

Manuscript Number:	GIGA-D-18-00435R3	
Full Title:	High-Resolution Computational Modeling of Immune Responses in the Gut	
Article Type:	Research	
Funding Information:	Defense Threat Reduction Agency (HDTRA1-18-1-0008)	Dr. Josep Bassaganya-Riera Dr. Raquel Hontecillas
Abstract:	<p>Background: <i>Helicobacter pylori</i> causes gastric cancer in 1-2% of cases, but is also beneficial for protection against allergies and gastroesophageal diseases. An estimated 85% of <i>H. pylori</i>-colonized individuals do not present any detrimental effects. To study the mechanisms promoting host tolerance to the bacterium in the gastrointestinal mucosa and systemic regulatory effects, we investigated the dynamics of immunoregulatory mechanisms triggered by <i>H. pylori</i> using a high-performance computing driven ENteric Immunity Simulator multiscale model. Immune responses were simulated by integrating an agent-based model, ordinary and partial differential equations.</p> <p>Results: The outputs were analyzed using two sequential stages: the first 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 time-varying impact of input parameters during initiation, peak and chronic phases of infection. The study identified epithelial cell proliferation and epithelial cell death as key parameters that control infection outcomes. In-silico validation showed that colonization with <i>H. pylori</i> 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 <i>H. pylori</i> infection identified epithelial cell proliferation as a key factor for successful colonization of the gastric niche and highlighted the role of tolerogenic dendritic cells and regulatory macrophages in modulating the host responses and shaping infection outcomes.</p>	
Corresponding Author:	Raquel Hontecillas, Ph.D. Virginia Polytechnic Institute and State University Blacksburg, Virginia UNITED STATES	
Corresponding Author Secondary Information:		
Corresponding Author's Institution:	Virginia Polytechnic Institute and State University	
Corresponding Author's Secondary Institution:		
First Author:	Meghna Verma, M.S.	
First Author Secondary Information:		
Order of Authors:	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.	

Order of Authors Secondary Information:	
Response to Reviewers:	<p>Once more, we would like to truly thank the reviewers and editor for the time and effort they are putting in reviewing our work. The fact that they've all brought up concerns about the grid dimensions and structure underlines the importance of computational biology as means of integrating biological concepts with biophysics and mathematics. Most biologists tend to focus on functions (i.e., gene expression or protein properties) and numbers, but they rarely consider the dimensions of the "biological space". The proposed revisions and clarifications better connect the biology, underlying mathematics and spatial/biophysical considerations. In response to the comments raised by the reviewers, we have revised the manuscript and provided explanations to further clarify the description of the grid.</p> <p>Reviewer reports:</p> <p>Reviewer #2 Comment 1: I have looked through the authors' responses and remain concerned about the grid dimensions. Redefining the grid in um instead of nm still leaves the entire simulation environment at 30um by 10um which is about on the order of a single cell (e.g. macrophage diameter ~20um https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1470168/). I remain concerned that defining the grid at such a small scale creates a biologically unrealistic disconnect between the diffusing cytokines (on a very small scale) and the cells that they influence (on a much larger and abstract scale). I tried to understand the dimensions of the diffusion constants for the cytokines but the parameter table S1 contains no units to be able to compare diffusion scales to cell movement scales.</p> <p>Response to Reviewer 2, Comment 1: We thank the reviewer for their concern. We agree that defining the grid in m leaves the simulation in a smaller scale and want to clarify that even though the units in the model are annotations, we understand that it is crucial to define the grid in a more biologically meaning way.</p> <p>We define the area in the model being simulated as a simulation environment with 30 mm x 10 mm two-dimensional grid. The size of an individual lattice site (previously referred to as grid cell in the paper) is 1 mm x 1mm. The scales described as in the previous version of ENISI-MSM (Mei et al., 2015) were kept unchanged. The table describing the scales used in (Mei et al., 2015) are also shown here in Table 2. As described in table shown below (adapted from Table 1) of our previous work (Mei et al., 2015), the spatial properties for cytokine diffusion defined in the range of millimeters were unchanged in the version of ENISI-MSM used in this paper.</p> <p>Scale</p> <p>Example scenario</p> <p>Spatial(m)</p> <p>Time(s)</p> <p>Technology</p> <p>Tool</p> <p>Intra-cellular</p> <p>Signaling pathways</p> <p>Nano (nm)</p> <p>Nano</p> <p>ODE</p> <p>COPASI</p> <p>Cellular</p> <p>Cell movement and subtype</p> <p>Milli (mm)</p> <p>Tens</p> <p>ABM</p> <p>ENISI</p> <p>Intra-cellular</p> <p>Cytokine-diffusion</p> <p>Milli (mm)</p> <p>Tens</p> <p>PDE</p> <p>ValueLayer</p> <p>Tissue</p>

Inflammation and lesions
 Centi (cm)
 Thousands
 Projection
 ENISI
 Table 2. The four scales of ENISI models, their spatial and temporal properties and modeling technologies and tools used for each scale. (Table 1 as adapted from (Mei et al., 2015))

We updated the manuscript accordingly, please refer to L213-L223.

We thank the reviewer for pointing this out. We updated Table S1 with units for clarity. Lastly, we want to clarify that the model deals with numbers and the units are annotations in the simulation hence the corrections in the dimensions above do not affect the simulation results in any way. All the values used in the code were internally consistent with the model.

Reviewer #3, Comment 1: In this review, I am looking at the more limited questions on the matter of units and scale, which have been raised by both Reviewer #1 and Reviewer #2.

Generally, I don't think the units matter, so long as they have been handled correctly (i.e., with correct conversion and internal consistency within the computational model). The units should be properly labeled in any parameter tables.
 I tend to use microns and minutes (or seconds for some problems) for this scale of problem, but there are others who just use cm and sec for everything. As long as the values are correct in the displayed units (and as long as the code used internally consistent values), I think it's purely aesthetic. That said, if something is being labeled on a multicellular level, then labeling 10 microns will be much more appropriate than 10,000 nm.

Response to Reviewer 3, Comment 1: We thank the reviewer for highlighting this point. We concur that the units don't matter for the simulation results, but want to best clarify the dimensions to make them relevant with the biology. We have labelled all the units in the parameter Table S1. The values used in the code were internally consistent with the model and handled correctly. For further clarity, we show the scales used in the previous ENISI models and the spatial and temporal properties as described in (Mei et al., 2015) and included the Table 1 from (Mei et al., 2015).

Scale
 Example scenario
 Spatial(m)
 Time(s)
 Technology
 Tool
 Intra-cellular
 Signaling pathways
 Nano (nm)
 Nano
 ODE
 COPASI
 Cellular
 Cell movement and subtype
 Milli (mm)
 Tens
 ABM
 ENISI
 Intra-cellular
 Cytokine-diffusion
 Milli (mm)
 Tens
 PDE
 ValueLayer

Tissue
Inflammation and lesions
Centi (cm)
Thousands
Projection
ENISI

Table 2. The four scales of ENISI models, their spatial and temporal properties and modeling technologies and tools used for each scale. (Adapted from (Mei et al., 2015)).

We updated the manuscript accordingly, please refer to L213-L223.

Reviewer #3, Comment 2: However, I want to further dig into Reviewer #2's concerns on grid sizes and scales.

As I read this draft, I see at least one potential source of confusion: this team appears to be very focused on mathematical and numerical methods. As such, they are using the word "cell" for both a biological cell and a computational lattice site. This is a really bad idea, and the authors should pick a better nomenclature (e.g., computational mesh or lattice site) to avoid this confusion. They should never use "cell" to mean anything other than a biological cell once they enter computational biology.

Response to Reviewer 3, Comment 2: The focus of the paper was to utilize an already published tool (Mei et al., 2015) to study *Helicobacter pylori* infection. So, a significant focus of the work was on the biology. However, we agree that the nomenclature should be clarified to avoid confusion from using the term "cell" for both actual biological cells and space/grid units. In the revised version of the manuscript we have substituted "grid cell" for "lattice site" and kept "cell" to refer to biological cells.

Reviewer #3, Comment 3: Next, they need to be clearer about what their grids represent. They should show a picture of the domain and meshing in their main text, and not just supplementary material. They should probably also clarify that they are simulating a cross-section of tissue in their model, rather than 3D or some top-down view. (At least from what I can tell.) (They seem to address this in text, but showing the mesh will provide better clarity.)

Response to Reviewer 3, Comment 3: We thank the reviewer for their valuable suggestion regarding this crucial point. We included a cartoon picture of the domain and mesh (as Fig. 2) in the main text as opposed to in the supplementary material (as shown below).

We highlighted that we are simulating a cross-section of tissue in the model and we redefined it as a simulation environment with 30 mm x 10 mm two-dimensional grid. The size of an individual lattice site (previously referred to as grid cell in the paper) is 1 mm x 1mm.

We updated the manuscript accordingly, please refer to L231-L235.

Reviewer #3, Comment 4: Assuming the authors now have a 30 micron x 10 micron domain, they can simulate at most one epithelial cell, if it's all in plane. But if it's a cross-section, I suppose they could have more. Perhaps many *h pylori* (which their size is about 1 micron), but not many mammalian cells. So, their computational domain is still not very clear to me, and they should just show it, with appropriate labeling. They seem to skip straight to population dynamics in their figures, but it would be very helpful if they showed one actual spatiotemporal simulation. This would make the nature and performance of their model much clearer.

Response to Reviewer 3, Comment 4: We thank the reviewer for their concern. With a redefined computational simulation environment of 30 mm x 10 mm, the epithelium is comprised of hundreds of epithelial cells.

For example, if the initial number of epithelial cells defined by the user is 12, the total number of epithelial cells amounts to = (30 x 1) dimension of epithelial compartment x 12 initial number = 360.

In addition to the figure of the grid environment, we included the screenshots of one actual *in silico* simulation of *H. pylori* infection to highlight the spatiotemporal aspects of

the modeling outputs. The screenshots were created using VisIt version 2.12 (Childs et al., 2012), an interactive visualization and analysis tool. As shown in Additional file Fig. S2 the screenshots represent the spatial distribution of different agent cells over time points (2, 4, 5 and 6) distributed across the 2D grid. Further, we presented the insets in Fig. S2 showing a zoomed in portion of the respective grids across the time steps 2, 4, 5 and 6.

We also want to clarify that the agents represented in the screenshots below are only for visual representation and do not represent the actual size of the biological cells.

We updated the manuscript accordingly and added Fig 3, please refer to L333-L338.

Fig S2. Time screenshots of a *Helicobacter pylori* infection modeled in a 30 mm (length) x 10 mm (width) two-dimensional grid. The thickness of the compartment is shown on the y-axis, such that: lumen spans (0 to 2) units, epithelium spans (2 to 3) units, lamina propria spans (3 to 8) units and gastric lymph node across (8-10) units on the scale. Two-dimensional distribution of different cell subsets over the time steps (ticks) 2, 4 (top panels), 5 and 6 (bottom panels) are shown. The insets in each image shows a zoomed in portion of the respective grids across the time steps 2, 4, 5 and 6. The agents represented in the screenshots below are only for visual representation and do not represent the actual size of the biological cells. Future refinements of the model will create agents of the actual sizes of cells.

