not included in the table. For example, the diffusivity of the cytokines are not listed.

We thank the reviewer for this valuable observation. We listed the diffusivity of the cytokines and updated the Table\_S1.

3. In Table 1 and Table 2, there is a T cell class named "Tr", which is not explained in the text. Please clarify.

**Response:** We thank the reviewer for pointing this out. The Tr cells are the type 1 regulatory (Tr1) T cells that are regulatory subset of T cells, whose expansion is dependent on environmental IL-10 (produced by Mreg). These are different than iTreg which are T cells differentiated from naïve

*T cell in presence of tolerogenic dendritic cells and TGF- $\beta$  cytokine. We clarified this point and updated the manuscript, please refer L208-211.*

4. The authors used a Gaussian emulator as surrogate model for the hybrid model. In line 582, the authors mentioned that performance is evaluated using diagnostic plots in Figure S4. Please clarify what the "Observed" data refers to. Are these the same simulations from the training set which the emulator fitted to, or are these new simulations done? If these are the training set results, the authors need to run simulations and emulation on a new testing set and evaluate the performance; if it's already done, please clarify how its done (range of parameters, number of simulations, etc.)

**Response:** *We thank the reviewer for your careful reading and bringing up the issues in the description of the original plot. Below please find our response to your comments.*

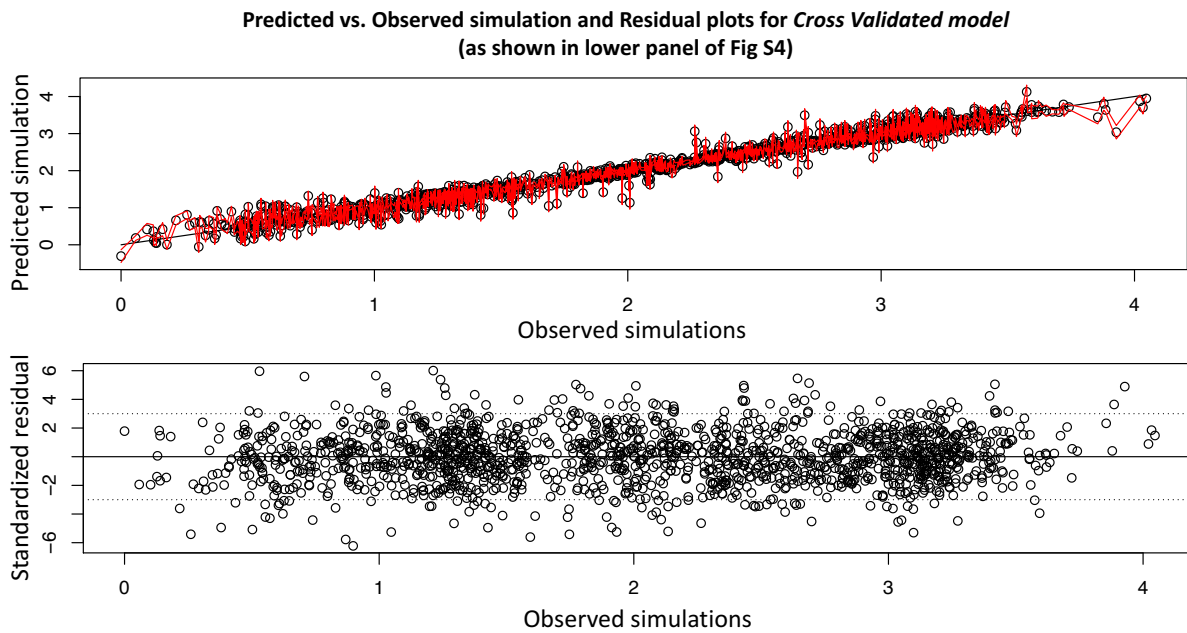
*First, the "observed" data, i.e. the 'x' axis in the first half of lower panel in Figure S4 (shown here below as Fig 1a) (please note in the revised manuscript the Figure S4 is now updated and referred to as by Fig S5.), refers to the observed output values of the simulations obtained after running the hybrid computer model, whereas the 'y' axis refers to the predicted values obtained from the cross validated model. These 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 second half of lower panel refers to the standard residual plot wherein the 'x' axis represents the observed values obtained from the simulation and the 'y' axis refers to the residual error ({error (predicted values – observed values) / standard deviation (error)}) obtained.*

*In fact, the models used for plotting are the cross-validated ones and are not fit using the entire dataset. Cross-Validation (section 7.10 of The Elements of Statistical Learning (Trevor, Robert et al. 2009), is a legitimate approach for model assessment and it is especially suitable in our case because the simulation data is expensive to obtain (each simulation takes ~9-10 minutes to run, thus 267 parameter sets with 20 replicates = 5,340 simulations. The entire simulation dataset took us about 2 months to obtain.*

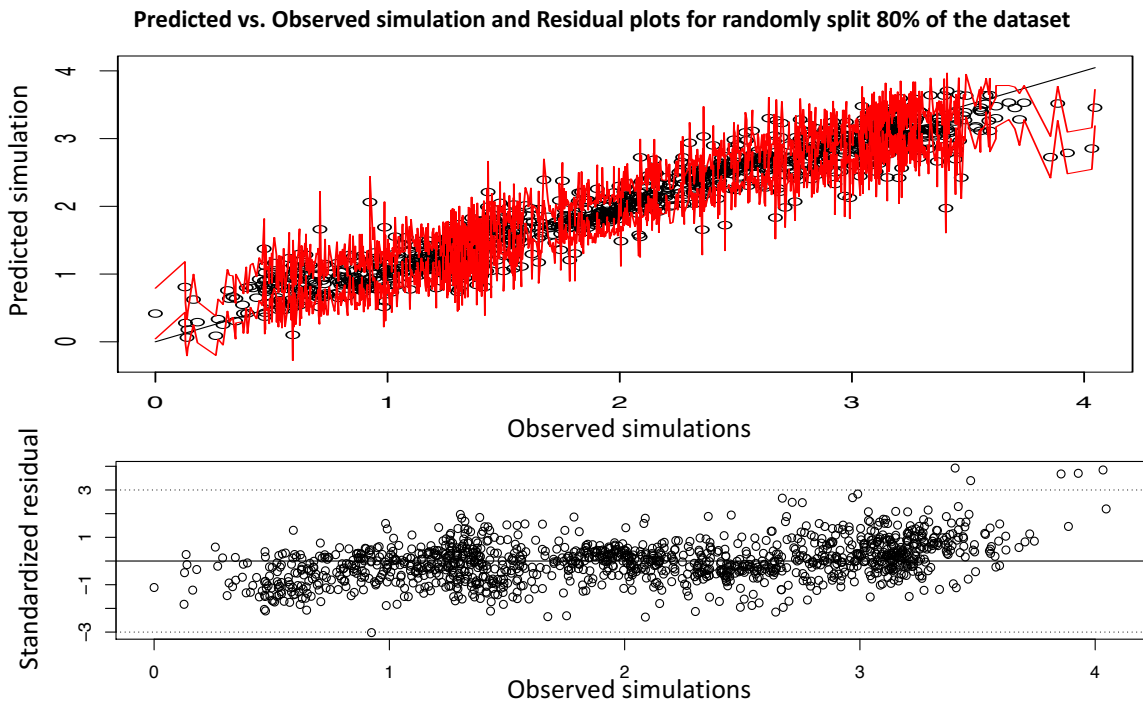
*Nevertheless, we would like to show that using separate testing and training dataset for model assessment we obtained similar conclusions as those using the cross validated model.*

*We randomly split the observed output simulation dataset for one of the datasets (Fig S4 "a"), Helicobacter pylori in Lamina propria into training (80%) and testing (20%) sets and built the Gaussian emulator using the mlegp package. As observed in the Fig 1b and Fig 1c below, we plotted the predicted (values predicted using mlegp) vs. the observed simulation data values for both the training set (top panel of Fig 1b) and testing sets (top panel of Fig 1c). In the top panels of Fig 1b, the black circles denote the cross validated prediction points for the training dataset. Similarly, the top panel of Fig 1c, the black open circles are obtained after plotting the predictions for testing dataset, made using the model trained on the 80% of the randomly split dataset, vs. the observed values (known) for the 20% of the randomly split used as testing dataset here. Additionally, we calculated the standardized residuals for each of the 80% and 20% randomly split datasets and plotted the standardized residual plots in the lower panels of Fig. 1b and Fig 1c respectively.*