Reviewer #3, Comment 5: On sizes and scaling, I fully agree with Reviewer #2: if this is indeed a 30 micron x 10 micron cross section, there's no way there are more than a handful of mammalian cells at any time in any simulation. If they have made a scaling argument (and there are such arguments that could be made if formulated clearly and rigorously), they'd better be clear about it. Any results that show thousands of mammalian cells in a 30 x 10 micron² domain are simply beyond biophysical plausibility.

Response Reviewer 3, Comment 5: We defined the simulation environment as a 30 mm x 10 mm two-dimensional grid that represents a cross section area of stomach tissue modeled here.

Reviewer #3, Comment 6: Again, just actually showing a simulation (either a movie, or some time snapshots, but showing locations of all the cell agents and substrate distributions) would help clarify things much more. No limit to the number of cells in a mesh site, while mathematically possible, does not make sense on such a small simulated domain. Even though the authors treated the cell agents as no size (infinitesimal points), there are physical limits, and moreover if each cell is absorbing / secreting things at appropriate rates, then there should be ridiculous amounts of secretion of growth factors and ridiculous depletion of growth substrates, if there is a huge overcrowding of hundreds or thousands of mammalian cells in a 1 micron x 1 micron lattice site.

Response to Reviewer 3, Comment 6: We thank the reviewer for their valuable suggestion and have included time snapshots for the simulation (time points 2, 4, 5 and 6) created using VisIt version 2.12 (Childs et al., 2012), an interactive visualization and analysis tool. Please refer to Additional file Fig S2 also included in the above response to Comment 5.

The hybrid multiscale modeling platform ENISI MSM is currently capable of scaling up to up to 109 agents, at which the memory (on a 32 GB node) was exceeded due to the large number of agents.

Reviewer #3, Comment 7: If that's what's going on, the authors really do need to take a step back and consider domains sufficiently large to capture hundreds or thousands of mammalian cells. This is not simply a matter of relabeling axes: it's a matter of

simulating a larger physical domain that is suited to the size of objects (mammalian cells) that they are considering, with biophysically reasonable parameters. The authors need to carefully review all their parameters (e.g., cell densities) to ensure they are correctly scaled and reasonable. If there is nobody on the team with sufficient domain expertise to review these parameters and results to check for reasonableness, it may be time to grow the team. Given that actual mammalian cells are much larger than the computational lattice sites, there must be constraints: if a lattice site is "occupied" by a mammalian cells, so are many (or most!) of the surrounding lattice sites. They would be no-fly zones for further mammalian cells. The common way to solve such problems is to use computational lattice sites that are of comparable size to the largest biological cells (e.g., 1 biological cell per lattice site, as a cellular automaton model), or use very large lattice sites (e.g., 100 micron x 100 micron) that can truly contain multiple cells.

Response to Reviewer 3, Comment 7: We thank the reviewer for their suggestion on the size of the lattice site.

In the current model, the size of the lattice site (referred as spatial grid previously) is 1mm x 1mm, capable of containing multiple cells.

The total number of agents in a compartment of size (length x width) is calculated as follows: $\text{number (agents)} \times \text{size (compartment)}$,

For example, if the user set the initial number of epithelial cell agents to be 12, the total number of epithelial agents within the epithelium compartment amounts to - $12 \times \text{initial_number} \times (30 \times 1) \times \text{size_epithelium_compartment} = 360$.

Reviewer #3, Comment 8: All is not lost, however. If for some reason that larger domain is computationally infeasible (hard to imagine), the authors really don't need a 1 micron mesh resolution for most of the effects here. The diffusion length scale of most chemokines and diffusing substrates would not require a 1 micron mesh resolution. (And numerical stability will improve if coarser mesh resolutions are used.) If the authors feel a 1 micron mesh is needed for the bacteria, they could easily use a separate mesh. This work looks interesting. I think the team has a great contribution to make, if they pay a bit closer attention to the biophysical limits of their system.

Response to Reviewer 3, Comment 8: The computational domain used here is of 30 x 10 units in size and the individual lattice site is 1 units for the simulation. The lattice site is a configurable run parameter and can be changed without modifying the model.

We thank the reviewer for their valuable suggestions, inputs and concerns and have tried to clarify the questions around the grid dimensions.

With these revisions and clarifications, we believe that the revised manuscript is acceptable for publication.

Thank you for considering this work.

References:

Childs, H., Brugger, E., Whitlock, B., Meredith, J., Ahern, S., Pugmire, D., ...Weber, G. (2012). VisIt: An End-User Tool For Visualizing and Analyzing Very Large Data. High Performance Visualization-Enabling Extreme-Scale Scientific Insight. Insight, 357-372.
 Mei, Y., Abedi, V., Carbo, A., Zhang, X., Lu, P., Philipson, C., ..Bassaganya-Riera, J. (2015). Multiscale modeling of mucosal immune responses. BMC Bioinformatics, 16 Suppl 12, S2. doi:10.1186/1471-2105-16-s12-s2

Additional Information:

Question

Response

Are you submitting this manuscript to a special series or article collection?

No

Experimental design and statistics

Yes

<p>Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. 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>	
<p>Resources</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 Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?</p>	<p>Yes</p>
<p>Availability of data and materials</p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript.</p> <p>Have you have met the above requirement as detailed in our Minimum Standards Reporting Checklist?</p>	<p>Yes</p>

[Click here to view linked References](#)

High-Resolution Computational Modeling of Immune Responses in the Gut

Meghna Verma^{1,2}, Josep Bassaganya-Riera¹, Andrew Leber¹, Nuria Tubau-Juni¹,
Stefan Hoops¹, Vida Abedi¹, Xi Chen³, Raquel Hontecillas^{1,*}

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

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

³Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA.

* Correspondence:

Dr. Raquel Hontecillas

Email: rmagarzo@vt.edu

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

Email address for all authors:

Meghna Verma: meghna89@vt.edu

Josep Bassaganya-Riera: jbassaga@vt.edu

Andrew Leber: ajleber@vt.edu

Nuria Tubau-Juni: nuriaj@vt.edu

Stefan Hoops: shoops@vt.edu

Vida Abedi: vidaabedi@gmail.com

Xi Chen: xchen6@vt.edu

Raquel Hontecillas: 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 is
9
10 3 also beneficial for protection against allergies and gastroesophageal diseases.
11
12 4 An estimated 85% of *H. pylori*-colonized individuals do not present any
13
14 5 detrimental effects. To study the mechanisms promoting host tolerance to the
15
16 6 bacterium in the gastrointestinal mucosa and systemic regulatory effects, we
17
18 7 investigated the dynamics of immunoregulatory mechanisms triggered by *H.*
19
20 8 *pylori* using a high-performance computing driven **EN**teric **I**mmunity **S**imulator
21
22 9 multiscale model. Immune responses were simulated by integrating an agent-
23
24 10 based model, ordinary and partial differential equations.

25
26 11 **Results:** The outputs were analyzed using two sequential stages: the first used
27
28 12 a partial rank correlation coefficient regression-based and the second employed
29
30 13 a metamodel-based global sensitivity analysis. The influential parameters
31
32 14 screened from the first stage were selected to be varied for the second stage.
33
34 15 The outputs from both stages were combined as a training dataset to build a
35
36 16 spatiotemporal metamodel. The Sobol' indices measured time-varying impact of
37
38 17 input parameters during initiation, peak and chronic phases of infection. The
39
40 18 study identified epithelial cell proliferation and epithelial cell death as key
41
42 19 parameters that control infection outcomes. *In-silico* validation showed that
43
44 20 colonization with *H. pylori* decreased with a decrease in epithelial cell
45
46 21 proliferation, which was linked to regulatory macrophages and tolerogenic
47
48 22 dendritic cells.

49 23 **Conclusion:** The hybrid model of *H. pylori* infection identified epithelial cell
50
51 24 proliferation as a key factor for successful colonization of the gastric niche and
52
53
54
55
56
57
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

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

27 **1. Background**

28 Computational modeling of the immune response dynamics can provide
29 novel insights and facilitate the systems level understanding of the interactions
30 at the gastric mucosa during infection. Ordinary differential equation (ODE-
31 based methods are deterministic and based on the average response of cells
32 over time. Dynamical models are used in immunology for system-level analyses
33 of CD4+ T cell differentiation [1], macrophage differentiation [2], immune
34 responses elicited by *Clostridium difficile* infection [3], co-infections [4], and in
35 cancer and immunotherapy [5]. However, ODE-based models lack the spatial
36 aspects and the features to study the organ and immune cell topology over time.
37 Agent-based models (ABM) employ a bottom-up approach that focuses on the
38 spatial and temporal aspects of individual immune cells, unlike the ODE-based
39 methods. This rule-based method includes agents that act as local entities which
40 interact locally with other agents, move in space, and follow set of rules
41 representing their role in a given system and contribute towards generating an
42 emergent behavior. Since, the immune system is a complex dynamical system
43 [6] wherein the components *i.e.*, the immune cells move in space and time
44 changing their location, ABMs are useful tools that can be employed to
45 understand biological mechanisms and the hidden insights.

46 *Helicobacter pylori* is a gram-negative bacterium that has persistently
47 colonized the human stomach since early evolution [7] [8] and is currently found
48 in over 50% [9] of the global population. *H. pylori* has co-evolved with humans
49 for thousands of years, such that an estimated 85% of the *H. pylori*-colonized

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

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

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

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

76 summarized in [17, 18]. The comparison between different multiscale modeling
77 tools and agent-based immune simulators, are discussed in [12, 19].

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

93 To understand the dynamics and emergent immunological patterns
94 described by this hybrid model, we employed sensitivity analysis (SA), an
95 important part of the model analysis used to explore the influence of varying
96 model parameters on the simulation outputs. The influence of the effects of
97 changes in parameter values on the model output explains the model dynamics
98 that underlay the outputs [21, 22]. Furthermore, SA examines the robustness of
99 the model output at a different range of parameter values that correspond to a
100 range of different assumptions. We employed global SA and conducted a two-
101 stage spatiotemporal global SA approach. First, we used a regression-based

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

method such as the partial rank correlation coefficient (PRCC) and screened the important input parameters that were shown to have the most influence on the output cell populations obtained from the hybrid model. Second, the screened input parameters from the first stage were varied to build a second stage parameter design matrix, and the computer simulations were again run using the hybrid ENISI model. The outputs from both analytics stages were combined and used as a ‘training dataset’ to build a spatiotemporal Gaussian process based metamodel. Finally, variance-based decomposition global SA was used to compute the Sobol’ indices and the most influential parameters over the course of infection were identified. The data analytics methods conducted on the hybrid model identified the epithelial cell parameters such as epithelial cell proliferation as the most influential ones, required for the successful colonization of *H. pylori* in the gastric microenvironment.

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 30 mm x 10 mm. An individual lattice site for our simulation is 1 mm x 1 mm, however, this is a configurable run parameter and can be changed without modifying the model. An individual lattice site is a unit wherein all the agents located within that location have the same cytokine environment, *i.e.*, for all the agents in that location, ENISI-MSM would send the same concentration of the cytokines to

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

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 movement of the agents. At the cellular scale, ENISI MSM, simulated epithelial cells, macrophages, dendritic cells (DC), CD4+ T cells and bacteria that are implemented as agents in the model. At the intracellular scale, calibrated ODE-based models of T cells [23] and macrophages [2] were used to represent the intracellular pathways controlling cytokine production. The CD4+ T cell ODE model was calibrated using the experimental data provided in the **Table S1** of [23]. The Particle Swarm algorithm implemented in COPASI was used to determine unknown model parameter values and fully calibrate the CD4+T cell ODE model, the details are described in [23]. The intracellular macrophage ODE model was calibrated using a combination of sourced and new data generated from *in vitro* macrophage differentiation studies, that were compiled into a dataset provided within S2 file of [2]. The parameter values are specified within

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

152 the previously published manuscripts - CD4+ T cell ODE model (Carbo,
153 Hontecillas et al. 2013) and macrophages [2]. The parameters of the calibrated
154 ODEs were kept unchanged, and the ABM parameters were calibrated by
155 approximating the output simulations such that they qualitatively resembled the
156 patterns observed in a mouse model of *H. pylori* infection [24], also described in
157 detail in section 3.1. Cytokines secreted by immune cells and their change in
158 concentration were modeled by PDE. The degradation value of the cytokines
159 and the diffusion constant determines the spread of the cytokine value of one
160 [lattice site](#) to its neighboring [lattice site](#) similar to as described in [our previous](#)
161 [work](#) [12]. The features of ABM, ODE, and PDE were combined to create a
162 multiscale modeling environment which spanned across different orders of
163 spatiotemporal scales. The model output contains information about the x and
164 y co-ordinate of the agents at every time point. The cytokines and internal
165 signaling pathways that drive functional fates of cells are well mixed within a cell,
166 *i.e.*, we have only temporal resolution within the cell during a time step. Since,
167 the model is capable of providing information regarding spatial co-ordinates over
168 time, we claim the model to be a spatio-temporal model.

170 The code for the hybrid model is freely accessible and can be downloaded
171 at <https://github.com/NIMML/ENISI-MSM>. The detailed instructions for the
172 usability, instructions on ‘how to run a simulation’ and codes for creating specific
173 examples presented here are presented in Additional file *S1*. The SciCrunch.org
174 database assigned research identification initiative ID (RRID) for ENISI-MSM is
175 RRID:SCR_016918. The design of the implementation of the code structure is
176 depicted in the Additional file **Fig S1**. The hybrid model is implemented in C++
177 and utilized the Repast HPC library (https://repast.github.io/repast_hpc.html)

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

[25]. For the ODEs, we utilized COPASI [26], an ODE-based modeling tool used in computational biology. The rules in the model that described the interaction of *H. pylori* with the gastric mucosa and the immune responses resulting from the infection are derived from the findings in our previously published studies [1, 2]. Specifically, this hybrid model reproduced the immune responses generated by the interaction *H. pylori* and the resident macrophages as shown in the mouse model of *H. pylori* infection [24]. The rules for each cell type in the *H. pylori* infection are summarized in **Table 1**. A pictorial representation of the rules is depicted in *Fig 1*. These cell types represented as agents, act according to the rules (as in **Table 1**) that are updated at discrete simulation cycle.

Fig 1. *Helicobacter pylori* infection schematic diagram of the hybrid ABM ODE model

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

ODE components in the hybrid model. The cellular agents are simulated in a two-dimensional grid space with their behavior defined by a set of rules during a course of *H. pylori* infection.

Model description

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

Spatial discretization

The model has a spatial discretization such that we define the area being simulated as a simulation environment with a two-dimensional grid whose size is 30 mm x 10 mm. An individual lattice site is 1mm x 1mm, however, this is a configurable run parameter and can be changed without modifying the model. We further want to clarify that the above units in the model are annotations and purely aesthetic. The scales described in the previous version of ENISI-MSM [12] were kept unchanged. The table describing the scales used in [12] are also shown here in Table 2.

<u>Scale</u>	<u>Example scenario</u>	<u>Spatial (m)</u>	<u>Time(s)</u>	<u>Technology</u>	<u>Tool</u>
<u>Intra-cellular</u>	<u>Signaling pathways</u>	<u>Nano (nm)</u>	<u>Nano</u>	<u>ODE</u>	<u>COPASI</u>
<u>Cellular</u>	<u>Cell movement and subtype</u>	<u>Milli (mm)</u>	<u>Tens</u>	<u>ABM</u>	<u>ENISI</u>
<u>Intra-cellular</u>	<u>Cytokine-diffusion</u>	<u>Milli (mm)</u>	<u>Tens</u>	<u>PDE</u>	<u>ValueLayer</u>
<u>Tissue</u>	<u>Inflammation and lesions</u>	<u>Centi (cm)</u>	<u>Thousands</u>	<u>Projection</u>	<u>ENISI</u>

Table 2. The four scales of ENISI models, their spatial and temporal properties and modeling technologies and tools used for each scale (Table 1 as adapted from [12]).

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

225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250

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

Fig 2. A pictorial representation of the spatial discretization of the 2D grid.

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

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

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

1
2
3
4 251
5
6 252
7
8
9 253
10
11 254
12
13 255
14
15 256
16
17
18 257
19
20 258
21
22 259
23
24 260
25
26 261
27
28
29 262
30
31 263
32
33 264
34
35 265
36
37
38 266
39
40 267
41
42 268
43
44 269
45
46
47 270
48
49 271
50
51 272
52
53 273
54
55 274
56
57
58 275
59
60 276
61
62
63
64
65

Time Step size

The time step size is 1 tick ~ 1 day which was obtained during the process of qualitatively comparing the output to the results from the mouse model of *H. pylori* infection. For e.g., the peak of resident macrophages in lamina propria (refer Fig 3b, d) is observed at ~21 days which is similar to the results obtained in Fig 2A described in [24] (also described in detail in section 3.1).

Updating

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

The COPASI setup for the solver used the LSODA (Livermore Solver for Ordinary Differential Equations) differential equation solver. The default values for the setup such as the - relative tolerance (1e-6), absolute tolerance (1e-12) and maximum internal steps of 10000 were maintained. The ENISI MSM sends the current concentrations of the cytokines to COPASI. COPASI uses those values to integrate the deterministic model for one tick, i.e., 1 day. The resulting time series of cytokine

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

277 concentrations are used to update the cytokine value in the ABM/PDE
278 system. COPASI simulates different model for each relevant cell type.
279 The ENISI MSM PDE solver uses a simple numerical scheme to solve the PDEs
280 (<https://github.com/NIMML/ENISI-MSM/tree/master/src/diffuser>) and process
281 distributed value layer (<https://github.com/NIMML/ENISI-MSM/blob/master/src/grid/ValueLayer.h>). The ValueLayer stores the value for a
282 grid space and provides methods to change the values of individual [lattice site](#).
283 The Diffuser is used to diffuse the values of the ValueLayer using diffusion (d)
284 and degradation (delta) constants as described in [12]. The diffusion constant
285 determines the migration of values of a [lattice site](#) to its neighboring [lattice site](#).
286 As implemented in ValueLayer library, the diffusion of cytokines follows the
287 equation shown below also described in Mei et al, 2015. Here, v_n is the value of
288 the [lattice site](#) itself at step n. The values of c_{delta} and c_d are degradation and
289 diffusion constant respectively.
290

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

0.3	1.2	0.3
1.2	-6.0	1.2
0.3	1.2	0.3

292
293 The PDE solver uses the above number scheme c_d^{neighbor} for the diffusion
294 process. The step size c_{delta} is automatically adjusted at the beginning of the
295 simulation based on the degradation and diffusion constants to avoid underflow

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

296 errors, *i.e.*, multiple PDE steps are in general executed per tick. The grid size is
297 the identical with the spatial discretization for the agents.

298 ***Movement***

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

312
313 The hybrid model simulations were run on an Ivy Bridge-EX E7-4890 v2 2.80
314 GHz (3.40 GHz Turbo) quad processor nodes. The code was parallelized such
315 that the simulation time on a single node with four parallel tasks, varied between
316 9-10 minutes. This runtime was based on the model parameters at the initiation
317 stage, which included the number of immune cell, bacteria, epithelial cells,
318 number of time steps, and size of the two-dimensional grid. To facilitate the
319 investigation of the mechanisms underlying host responses during *H. pylori*
320 infection, anatomical and functional compartments were spatially linked such that

the agents had both unidirectional and bidirectional movement. All the agents worked in a synchronous format wherein the two agent populations (macrophages and T cells) made function calls to their respective ODE models [2] [23]. These agents used the varying cytokine concentration (*i.e.*, environment variable) in their grid spaces as inputs to the ODE model, and these models were run using COPASI [26].

Table 3 shows information on the agents and the states that they can acquire.

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

Table 3. 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”.

Further, we included the screenshots of one actual in silico simulation of *H. pylori* infection to highlight the spatiotemporal aspects of the modeling outputs. The time snapshots were created using *VisIt* version 2.12 [27], an interactive

1
2
3
4 337 visualization and analysis tool. As shown in Additional file, Fig S2, the
5
6 338 screenshots at time points 2, 4, 5 and 6 represent the spatial distribution of
7
8 339 different agent cells over time distributed across the 2D grid.
9
10 340

13 341 **2.2 Global sensitivity analysis**

15 342
17 343 To conduct the global SA, we determined a list of 38 parameters to be varied
19
20 344 that were selected based on the calibration process (wherein the parameters
21
22 345 that did not show a lot of variation were not included). A range of values
23
24 346 (maximum and minimum) was specified for each of the parameters (refer to
25
26 347 Additional file **Table S1**) by expert judgment, summarized by bounded intervals.
27
28 348 The practice of using expert judgment is known in the SA field as supported in
29
30 349 [28]. As discussed in [29], one of the challenges encountered using ABM is the
31
32 350 process of determining the parameter values, for *e.g.* this may include the lack
33
34 351 of the availability of experimental techniques to measure such parameters. The
35
36 352 values of the parameters for the model presented here are obtained via the best
37
38 353 guess based on the qualitative comparison of the computer model outputs with
39
40 354 that of the experimental results obtained from the mouse model of *H. pylori*
41
42 355 infection (Viladomiu, Bassaganya-Riera et al. 2017) (as described in detail in
43
44 356 Section 3.1). Since, the source of the parameters is not known we estimated the
45
46 357 values to fit the data obtained from the mouse model of infection.
47
48
49
50

51 358 The values of these parameters were normalized within the range of 0 and 1
52
53 359 for SA purposes. We employed a two-stage metamodeling methodology to
54
55 360 determine the influence of each input parameter to the model output, in a high
56
57 361 dimensional screening setting inspired by [30]. The step-wise procedure is
58
59 362 described in the Additional file, Fig S3. All the files for global SA are freely
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

363 accessible and can be downloaded at [https://github.com/NIMML/Sensitivity-](https://github.com/NIMML/Sensitivity-Analysis)
364 [Analysis](https://github.com/NIMML/Sensitivity-Analysis).