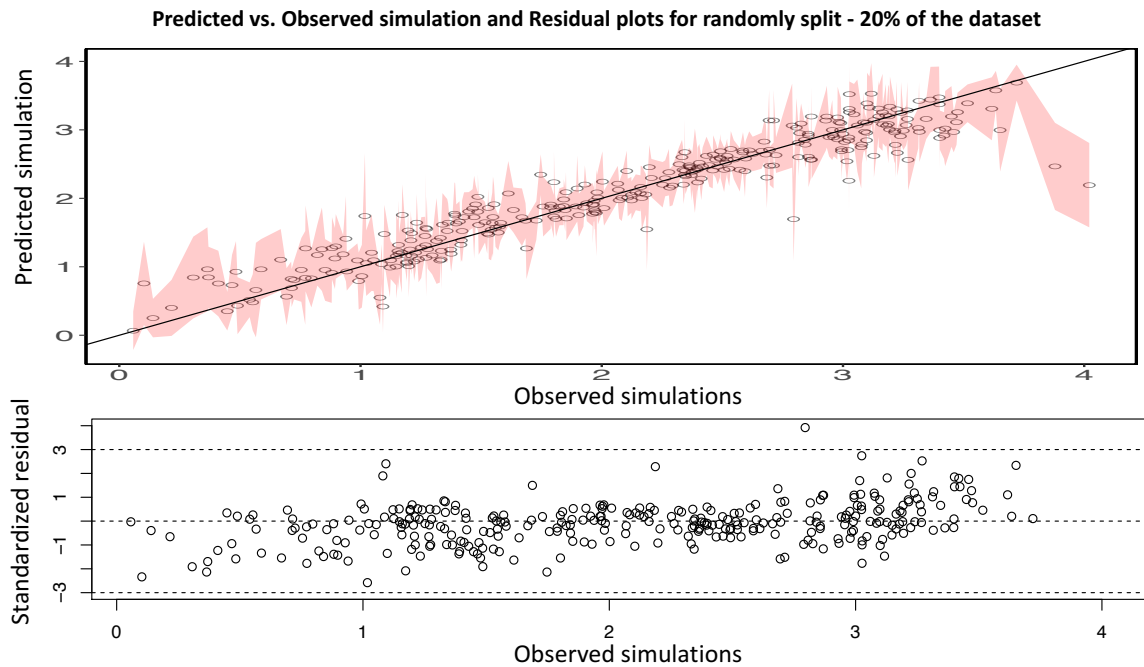
*As observed in the bottom panels of the Fig 1a, 1b and 1c the amount of standard residuals obtained for the cross validated model (Fig S4 a) from the paper and also the one mentioned in previous paragraph), the training dataset (80% randomly split dataset), and testing dataset (20% randomly split dataset) respectively, were similar. Thus, here we demonstrated that the results obtained from the cross-validated model built using mlegp (from Figure S4 a) and as shown here in Fig 1a) were similar to the results obtained using the cross-validation technique by randomly splitting the data into 80% and 20% (shown here in Fig 1b and Fig 1c).*



*Fig 1a. Original plot from Fig S4 a). The plot shows the predicted vs. observed simulation values for the Cross Validated (CV) model (top panel) and residual error plot for the CV model (bottom panel).*



*Fig 1b. The plot shows the predicted vs. observed simulation values for the randomly split 80% of the dataset (top panel) and residual error plot for the randomly split 80% of the dataset (bottom panel).*



*Fig 1b. The plot shows the predicted vs. observed simulation values for the randomly split 20% of the dataset (top panel) and residual error plot for the randomly split 20% of the dataset (bottom panel).*

*We clarified that the observed data refers to observed simulation values and recreated the Figure S4 (now updated to Fig S5 with updated legends. Please refer to L1045-L1058 in the manuscript.*

**Reviewer #2:** *The authors present results from a multi-scale hybrid model of host immune responses to *H pylori* exposure in the gut. The paper addresses outstanding questions in this complex system and overall the results are interesting. Some comments/questions to be addressed are outlined below.*

1. A key component of the introduction ("double edge sword, p 1 line 4") as well as in the discussion (p28 line 672 "dual role as pathogen and beneficial organism") mentions the conflicting roles of H pylori infection - however the results do not clearly connect to help answer this dichotomy. More detailed analysis/discussion of the results should be provided to clarify the conclusion or the focus of the intro/discussion should be adjusted to relate more closely to the results currently presented.