365 The two-stage global SA is described in detail in the below section. To
366 summarize, for the first stage the input parameter matrix was designed using the
367 method described in Moon, Dean et al. 2012 and simulations were run using the
368 hybrid computer model. The simulation output from the first stage was analyzed
369 using PRCC as it was computationally efficient, and the active inputs (significant
370 effect) were screened to reduce the input parameter space. Second, the active
371 parameters were varied whereas the inactive parameters from the first stage
372 were maintained at a nominal value for the input parameter matrix design to be
373 employed for the second stage. Third, the simulation outputs from both stages
374 were combined and used as a training dataset to fit a spatio-temporal
375 metamodel. Fourth, the unknown model parameters for the spatio-temporal
376 metamodel were estimated using the maximum log-likelihood function. The
377 spatio-temporal metamodel was used as a substitute for the hybrid computer
378 model, and the variance-decomposition method was used to compute the Sobol'
379 total and first-order indices. Overall, we employed both approaches, PRCC
380 based (for screening) and Sobol' indices calculation to perform a complete global
381 SA of the hybrid computer model. The following sections, describe a detailed
382 step by step explanation of the procedure.

383
384 *Design of two-stage experiments and analysis*

385
386 The input for the hybrid model are varying parameter values obtained from
387 the design matrix and the output are the number of cells (agents) that vary over
388 time. The first stage experiment was focused on the screening of the input

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 variables to reduce the number of input parameters to vary for the SA and to limit
390 the computational cost. Computational costs are often a limiting factor that play
391 an important role in the inclusion of model parameters in the SA [21]. For the
392 design, we assumed the total number of input parameters under consideration
393 as d (in our case, 38). With an assumption of a maximum of 50% active inputs
394 that is aimed to improve the screening performance, the number of runs for stage
395 1, was fixed to $n_1 = 4d$, such that $n_1 > 5 \cdot d \cdot 0.5 = 2.5d$ as in [30]. To construct a n_1
396 $\times (n_1 - 1)$ preliminary input parameter design matrix, X^* , needed to be constructed
397 ([30]). The input parameter design matrix for first stage sampling was drawn from
398 X^* .

399 The algorithm for the first stage design generated a design matrix $X^{(1)}$ that
400 satisfied the below three listed properties as in [30]

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

409
410 First stage sampling plan:

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

1
2
3
4 414 In our case, the model comprised of $d = 38$ input variables. The total number of
5
6 415 distinct input parameter design points obtained using the above procedure was
7
8
9 416 $n_1 = 152$ ($4 \cdot d = 4 \cdot 38$). To account for the variability in the output, we run 20
10
11 417 replicates (r). Thus, the total number of simulations run using the hybrid model
12
13 418 computer simulator with $X^{(1)}$ as input parameter design matrix, were $r \times n_1 = 20$
14
15 419 $\times 152 = 3040$.

16 420
17
18
19
20 421 First stage analysis

21
22
23 422 We analyzed the outputs from first stage analysis and screened the active
24
25 423 inputs from using PRCC. To measure the effect of input parameter on output, we
26
27
28 424 performed both PRCC and the spearman rank correlation coefficient (SRCC)
29
30 425 analysis. PRCC and SRCC were chosen because they were computationally
31
32 426 efficient (accounting for the low computational budget). A correlation analysis
33
34 427 provides a measure of the strength of linear association between input and
35
36
37 428 output variable [31]. A correlation coefficient between x_j and y is calculated as
38
39 429 follows:

40
41
42
43 430
$$r_{x_j y} = \frac{Cov(x_j, y)}{\sqrt{Var(x_j)Var(y)}} = \frac{\sum_{i=1}^N (x_{ij} - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_{ij} - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

44
45
46
47
48
49 431
$$j = 1, 2, \dots, k.$$

50
51
52 432 where $Cov(x_j, y)$ stands for the covariance between x_j and y , and $Var(x_j)$ and
53
54 433 $Var(y)$ are the variance of x_j and y respectively.

1
2
3
4 434 PRCC is performed when i) a non-linear but monotonic relation exists
5
6 435 between the input and outputs, and ii) when little or no correlation exists between
7
8
9 436 the input variables (which is guaranteed by the property (i) of our input parameter
10
11 437 matrix, $X^{(1)}$ described above). As described in Marino, Hogue et al. 2008, the
12
13 438 PRCC between rank transformed x_j and y is the CC between the two residuals
14
15 439 $(x_j - \widehat{x}_j)$ and $(y_j - \widehat{y}_j)$ where \widehat{x}_j and \widehat{y}_j are rank transformed and follow the linear
16
17
18 440 regression models as follows:

19
20
21 441
$$\widehat{x}_j = c_o + \sum_{\substack{p=j \\ p \neq j}}^k c_p x_p \text{ and } \widehat{y}_j = c_o + \sum_{\substack{p=j \\ p \neq j}}^k c_p x_p .$$

22
23
24
25
26 442 We performed the PRCC analysis on the outputs obtained from the hybrid
27
28 443 computer model with $X^{(1)}$ as an input, using ‘*epi.prcc*’ package in R ([https://cran.r-](https://cran.r-project.org/web/packages/epiR/epiR.pdf)
29
30 444 [project.org/web/packages/epiR/epiR.pdf](https://cran.r-project.org/web/packages/epiR/epiR.pdf)). The significance test evaluated the
31
32
33 445 strength of influence each input parameters and assessed if the PRCC
34
35 446 coefficients were significantly different than zero [31]. We run the PRCC analysis
36
37 447 for 13 output cell populations (*Fig 4* shows data for two output populations and
38
39 448 the rest of the data not shown) and identified the active input parameters using
40
41
42 449 the significance test. PRCC and SRCC produced identical outputs, hence results
43
44 450 from SRCC are not shown here. If an input parameter was shown to be
45
46 451 significant ($P < 0.05$) in one of the 13 output cell populations, it was considered
47
48 452 as an active input for the second stage input parameter design matrix.
49
50
51 453 Additionally, domain expert knowledge was employed to include additional
52
53 454 parameters, based on the biological significance, that were otherwise shown to
54
55 455 be non-significant. In all, based on the PRCC analysis performed on the outputs
56
57 456 obtained from the first stage simulations and domain expert knowledge, we
58
59 457 chose 23 input parameters as active inputs for the second stage (see Additional

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 file Fig. S4). Thus, PRCC screened inputs at significance level $p < 0.05$ and
459 inputs based on expert knowledge were selected as active inputs to be varied
460 for the second stage sampling plan.

461 Second stage sampling plan:

462 The number of active inputs obtained from the first stage analysis amounted
463 to 23 parameters out of the initial set of 38 parameters. We followed the design
464 described in [30] for the second stage and the number of design points amounted
465 to, $n_2 = 100\% * 5 * a$ where 'a' stands for the number of active inputs from the first
466 stage. This resulted into $n_2 = 23 * 5 = 115$ parameters combinations for the second
467 stage input parameter design matrix. Since outputs from both stages are to be
468 combined for second stage analysis, per [30], the design for the second stage
469 was chosen to build on top of $X^{(1)}$. The sampling phase design algorithm ensured
470 that the columns satisfied the properties (i) (uncorrelated design points) and (ii)
471 (between values 0 and 1) as listed in the previous section. We constructed the
472 115 x 38 (115 parameter setting and 38 parameters) design matrix for the second
473 stage that incorporated the 23 active inputs obtained from the PRCC screening
474 in the first stage output analysis. After combining the design points from both the
475 stages, the parameter design matrix X with space filling properties contained 267
476 (152 from the first stage and 115 from the second stage) design points.

477 Second stage analysis

478 We run the computer code for the hybrid model with the second stage input
479 parameter design matrix (with 115 (n_2) design points), for 20 (r) replicates, which
480 amounted to 115 x 20 (2300) runs. The outputs from the first stage (152 x 20

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

481 runs) and second stage (115 x 20 runs) were combined to provide the training
482 data to build a spatio-temporal metamodel. For the second stage analyses, we
483 utilized a metamodeling-based approach. Metamodels are surrogate models that
484 can be used as a substitute for the simulation model [32]. The use of metamodels
485 reduces the computational budget, cost of analysis, and are useful options in
486 cases when the simulation model is expensive to run (in our case 9-10 minutes
487 for 1 design point) [32]. The various metamodeling techniques used to build
488 surrogates for a computer model output include linear regression models, neural
489 networks, high dimensional model representation methods, Gaussian process
490 (GP) regression models, polynomial chaos expansion and more that are
491 discussed in length in [33, 34]. Amongst these, GPs are one of the most popular
492 emulators as it allows modeling of fairly complex functional forms. The GPs not
493 only provide prediction at a new point but also an estimate of the uncertainty in
494 that prediction [33]. A GP is a stochastic process for which any finite set of y-
495 variables has a joint multivariate Gaussian distribution [35] [33]. Suppose, $y_j(w)$,
496 the simulation response obtained on the j th simulation replicate, at a design point
497 $w = (X^T, t)^T \in \chi \times T$, it can be described as follows:

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

499 where $Y(w)$ represents the mean function of $y_j(w)$, the quantity of interest that
500 we intend to estimate at any design point w . The β_0 is a constant trend term and
501 is assumed to be unknown. The input parameter $X \in \chi \subset \mathbb{R}^d$ and the time $t \in$
502 $T \subset \mathbb{R}_+$; and X is independent of t . The $\varepsilon_j(w)$ are represents the sampling
503 variability inherent in a stochastic simulation, that are that are assumed to be

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

504 independent and identically distributed across the replications at any given
505 design point [36].