***Response:*** We thank the reviewer for this valuable suggestion. This study addresses the dichotomy in the introduction but mostly focuses on investigating the dynamics that promote the tolerance to the bacterium in the gastrointestinal mucosa and its systemic immunoregulatory effects. We view the dichotomy represented by the beneficial effects of regulatory responses (immune tolerance to the bacterium) in lesion development versus the detrimental actions of effector responses. Since, the majority of H. pylori-colonized individuals, approximately 85%, do not present any detrimental effect, we wanted to contribute towards the further investigation of the dynamics of immunoregulatory mechanisms underlying H. pylori infection using computational modeling. We emphasized the need for investigation of the immunoregulatory role and the adjusted the focus of the introduction and discussion to relate more closely to the results highlighting regulatory immune cells here. We updated the manuscript accordingly, please refer to L4-L7, L52-L54, L62-L65 in introduction and L850-L854 in discussion.

2. Section 3.4 and p 29 line 694 discuss the involvement of regulatory macrophages and tolerogenic DCs on the colonization of H pylori. These conclusions appear to be drawn based on correlation between responses in H pylori and macrophage/DC populations upon epithelial cell proliferation adjustment (Fig 5). A causal connection between the macrophages/DCs and H pylori is not made (or is not clear to me from the text). If such a connection is embedded in the mechanisms included in Table 1 it should be outlined in the results section where the conclusion is made otherwise simulations targeting the macrophage/DC populations would be needed to confirm this hypothesis.

***Response:*** We thank the reviewer for pointing this out. We want to clarify that computational modeling based studies are capable of providing predictive modeling derived insights, however, any definitive causal connection should be validated in an experimental or clinical setting. In this study, based on the results obtained from the metamodel based global SA, the epithelial cell proliferation parameter was shown to an impact on the H. pylori population.

*Following these findings which highlighted the importance of epithelial cell proliferation, the biological hypothesis derived from this prediction is that the epithelial cell proliferation parameter 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 a varying range of values (0.9-0.1) 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 computational model, as we varied the epithelial cell proliferation parameter. We analyzed the outputs from the hybrid computer model and observed 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. Overall, these cell populations varied due to the variation in the epithelial cell proliferation parameter.*

*We want to clarify that such connection was not embedded in the mechanisms included in Table 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 caused a change in regulatory and tolerogenic dendritic cell population. 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 a mechanism that involves the regulatory macrophages and tolerogenic cells. This is in line with our own conclusions drawn from a previous paper (Viladomiu, Bassaganya-Riera et al. 2017) 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.

We updated the manuscript with the detailed clarification, please refer to L788-826 and L908-L915.

3. Clarity is needed on some parts of the methods description:

3.1 P6, line 131: what are the units of the grid dimensions given. Are these the dimensions of a single grid cell or the entire grid? How are the 4 compartments separated on the grid?

**Response :** We thank the reviewer for pointing this out. These are the dimensions of the entire grid. An individual grid cell is 1nm x 1nm. The 4 compartments are separated by border implementation such that the dimensions of the 4 compartments are lumen (2nm), epithelium (1nm), lamina propria (5nm) and gastric lymph node (2nm). The following compartments are adjacent to each other:

- Lumen - epithelium
- Epithelium - lamina propria
- Lamina propria – gastric lymph node

We updated the manuscript with detailed model description, please refer to L222-L232. We also added a figure describing the grid in the Additional file Fig S2.

3.2 P6 line 149: what data were the ODEs calibrated to? Is there a reference?

**Response:** The CD4+ ODE model was calibrated using the experimental data provided in the Table S1 of the reference - Carbo, Hontecillas et al. 2013. The Particle Swarm algorithm implemented in COPASI was used to determine unknown model parameter values and fully calibrate the model. The intracellular macrophage ODE model was calibrated using a combination of sourced and new data generated from in vitro macrophage differentiation studies, compiled into a dataset provided within S2 file of Leber, Bassaganya-Riera et al. 2016.

We accordingly updated the manuscript, please refer to L155-L166.

3.3 P6 line 150, and p22 line 524: ABM parameters were calibrated to "qualitatively resemble" the patterns observed in in vivo model. What patterns? What is considered to be qualitatively similar enough? Do the simulations reproduce the dynamics as well and the endpoint experimental observations? Inclusion of experimental data alongside the simulations in figure 2 or a description of the key dynamics (e.g. fold-changes, peak values etc.) would go a long way in communicating confidence in the model parameters.

**Response:** We thank the reviewer for their valuable suggestion. The values of the parameters are obtained based on the qualitative comparison of the model outputs with the experimental

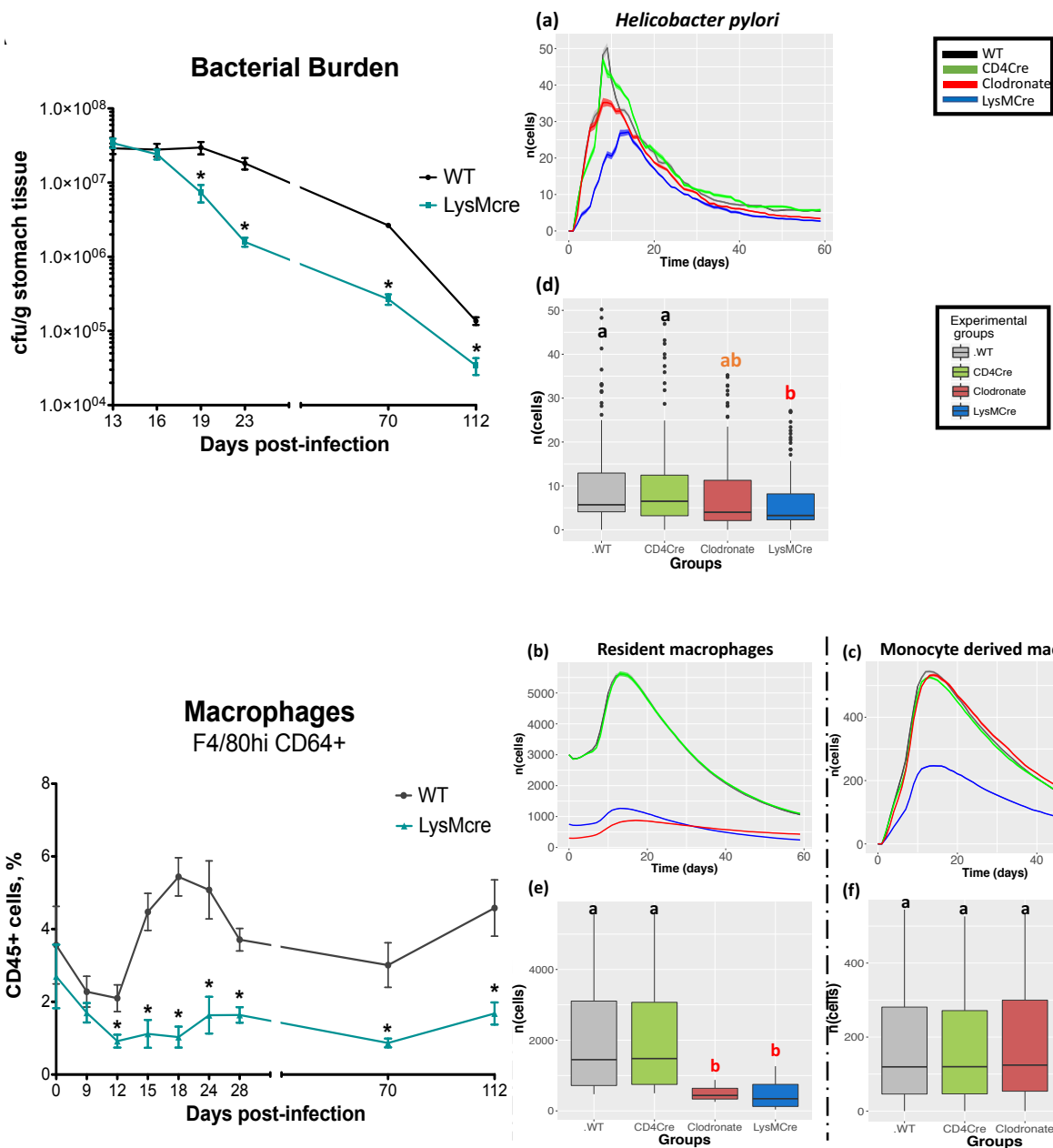


*results obtained from the mouse model of H. pylori infection. The simulations reproduced similar dynamics as described below -*

*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 (lacking PPAR $\gamma$  in myeloid cells) was observed as shown in Fig 1A of (Viladomiu, Bassaganya-Riera et al. 2017). The decrease in bacterial burden led to a significant and sustained lower colonization levels when compared to WT and CD4Cre (lacking PPAR $\gamma$  in T cells). Similar to the results observed in the mouse model, we observed a significant decrease (Fig 2a,d) 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 Fig. 2A, 2E of (Viladomiu, Bassaganya-Riera et al. 2017). 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 (Fig 2b,e and Fig 2c,f) in the number of resident macrophages as well as monocyte derived macrophages in the simulated WT groups in comparison to the simulated LysMcre group. As shown below, the peak of resident macrophages in lamina propria (refer Fig 2b of this paper) was observed at ~16-21 days which was similar to the peak observed in the CD64+F480hi macrophages at day 21, in Fig 2a described in (Viladomiu, Bassaganya-Riera et al. 2017). We included the experimental data alongside the simulation and revised the Fig 2 as shown below.*

*We accordingly updated the manuscript, please refer to L344-L350, L654-L671 and updated legend for Fig 2, L647-L649.*



3.4 P11 line 246: the authors state that they perform global SA of the hybrid computer model. I believe they mean the metamodel here?