506 The term $M(w)$ represents a stationary Gaussian process with mean = 0 and
507 covariance between any points was modeled as the Gaussian covariance
508 defined in [37]. Thus, the covariance between any design points $w_a =$
509 $(X_a^T, t_a)^T$ and $w_b = (X_b^T, t_b)^T$ in the random field can be modeled as-

$$510 \quad Cov(M(w_a), M(w_b)) = \Gamma^2 \exp(-\sum_{r=1}^d \theta_r (X_{ar} - X_{br})^2) R(t_a - t_b; \gamma), \quad (2)$$

511 wherein, $\exp(-\sum_{r=1}^d \theta_r (X_{ar} - X_{br})^2)$ models the spatial correlation between two
512 input design points X_a and X_b in the input parameter space, whereas
513 $R(t_a - t_b; \gamma)$ also given by $\exp(-\sum_{r=1}^d \gamma_r (t_{ar} - t_{br})^2)$ models the temporal
514 correlation between time points t_a and t_b . The parameters θ and γ represents
515 the rate at which i) spatial correlation decreases as the points move farther in
516 space with the same time index, and ii) temporal correlation decreases as the
517 time points are farther apart in time at the same input vector, respectively. Both
518 the spatial correlation and temporal correlation are modeled using the Gaussian
519 covariance. The parameter Γ^2 can be interpreted as the variance of $M(w)$ for all
520 w . The input parameter design consists of $((w_a, n_i)_{i=1}^k)$ design points to run
521 independent simulations with replicates applied to each of the design points. Let,
522 $k \times 1$ denote a vector of sample averages of simulation responses given by $\bar{y} =$
523 $(\bar{y}(w_1), \bar{y}(w_2), \dots, \bar{y}(w_k))^T$, where in $\bar{y}(w_i)$ is the resulting estimate of
524 performance measure obtained at design point w_i and $\bar{\varepsilon}(w_i)$ is the sampling
525 variability inherent in a stochastic simulation (Ankenman, Nelson, & Staum,

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

526 2010). The equations associated with $\bar{y}(w_i)$ and $\bar{\varepsilon}(w_i)$ are described below in
527 equation (3):

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

530 Similar as in (Ankenman, Nelson, & Staum, 2010), shown below in equation (4),
531 let Σ_M be the $k \times k$ covariance matrix across all design points and let $\Sigma_M(w_o, \cdot)$ be
532 the $k \times 1$ vector, $(\text{Cov}[M(w_0, w_1)], \text{Cov}[M(w_0, w_2)], \dots, \text{Cov}[M(w_0, w_k)]^T$ that contains
533 spatial covariance between the k design points and a given prediction point w_o .
534 Also, let Σ_ε be the $k \times k$ covariance matrix of the vector of simulation errors
535 associated with the vector of point estimates \bar{y} , across all design points. As
536 described in [36], the best linear predictor $Y(w_o)$ that has the minimum mean
537 squared error (MSE) among all linear predictors at a given point $w_o =$
538 $(X_o^T, t_o)^T$ can be given by equation (4):

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

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

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

543 To implement the metamodeling approach as described above, the unknown
544 model parameters are estimated through maximizing the log-likelihood function.
545 The underlying standard assumption is that $(Y(w_o), \bar{y}^T)^T$ follows a multivariate
546 normal distribution, for e.g., see [36] and [38]. The function implemented in the

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

547 *mlegp* package in R [39] is used for the estimation of the parameters. Once the
548 parameters are estimated the prediction then follows equations (4) and (5).

549
550 Sensitivity index calculation

551
552 To determine the effect of input variables on the output, we employed the
553 variance decomposition method. These methods involve the decomposition of
554 the variance of the output as a sum of the variance produced by each input
555 parameter [35].

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

564
565
$$SI_1^{X_i} = \frac{V[E(Y|X_i)]}{V(Y)},$$

566
567 and

568
$$SI_{tot}^{X_i} = \frac{V[E(Y|X_{\sim i}])}{V(Y)}.$$

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

570 The Sobol' first order sensitivity index $SI_1^{X_i}$ measures the impact of one single
571 parameter on the model output, whereas the Sobol' total order index measures
572 the influence of X_i including all the interactions with other parameters. The First-
573 order indices were computed using the Sobol-Saltelli's method as described in
574 [40] [42] whereas, the total order indices were computed using Sobol-Jansen as
575 in [40, 43].

3. Results

3. 1 Hybrid model simulations produce similar immune response dynamics observed in previously published experimental data

581 We first aimed to simulate the findings observed in previous gut models [24]
582 to ensure that we obtained similar response dynamics from the hybrid ENISI
583 model of *H. pylori* infection. As in [24], to demonstrate that the gastric mucosa
584 harbors a system of macrophages that contribute to the outcome of *H. pylori*
585 infection, we created an *in-silico* Peroxisome proliferator-activated receptor
586 gamma (PPAR γ) macrophage-specific knockout (KO) model. PPAR γ is an
587 important transcription factor that controls the expression of genes that
588 contribute to the inflammatory response once this is initiated. To disrupt the
589 downregulation of pro-inflammatory responses, we simulated a PPAR γ KO
590 system in either macrophage or T cell populations and compared the response
591 to a wild-type system. In the model, we created three different macrophage
592 populations, comprised of, "resident" macrophage agents that mimic the

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

594 properties of the F4/80hi CD11b+ CD64+ CXCR1+ macrophages reported in
595 [24], monocyte-derived (infiltrating) and macrophage populations with regulatory
596 (M2, or alternatively activated) and pro-inflammatory function (M1 or classically
597 activated) (see *Table 3*).

598 We simulated an *in-silico H. pylori* infection by creating four groups, i) a control -
599 WT (representing a wild-type group), ii) CD4Cre (T cell specific PPAR γ KO-lacks
600 PPAR γ gene in all CD4 T cells), iii) LysMCre (Myeloid cell specific PPAR γ KO-
601 lacks PPAR γ gene in all macrophages) and clodronate group (simulating the
602 removal of macrophages by chemical depletion via clodronate treatment). To
603 simulate the CD4Cre group, the probabilities of a naive T cell transitioning to an
604 iTreg cell ($p_{nTtoiTreg}$) and Th17 cell differentiating to iTreg ($p_{Th17toiTreg}$)
605 were reduced to 5% and 10% of the control value, respectively (refer to *Table*
606 *S1*). As described in [23], to simulate the LysMCre experimental conditions, the
607 probabilities of i) a monocyte transitioning to a regulatory macrophage
608 ($p_{Mregdiff}$) and ii) immature dendritic cells switching to tolerogenic dendritic
609 cells ($p_{iDCtoiDC}$) were reduced approximately to 60% and 30% of the control
610 value, respectively (refer to *Table S1*). A complete set of parameter for each of
611 the biological KOs are included as separate columns in *Table S1*. Lastly, the
612 removal of macrophages by clodronate were simulated by decreasing the initial
613 numbers of the macrophage population including the resident macrophages. The
614 rationale to include the clodronate group (macrophage removal) was to evaluate
615 if depletion of phagocytic cells (terminology with respect to model, *i.e.*,
616 monocytes, resident, monocyte-derived macrophages and inflammatory
617 macrophages) would affect *H. pylori* colonization levels, as we have previously

1
2
3
4 618 reported in an *in vivo* model [24]. Further, to simulate the myeloid cell PPAR γ KO
5
6 619 system, the initial population of resident macrophages were also reduced.

7
8
9 620 All the groups were initialized with equal loads of *H. pylori* agents. Ten
10
11 621 replicates of the simulations were performed for each of the input parameter
12
13 622 settings specific to each group. The outputs were averaged, and standard error
14
15 623 of the means were plotted as ribbons (shaded regions) across the graphs. After
16
17 624 running the ten replicates of the time series *in-silico* simulation, the hybrid model
18
19 625 showed significantly ($p < 0.05$) higher levels of *H. pylori* in the WT and CD4Cre
20
21 626 groups as compared to LysMCre KO and macrophage-depleted groups (*Fig. 3*,
22
23 627 panel a and d).
24
25
26
27 628
28
29 629
30

31 630 ***Fig 3. Time course simulations representing the immune response during***
32
33 631 ***Helicobacter pylori infection.***

34
35 632 *The upper half of the plot in both the panels shows the dynamics of the*
36
37 633 *population cells*
38
39 634 *over time representing the number of cells (y-axis) versus time (x-axis) in a WT*
40
41 635 *(black), CD4Cre (green), clodronate (red) and LysMCre (blue) simulated in-silico*
42
43 636 *groups during H. pylori infection. The cell populations include - a) H. pylori; b)*
44
45 637 *the resident macrophages and, c) monocyte-derived macrophages in the lamina*
46
47 638 *propria compartment. The figures in the lower half (d-f) of both the panels, show*
48
49 639 *the results for statistical comparison between the groups using ANOVA with the*
50
51 640 *post-hoc analysis. The letters 'a', 'ab' and 'b' represent statistically significant*
52
53 641 *differences ($P < 0.05$) between the groups obtained after running the Tukey's*
54
55 642 *Honestly Significant Difference. A side by side comparison with the bacterial load*
56
57
58
59
60
61
62
63
64
65

1
2
3
4 643 and macrophage population as observed in the mouse model of *H. pylori*
5
6 644 infection are also included.
7
8
9 645

10
11 646 In addition to the increase in *H. pylori*, WT and CD4Cre *in-silico* experimental
12
13 647 groups had a higher resident as well as monocyte-derived regulatory
14
15 648 macrophages as compared to clodronate (macrophage depleted) and LysMCre
16
17 649 groups (Fig. 3b-c, e-f). The results in the mouse model indicated that between
18
19
20 650 weeks 2 and 3 post-infection a decrease in bacterial burden in the stomach of
21
22 651 LysMcre mice was observed as shown in Fig. 1A of Viladomiu, Bassaganya-
23
24 652 Riera et al. 2017. The decrease in bacterial burden led to a significant and
25
26 653 sustained lower colonization levels when compared to WT and CD4Cre. Similar
27
28
29 654 to the results observed in the mouse model, we observed a significant decrease
30
31 655 (Fig. 3a, d) in the bacterial burden in the simulated LysMcre group as compared
32
33 656 to the simulated WT and CD4cre groups. Furthermore, the results from the
34
35 657 mouse model indicated that a significant increase in numbers of F4/80hiCD11b+
36
37 658 CD64+ CX3CR1+ cells (here referred to as resident macrophages in this paper),
38
39
40 659 was observed in WT mice in comparison with LysMcre mice as shown in Fig. 2A,
41
42 660 2E of Viladomiu, Bassaganya-Riera et al. 2017. These cells accumulated in the
43
44 661 stomach mucosa starting on day 14 post-infection in the WT mice but not in the
45
46 662 LysMcre mice. We observed a similar increase (Fig. 3b,e and Fig. 3c,f) in the
47
48
49 663 number of resident macrophages as well as monocyte derived macrophages in
50
51 664 the simulated WT groups in comparison to the simulated LysMcre group. We
52
53 665 estimated the parameter values to fit the data obtained from the mouse model of
54
55 666 *H. pylori* infection. Thus, the observations were qualitatively similar to the
56
57
58 667 findings in [24], where the stomach of WT mice was enriched in a population of
59
60 668 F4/80+CD11b+CD64+ myeloid cells, compared to LysMCre mice.
61
62
63
64
65

1
2
3
4 669 Overall, with the results in Fig. 3, we showed the ability of the hybrid model
5
6 670 to replicate the experimental results in [24], and this preliminary data was used
7
8
9 671 as a base calibration setting for SA and other *in-silico* findings.

10 11 12 672 **3.2 Partial correlation coefficient analysis screened the influential** 13 14 673 **parameters**

15
16
17 674
18
19 675 To reduce the computational complexity of varying an input parameter space
20
21 676 of 38 parameters, we divided the SA process in two stages. For first-stage
22
23 677 analysis, we utilized the PRCC regression-based SA method to screen the
24
25 678 influential inputs and used it for the second stage design of the experiments (refer
26
27 679 Methods 2.2). Using PRCC, we determined the impact of the input parameters
28
29 680 on the output cell populations in the model. The parameters with significant
30
31 681 correlation with *H. pylori* in the gastric lamina propria compartment and resident
32
33 682 macrophages are shown in Fig. 4, along with their PRCC values. The bars in
34
35 683 blue, highlight the parameters that are significantly different than 0, at $P < 0.05$
36
37 684 compared to grey bars which are not significant. It is important to note that at this
38
39 685 stage the analysis using PRCC was non-temporal.

40
41
42 686 The SA from first stage results showed that the epithelial damage due to
43
44 687 infectious bacteria (*epiinfbctdam*) with a coefficient value of (~0.2), was positively
45
46 688 correlated with the colonization of *H. pylori* in the lamina propria compartment,
47
48 689 indicating the important role of epithelial cell damage during the course of
49
50 690 infection, similar to our findings obtained in [44]. Another parameter included the
51
52 691 probability of the release of IL-6 (*IL6*) with a coefficient value within the range
53
54 692 (0.3-0.4).
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 693 Next, the epithelial cell damage parameters ($epiinf\bctdam = (0.2-0.3)$,
5
6 694 $epiTh17dam = 0-0.2$) were shown to have positive influence on the resident
7
8
9 695 macrophage cells whereas, the T cell type transition parameters ($p_{iTregtoTh17}$
10
11 696 $= (0.3 - 0.4)$ and $p_{Th17toiTreg} = (0.1 - 0.2)$) showed a negative impact on the
12
13 697 resident macrophages. Similarly, we performed the PRCC analysis for all the cell
14
15 698 populations under consideration during the infection (not shown).
16
17
18 699
19
20 700

21
22 **Fig 4. Bar plots for the partial rank correlation coefficients.**

23
24 702 *The magnitude of the bar-plot indicates the value of the partial rank correlation*
25
26 703 *coefficient. The blue bar indicated the input parameters shown to be significantly*
27
28
29 704 *different than 0, at $P < 0.05$ as influential whereas the grey bars indicate the non-*
30
31 705 *influential parameters on a) *H. pylori* and b) resident macrophages, in the lamina*
32
33 706 *propria compartment.*
34

35 707
36
37
38 708 The significant parameters (marked in blue bars) obtained from the SA of the
39
40 709 output from first stage design of experiments (152 parameter settings with 20
41
42 710 replicates, refer Methods 2.2), were selected to be varied for the second stage
43
44 711 design. All the selected inputs are shown in Additional file Fig. S4. In all, we
45
46
47 712 obtained 23 active inputs from the first stage.
48
49

50 **3.3 Metamodel based spatio-temporal sensitivity analysis**

51
52
53 714 The outputs obtained after running the first (152 x 20 runs) and second (115
54
55 715 x 20 runs) stage simulations, wherein x20 denotes the 20 replicates, were
56
57 716 combined to be used as a training dataset. The combined output was utilized to
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

717 build a Gaussian process based spatiotemporal metamodel (refer Methods 2.2),
718 using *mlegp* package in R [39].

719
720 The outputs from the training dataset were sub-divided into 6 datasets,
721 corresponding to six time periods (Days 1-14, 15-21, 22-30, 31-42, 43-90, 91-
722 201) and averaged across these periods. The sub-division of output across the
723 time periods, aided the temporal analysis over the initiation (Day 1-14), peak of
724 infection (Days 15-30) and chronic phase (post Day 31) stages as in [24]. We
725 then fit a Gaussian process model (with nugget) and evaluated the performance
726 of the fitting of the metamodel for *H. pylori*, resident macrophages, and
727 monocyte-derived macrophages in *lamina propria* compartment, and tolerogenic
728 DC in the gastric lymph node, using the diagnostic plots (see figures in Additional
729 file, *Fig. S5*). After fitting the models, we performed variance based global SA by
730 computing the Sobol' total order and first order sensitivity index (refer Methods
731 2.2). The estimates of the Sobol' total order indices for the input parameters
732 calculated over the six time periods are shown in *Fig. 5* (a-d).

733
734
735 ***Fig 5. Heat-maps of Sobol' total order index for the input parameters across***
736 ***different output populations.***

737 *The values in the heat-map indicate the Sobol' total order sensitivity index*
738 *obtained from the metamodel, for the 38 input parameters with respect to the cell*
739 *populations. The values with darker color indicate a stronger influence on the cell*
740 *population as compared to the ones with lighter shade that indicate non-*
741 *influential parameters for the cell populations - a) *H. pylori*, b) monocyte-derived*
742 *macrophages, c) resident macrophages, in the lamina propria compartment and*

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

d) tolerogenic DCs, in the gastric lymph node compartment. The indexes are calculated over six time points ranging across the three stages of infection, including initiation (Day 1-14), peak (Days 15-42) and recovery stages (Days 43-201).

As shown in Fig. 5a, the metamodel based global SA showed that the input parameters, epithelial cell proliferation (*EpiProlifer*) and epithelial cell death (*EpicellDeath*) had the strongest impact on the population of *H. pylori* in lamina propria compartment. As time progressed from initiation of the infection (Days 1-14), through peak (Days 15-30), the epithelial cell proliferation had a continued impact on the colonization of *H. pylori*. Next, the influence of the probability of epithelial cell death decreased over the course of infection. Further, Fig. 5b highlighted the impact of epithelial cell proliferation (*EpiProlifer*) and epithelial cell death (*EpicellDeath*) on the monocyte-derived macrophages.

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

Finally, for the tolerogenic DCs in Fig. 5d, we observed that the epithelial cell death (*EpicellDeath*) seemed to have an impact. Another parameter that stands for the probability of naive T cell transitioning to iTreg cell (*nTtoiTreg*) was shown

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

769 to have an impact on the tolerogenic dendritic cells. Tolerogenic dendritic cells
770 are involved in the rule that transitions the naive T cells to iTreg cells in the gastric
771 lymph node, and the stronger impact of the *nTtoiTreg* during the initiation and
772 peak stages of the infection highlights the role of the tolerogenic dendritic cells
773 during the course of infection.

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

777 **3.4 Effect of different ranges of epithelial cell proliferation**

778 An interesting prediction derived from the metamodel based global SA is that
779 epithelial cell proliferation is one of the parameters that has a strong impact on
780 the size of *H. pylori* population. The biological hypothesis derived from this
781 prediction is that the epithelial cell proliferation is responsible for the higher
782 colonization of *H. pylori*. Prior to conducting any experimental studies, we wanted
783 to explore the hypothesis using our hybrid computer model *in silico* and study
784 the model outputs obtained after we changed the epithelial cell proliferation
785 parameter. Thus, we varied the epithelial cell proliferation parameter across
786 different ranges (0.1-0.9, with 0.6 being the value for baseline conditions) and
787 ran the simulations using the hybrid model and studied its effect on the different
788 output cell population (obtained after running the simulations). These outputs
789 were the ones obtained after running the simulation using the hybrid computer
790 model, as we varied the epithelial cell proliferation parameter. We analyzed the
791 outputs from the hybrid computer model and interestingly, observed that upon
792 decreasing the *Epiprolifer* from a range of values 0.9-0.1, the output cell
793 populations with regulatory function, namely regulatory macrophages and

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

794 tolerogenic dendritic cells were found to vary. We observed a decreasing effect
795 (Fig. 6a-d) on *H. pylori*, monocyte-derived macrophages, resident macrophages
796 in the lamina propria compartment and tolerogenic dendritic cells in gastric lymph
797 node. Overall, these cell populations varied due to the variation in the epithelial
798 cell proliferation parameter.

799 For clarification, such connection was not embedded in the mechanisms
800 included in Table 1 but it represents an emergent behavior from the simulations
801 predicting the involvement of regulatory and tolerogenic dendritic cells in the
802 mechanisms of immunoregulation during *H. pylori* infection. Finally, the
803 simulations targeting the epithelial cell proliferation resulted in changes in
804 regulatory and tolerogenic dendritic cell populations. This shows that the
805 simulations indirectly targeted the regulatory and tolerogenic dendritic cell
806 population. Thus, we hypothesize that epithelial cell proliferation might be
807 responsible for the higher colonization of *H. pylori* through an immunoregulatory
808 mechanism that involves regulatory macrophages and tolerogenic cells. This is
809 in line with our own conclusions drawn from a previous paper [24] where we
810 show that the presence of cells with regulatory phenotype favor higher levels of
811 *H. pylori* colonization. The results from the sensitivity analysis presented in this
812 paper suggest that epithelial proliferation might be a crucial part of the
813 mechanisms by which these regulatory responses are induced and that there is
814 a link between these parameters. The exact biological process however cannot
815 be inferred from the current model and it will be investigated in follow-up in vivo
816 studies.

817
818
819

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

820 **Fig 6. In silico study of the effect of epithelial cell proliferation parameter**
821 **on the cell populations.**

822 *The plots show the effect of varying epithelial cell proliferation (p_Epiprolifer)*
823 *parameter (with values 0.1, 0.5, 0.6(WT), and 0.9) on the output cell population*
824 *of a) H. pylori, b) tolerogenic dendritic cells, c) resident macrophages and d)*
825 *monocyte-derived macrophages. The parameter has a decreasing effect on the*
826 *cellular populations under consideration, wherein a decrease in the parameter*
827 *value, decreases the abundance of the cells over time. The lower half of the*
828 *figures (a-d), show the results for statistical comparison between the groups*
829 *using ANOVA with the post-hoc analysis. The letters 'a', 'b', 'c', and 'bc' represent*
830 *statistically significant differences (P<0.05) between the groups obtained after*
831 *running the Tukey's Honestly Significant Difference.*

832
833 The *in silico* findings suggested the involvement of regulatory macrophages
834 (both resident as well as monocyte-derived) and tolerogenic DC on the
835 colonization of *H. pylori* in the gastric lamina propria. This highlighted and
836 validated the role of epithelial cell proliferation as one of the main factor affecting
837 *H. pylori* levels in the gastric niche.

838

839 4. Discussion

840 *H. pylori* is the dominant indigenous bacterium of the gastric microbiota. In the
841 majority of individuals, *H. pylori* colonizes the stomach without causing adverse
842 effects, with little to no activation of inflammatory pathways. However, certain
843 members of the population lose immune tolerance to the bacterium thereby

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

844 contributing to the development of chronic gastric diseases. The immunological
845 mechanisms underlying its ability to persist in a harsh acidic gastric environment and
846 its dual role as a pathogen and beneficial organism remain unknown. A subset of
847 macrophages helps create a regulatory microenvironment that promotes the chronic
848 colonization of *H. pylori* [24]. However, the immune regulatory mechanisms are
849 incompletely understood. Computational models of the immune system featuring
850 immune responses are powerful tools for testing the different ‘what-if’ scenarios.
851 Multiscale models of the immune response are attractive in terms of modeling the
852 responses at different spatiotemporal scales [45].