**Response:** We thank the reviewer for pointing this out. Although, a metamodel was built using the hybrid computer model, overall the global SA that included two stages –i) screening the influential inputs using PRCC (which was performed on the outputs from hybrid computer model simulations) and building a metamodel (using the outputs from the hybrid computer model)

followed by calculating the Sobol' indices. Hence, we stated that we performed the global SA of the hybrid computer model.

3.5 P 21 line 480 and 484: parameter values were 'reduced' to emulate biological KOs. By how much were the parameters reduced?

**Response:** We thank the reviewer for pointing this out. We added new columns in Table S1 with the values of the parameters used to emulate the biological KOs. A complete set of parameter for each of the biological KOs are included as separate columns in Table S1.

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 baseline (WT) value respectively (refer Table S1). As described in (Carbo, Hontecillas et al. 2013), to simulate the LysMCre experimental conditions, the probabilities of i) a monocyte transitioning to a regulatory macrophage ( $p_{MonotoMreg}$ ) and ii) immature dendritic cells switching to tolerogenic dendritic cells ( $p_{iDCtotDC}$ ) were reduced approximately to 60% and 30% of the baseline (WT) value, respectively (refer Table S1).

We updated the manuscript with the above listed values, please refer to L602-615.

3.6 The *in vivo* model is mentioned several times before it is clarified to be a mouse model.

**Response:** We thank the reviewer for their valuable suggestion. We updated the manuscript and clarified that the *in vivo* model is a mouse model.

**Comment from the Editor:** Further, our series Guest Editor, Paul Macklin has had a quick look at the manuscript from a reproducibility point-of-view and suggests that you include somewhere (e.g., in supporting info) the specific examples for this paper, including detailed instructions on how to create the specific examples presented. Note that our curators also asked for detailed instructions on how to require detailed instructions for usability - not just code.

In addition, please register any new software application in the SciCrunch.org database to receive a RRID (Research Resource Identification Initiative ID) number, and include this in your manuscript. This will facilitate tracking, reproducibility and re-use of your tool.

**Response:** The RRID (Research Resource Identification Initiative ID) number as assigned by the SciCrunch.org database is SCR\_016918. We included this in the manuscript, please refer to L180-L182.

Detailed instructions for the usability are described below and also included in Additional file S1. Detailed instructions on how to create the specific examples presented here are also included. We accordingly updated the manuscript, please refer to L178-180.

### **Additional file S1**

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**) 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 -
  - i. `git clone --recursive https://github.com/NIMML/ENISI-Dependencies`
4. Change the path to the ENISI-Dependencies folder: `cd ENISI-Dependencies`
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 –
  - i. `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_MSM_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. 1<sup>st</sup> column: time points information (*i.e. Ticks*),
  - b. 2<sup>nd</sup> column mean values and
  - c. 3<sup>rd</sup> 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 - *PRCC\_activeinactiveinputs-added.xlsx*.

### Stage 2 Initialization

---

- 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* file.
- 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. 1<sup>st</sup> column: time points information (*i.e.* *Ticks*),
  - b. 2<sup>nd</sup> column mean values and
  - c. 3<sup>rd</sup> 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*.

## References

- Mei, Y., V. Abedi, A. Carbo, X. Zhang, P. Lu, C. Philipson, R. Hontecillas, S. Hoops, N. Liles and J. Bassaganya-Riera (2015). "Multiscale modeling of mucosal immune responses." BMC Bioinformatics **16 Suppl 12**: S2.
- Trevor, H., T. Robert and F. JH (2009). *The elements of statistical learning: data mining, inference, and prediction*, New York, NY: Springer.
- Viladomiu, M., J. Bassaganya-Riera, N. Tubau-Juni, B. Kronsteiner, A. Leber, C. W. Philipson, V. Zoccoli-Rodriguez and R. Hontecillas (2017). "Cooperation of Gastric Mononuclear Phagocytes with *Helicobacter pylori* during Colonization." J Immunol **198**(8): 3195-3204.