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

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

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

875 Our group also showed another mechanism used by *H. pylori* to create a gut
876 microenvironment that involved the induction of IL-10-driven regulatory mechanism
877 mediated by CD11b⁺F4/80^{hi}CD64⁺CX₃CR1⁺ mononuclear phagocytes, which
878 facilitated bacterial colonization [24]. Additionally, in this paper, we reported that
879 regulatory macrophages were involved in the process of colonization with *H. pylori*
880 when we varied the epithelial cell proliferation parameter *in-silico*. Zhang et al.,
881 demonstrated that *H. pylori* directed active tolerogenic programming of DCs that
882 favored chronic bacterial colonization, by altering the balance of Th17/Treg cells [49].
883 Rizzuti, Ang et al., demonstrated *H. pylori*-mediated IL-10 release caused the
884 activation of signal transducer and activator of transcription 3 (STAT3) in DC. This
885 activation of STAT3 via IL-10 release was shown to induce the production of
886 tolerogenic DC phenotype [50]. The findings from this paper also indicated the
887 involvement of tolerogenic DCs in affecting the mucosal levels of *H. pylori*.
888 Therefore, the literature combined with the results from this study, collectively
889 suggest that during *H. pylori* infection, the epithelial cell favors the colonization of *H.*
890 *pylori* by creating a regulatory microenvironment. This process is mediated by the
891 regulatory macrophages and tolerogenic programming of DC. Based on the results
892 from this paper and findings from the literature, this leads us to propose that the
893 induction of IL-10 by the regulatory macrophages is potentially involved in directing
894 the tolerogenic programming of DC. All experimental evidence combined with our

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

895 model prediction suggest the action of an underlying biological mechanism that links
896 the presence of *H. pylori* in the gastric mucosa with changes in the rates of epithelial
897 cell proliferation which ultimately affects the levels of colonization. Our prediction
898 points towards a link between epithelial cell proliferation and the action of tolerogenic
899 dendritic cells and regulatory macrophages. The exact cellular mechanism induced
900 during this process however cannot be inferred from the current model and it will be
901 investigated in follow-up in vivo studies.

902
903 At its current stage, the hybrid ENISI model reproduces the overall immune
904 system dynamics observed during an *H. pylori* infection. The parameters of
905 calibrated ODEs were kept unchanged, whereas the ABM parameters were
906 calibrated by qualitatively matching the patterns of the output simulations as
907 observed in an *in vivo* model of *H. pylori* infection [24]. For ABM, its calibration and
908 validation remain the major key issues, discussed elsewhere [21] [51] [52]. Further,
909 developing targeted methods of SA have been identified as an important challenge
910 in the field [21, 53, 54]. In this paper, we highlighted the use of SA methods with a
911 two-stage global SA framework comprised of first, screening the input parameters
912 (using PRCC) and second, building of a surrogate model (using GP) of the hybrid
913 model, to understand the emergent behavior of the represented system. It is
914 important to note that each SA method known, has its own merits and produces
915 useful information however none provide a complete picture of the emergent model
916 behavior [21]. First, we employed PRCC methods as the initial step in our two staged
917 SA that aided the screening of active inputs and reduced the parameter space. The
918 choice of PRCC was advantageous and justified by the low computational cost and
919 low complexity in the computation of the coefficients. Another advantage of the
920 regression-based PRCC method is that the complex output from our 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

921 was condensed into a descriptive relationship that can be described by statistical
922 measures such as R^2 [21]. As described in [21] the results from PRCC are good
923 descriptors of the outputs produced if the regression function constitutes a good fit
924 to the output [21]. However, if the function does not yield a good fit, the regression-
925 based SA are proven to be useful in screening the influential parameters for further
926 analysis [21], as described in our analysis.

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

942 Some limitations of the model include implementation through a two-dimensional
943 grid system and including all cells of the same size. Although we parallelize the
944 computation of the hybrid model output, the large number of simulations required for
945 the global SA compensates for the benefits of parallelization. To improve the
946 calibration process and overall usability of the model, the data required for model

1
2
3
4 947 calibration would include tissue biopsies from people infected with *H. pylori* that can
5
6 948 be used to quantify the cells and take into account their spatial arrangement. The
7
8 949 current version is also limited in terms of the interactions that are based on epithelial
9
10 950 cells and DC as they are strictly rule-based. The building of ODE models for these
11
12 951 cells and integrating them with the ABM model will help capture the dynamics of
13
14 952 epithelial cells and DC more in-depth. Overall the immunoregulatory mechanisms
15
16 953 underlying the chronic colonization of *H. pylori* and the predictive capacity of the
17
18 954 model can be further improved by incorporating cell-specific models for epithelial
19
20 955 cells and DC.
21
22

23
24 956 In summary, a high-resolution, hybrid, multiscale spatiotemporal stochastic
25
26 957 model of *H. pylori* infection was built and global SA was performed. The results from
27
28 958 the global SA highlight the key role played by epithelial cells in affecting the levels of
29
30 959 *H. pylori* colonization. The *in-silico* validation of varying the epithelial cell proliferation
31
32 960 parameter demonstrated the involvement of regulatory macrophages and the
33
34 961 tolerogenic DC. The next steps aimed to enrich the model will involve the validation
35
36 962 of the findings *in vivo* to study the underlying mechanisms involved in the successful
37
38 963 immune evasion by *H. pylori*. The computational modeling predictions will be further
39
40 964 validated experimentally and clinically.
41
42
43

44 965

46 966 **5. Potential Implications**

48
49 967 The computational model of the gut contains high-resolution information
50
51 968 processing representations of immune responses that are generalizable for other
52
53 969 infectious and autoimmune diseases. Complex diseases such as autoimmune
54
55 970 disorders, infectious diseases, and cancer all require integration of the multiscale
56
57 971 level data, information and knowledge, ranging from genes, proteins, cells, tissue to
58
59 972 organ level. The ENISI model of the gut presented here can be generalized to other
60
61
62

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

973 diseases by implementing the agents and rules specific to that disease, plus
 974 recalibrating the model based on data that are specific to the new indication. Since
 975 ABMs have modular architectures, an addition of new agent-types and modification
 976 of rules can be done without restructuring the entire simulation setup [19]. The use
 977 of ABM in such hybrid models not only facilitates the implementation of already
 978 known mechanisms but also helps validate and predict any unforeseen new
 979 mechanisms using data analytics methods such as global SA to analyze emerging
 980 behaviors at the systems level. The finer details regarding intracellular and
 981 intercellular interactions that contribute towards the nonlinear and complex behavior
 982 of the gut can also be studied by integrating the intracellular ODE models as
 983 implemented here.

984
985 **Tables**

Name of Agent	Agent Type	Rules
<i>Helicobacter pylori</i>	<i>H. pylori</i>	<ul style="list-style-type: none"> - Moves across the epithelial cell border if near damaged epithelial layer - Proliferates in the lumen and lamina propria - Dies (removed from the simulation) in lamina propria and in the lumen due to the damage of epithelial cells by Th1 or Th17 cells
Macrophages	Monocyte	<ul style="list-style-type: none"> - Proliferates in presence the of effector dendritic cells or damaged epithelial cells - Proliferates in the lamina propria - Differentiates to regulatory macrophage in based on the output from the Macrophage ODE - Differentiates to inflammatory macrophages in presence of IFN-γ - Dies naturally (removed from the model)
	Resident	<ul style="list-style-type: none"> - Proliferates in the presence of <i>H. pylori</i> - Secretes IL10 - Dies naturally - Dies due to Th1 and Tr cells
	Regulatory	<ul style="list-style-type: none"> - Proliferates and removes bacteria - Dies - Secretes IL10
	Inflammatory	<ul style="list-style-type: none"> -Proliferates in the presence of damaged epithelial cell

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

		-Dies naturally
Dendritics	Immature	-Moves from lamina propria to epithelium compartment and from the epithelium to the lamina propria - Differentiates to tolerogenic dendritic cell in the presence of tolerogenic bacteria, both in epithelium and lamina propria - Differentiates to effector dendritic cell in the presence of <i>H. pylori</i> - Proliferates in lamina propria and gastric lymph node - Dies naturally
	Effector	- Moves from lamina propria to gastric lymph node - Moves form epithelium to lamina propria - Secretes IL6 and IL12 - Dies naturally
	Tolerogenic	- Moves from lamina propria to gastric lymph node - Moves from epithelium to lamina propria - Secretes TGF- β - Dies naturally
T cells	Naïve	In the presence of effector dendritic cells: - Differentiates to Th1 in the presence of IFN- γ or IL12 - Differentiates to Th17 in the presences of IL6 or TGF- β In the presence of tolerogenic dendritic cells: - Differentiates to iTreg in the presence of TGF- β - Differentiates to Tr in the presences of IL10 - Dies naturally
	Th1	- Secretes IFN- γ - Moves from gastric lymph node to lamina propria - Proliferates in lamina propria and gastric lymph node - Dies naturally
	Th17	- Secretes IL17 - In the presence of tolerogenic dendritic cell, transition to iTreg cells - Moves from gastric lymph node to lamina propria - Proliferates in lamina propria and gastric lymph node - Dies naturally
	iTreg	- Secretes IL10 - In the presence of tolerogenic dendritic cell, transition to iTreg cells - Moves from gastric lymph node to lamina propria - Proliferates in lamina propria and gastric lymph node - Dies naturally
	Tr	- Secretes IL10 - Dies naturally - Proliferates in the lamina propria
Epithelial	Healthy	-Damaged due to infectious bacteria -Damaged due to Th1 and Th17 cells -Proliferates -Secretes IL6 and IL12 -Dies naturally
	Damaged	-Transitions to healthy state in the presence of IL10

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

		-Dies naturally
Bacteria	Infectious	- Dies due to Th1 or Th17 or inflammatory macrophages or damaged epithelial cells - Dies naturally - Proliferates in the lamina propria
	Tolerogenic	- Moves from lumen to the epithelium in the presence of damaged epithelial cells - Becomes infectious if moves in the lamina propria compartment - Proliferates in lumen and lamina propria - Dies naturally

986

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

988

989

990 **Additional Files**

991 File S1

992 Fig S1

993 Table S1

994 Fig S2

995 Fig S3

996 Fig S4

997 Fig S5

998

999 **File S1** – The detailed instruction to Install ENISI MSM (Step I), Run a simulation
1000 (Step II) and Conduct Sensitivity Analysis (Step III) are described.

1001

1002 **Fig S1. Design implementation of the hybrid multiscale model used to**
1003 **simulate Helicobacter pylori infection**

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

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. Time screenshots of a Helicobacter pylori infection modeled in a 30 mm (length) x 10 mm (width) two-dimensional grid. The thickness of the compartment is shown on the y-axis, such that: lumen spans (0 to 2) units, epithelium spans (2 to 3) units, lamina propria spans (3 to 8) units and gastric lymph node across (8-10) units on the scale. Two-dimensional distribution of different cell subsets over the time steps (ticks) 2, 4 (top panels), 5 and 6 (bottom panels) are shown. The insets in each image shows a zoomed in portion of the respective grids across the time steps 2, 4, 5 and 6. The agents represented in the screenshots below are only for visual representation and do not represent the actual size of the biological cells.

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

Fig S4. The active and inactive inputs selected from the stage 1 analysis. The rows represent the input parameters and columns represent the output cell populations. The green boxes highlight the 'active' input parameters (row) that are shown to have a significant influence (calculated based on the results

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

obtained from partial correlation coefficient analysis), on an output cell (columns) under consideration.

Fig S5. Diagnostic and residual plots obtained for the Gaussian processes fitted metamodels

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

Data and materials

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

The data sets and files supporting the results of this article are available in the ENISI-MSM
GitHub repository, RRID: SCR_016918 <https://github.com/NIMML/ENISI-MSM>.

Availability of source code and requirements

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

Declarations

List of abbreviations

- ABM – Agent based model
- DC – Dendritic cells
- ENISI MSM – Enteric Immunity Simulator Multi-scale Modeling
- GLN – gastric lymph node
- GP - Gaussian process
- H. pylori* – *Helicobacter pylori*
- HPC – High performance computing
- LP – Lamina propria
- ODE – Ordinary Differential Equation
- PDE – Partial Differential Equation
- SA – Sensitivity analysis

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

1077 PRCC - Partial rank correlation coefficient

1078 **Consent for publication**

1079 Not applicable.

1080 **Competing interests**

1081 The author(s) declare that they have no competing interests.

1082 **Funding**

1083 This work was supported by the Defense Threat Reduction Agency (DTRA) grant
1084 HDTRA1-18-1-0008 to JBR and RH and funds from the Nutritional Immunology and
1085 Molecular Medicine Laboratory (www.nimml.org). The funding body had no role in
1086 the design of the study, data collection, analysis, interpretation of data and writing of
1087 the manuscript.

1088 **Authors' contributions**

1089 MV, RH and JBR formulated the model, implemented, performed the simulations,
1090 analyzed model-generate outputs, made the figures and wrote the manuscript. MV,
1091 AL, JBR, RH, and SH formulated the model. SH, AL and VA implemented the code
1092 architecture and benchmarked the parallel version of the hybrid model. XC and MV
1093 wrote the codes for global sensitivity analysis and generated the design matrices.
1094 NTJ generated macrophage and *H. pylori* experimental data. JBR, VA, and RH
1095 supervised the project. JBR and RH edited the manuscript. JBR, AL, NTJ, SH, VA,
1096 XC and RH participated in discussions on the model and results. All authors provided
1097 critical feedback on the project.

1098 **Acknowledgements**

1099 Not applicable.

1100

1101 **References**

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

1. Carbo A, Bassaganya-Riera J, Pedragosa M, Viladomiu M, Marathe M, Eubank S, et al. Predictive computational modeling of the mucosal immune responses during *Helicobacter pylori* infection. *PLoS One*. 2013;8(9):e73365. doi: 10.1371/journal.pone.0073365.
2. Leber A, Bassaganya-Riera J, Tubau-Juni N, Zoccoli-Rodriguez V, Viladomiu M, Abedi V, et al. Modeling the Role of Lanthionine Synthetase C-Like 2 (LANCL2) in the Modulation of Immune Responses to *Helicobacter pylori* Infection. *PloS one*. 2016;11(12):e0167440. Epub 2016/12/10. doi: 10.1371/journal.pone.0167440. PubMed PMID: 27936058; PubMed Central PMCID: PMC5147901.
3. Leber A, Viladomiu M, Hontecillas R, Abedi V, Philipson C, Hoops S, et al. Systems Modeling of Interactions between Mucosal Immunity and the Gut Microbiome during *Clostridium difficile* Infection. *PloS one*. 2015;10(7):e0134849. Epub 2015/08/01. doi: 10.1371/journal.pone.0134849. PubMed PMID: 26230099; PubMed Central PMCID: PMC4521955.
4. Verma M, Erwin S, Abedi V, Hontecillas R, Hoops S, Leber A, et al. Modeling the Mechanisms by Which HIV-Associated Immunosuppression Influences HPV Persistence at the Oral Mucosa. *PloS one*. 2017;12(1):e0168133. Epub 2017/01/07. doi: 10.1371/journal.pone.0168133. PubMed PMID: 28060843; PubMed Central PMCID: PMC5218576.
5. Qomlaqi M, Bahrami F, Ajami M, Hajati J. An extended mathematical model of tumor growth and its interaction with the immune system, to be used for developing an optimized immunotherapy treatment protocol. *Mathematical biosciences*. 2017;292:1-9. Epub 2017/07/18. doi: 10.1016/j.mbs.2017.07.006. PubMed PMID: 28713023.
6. Vodovotz Y, Xia A, Read EL, Bassaganya-Riera J, Hafler DA, Sontag E, et al. Solving Immunology? *Trends in immunology*. 2017;38(2):116-27. doi: 10.1016/j.it.2016.11.006. PubMed PMID: PMC5695553.
7. Kusters JG, van Vliet AH, Kuipers EJ. Pathogenesis of *Helicobacter pylori* infection. *Clinical microbiology reviews*. 2006;19(3):449-90. Epub 2006/07/19. doi: 10.1128/cmr.00054-05. PubMed PMID: 16847081; PubMed Central PMCID: PMC1539101.
8. Mane S, Dominguez-Bello M, Blaser M, Sobral B, Hontecillas R, Skoneczka J, et al. Host-interactive genes in Amerindian *Helicobacter pylori* diverge from their Old World homologs and mediate inflammatory responses. *Journal of bacteriology*. 2010;192(12):3078-92.
9. Cover TL, Blaser MJ. *Helicobacter pylori* in health and disease. *Gastroenterology*. 2009;136(6):1863-73. doi: 10.1053/j.gastro.2009.01.073. PubMed PMID: PMC3644425.
10. Bassaganya-Riera J, Dominguez-Bello MG, Kronsteiner B, Carbo A, Lu P, Viladomiu M, et al. *Helicobacter pylori* colonization ameliorates glucose homeostasis in mice through a PPAR γ -dependent mechanism. *PLoS One*. 2012;7(11):e50069. doi: 10.1371/journal.pone.0050069.
11. Oertli M, Sundquist M, Hitzler I, Engler DB, Arnold IC, Reuter S, et al. DC-derived IL-18 drives Treg differentiation, murine *Helicobacter pylori*-specific immune tolerance, and asthma protection. *The Journal of clinical investigation*. 2012;122(3):1082-96. Epub 2012/02/07. doi: 10.1172/jci61029. PubMed PMID: 22307326; PubMed Central PMCID: PMC3287234.
12. Mei Y, Abedi V, Carbo A, Zhang X, Lu P, Philipson C, et al. Multiscale modeling of mucosal immune responses. *BMC Bioinformatics*. 2015;16 Suppl 12:S2. Epub 2015/09/04. doi: 10.1186/1471-2105-16-s12-s2. PubMed PMID: 26329787; PubMed Central PMCID: PMC4705510.
13. Gong C, Milberg O, Wang B, Vicini P, Narwal R, Roskos L, et al. A computational multiscale agent-based model for simulating spatio-temporal tumour immune response to PD1 and PDL1 inhibition. *Journal of the Royal Society, Interface*. 2017;14(134). Epub 2017/09/22. doi: 10.1098/rsif.2017.0320. PubMed PMID: 28931635; PubMed Central PMCID: PMC5636269.
14. Wang Z, Birch CM, Sagotsky J, Deisboeck TS. Cross-scale, cross-pathway evaluation using an agent-based non-small cell lung cancer model. *Bioinformatics (Oxford, England)*.

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

2009;25(18):2389-96. Epub 2009/07/07. doi: 10.1093/bioinformatics/btp416. PubMed PMID: 19578172; PubMed Central PMCID: PMCPMC2735669.

15. Marino S, El-Kebir M, Kirschner D. A hybrid multi-compartment model of granuloma formation and T cell priming in tuberculosis. *Journal of theoretical biology*. 2011;280(1):50-62. Epub 2011/03/30. doi: 10.1016/j.jtbi.2011.03.022. PubMed PMID: 21443879; PubMed Central PMCID: PMCPMC3740747.

16. Solovyev A, Mi Q, Tzen YT, Brienza D, Vodovotz Y. 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 computational biology*. 2013;9(5):e1003070. Epub 2013/05/23. doi: 10.1371/journal.pcbi.1003070. PubMed PMID: 23696726; PubMed Central PMCID: PMCPMC3656105.

17. Cappuccio A, Tieri P, Castiglione F. Multiscale modelling in immunology: a review. *Briefings in bioinformatics*. 2016;17(3):408-18. Epub 2015/03/27. doi: 10.1093/bib/bbv012. PubMed PMID: 25810307.

18. Bassaganya-Riera J. *Computational Immunology: Models and Tools*: Academic Press; 2015.

19. An G, Mi Q, Dutta-Moscato J, Vodovotz Y. Agent-based models in translational systems biology. *Wiley interdisciplinary reviews Systems biology and medicine*. 2009;1(2):159-71. Epub 2010/09/14. doi: 10.1002/wsbm.45. PubMed PMID: 20835989; PubMed Central PMCID: PMCPMC3640333.

20. Abedi V, Hontecillas R, Hoops S, Liles N, Carbo A, Lu P, et al., editors. ENISI multiscale modeling of mucosal immune responses driven by high performance computing. 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2015 9-12 Nov. 2015.

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

22. Ligmann-Zielinska A, Kramer DB, Spence Cheruvelil K, Soranno PA. Using uncertainty and sensitivity analyses in socioecological agent-based models to improve their analytical performance and policy relevance. *PLoS One*. 2014;9(10):e109779. Epub 2014/10/24. doi: 10.1371/journal.pone.0109779. PubMed PMID: 25340764; PubMed Central PMCID: PMCPMC4207681.

23. Carbo A, Hontecillas R, Kronsteiner B, Viladomiu M, Pedragosa M, Lu P, et al. Systems modeling of molecular mechanisms controlling cytokine-driven CD4+ T cell differentiation and phenotype plasticity. *PLoS computational biology*. 2013;9(4):e1003027. Epub 2013/04/18. doi: 10.1371/journal.pcbi.1003027. PubMed PMID: 23592971; PubMed Central PMCID: PMCPMC3617204.

24. Viladomiu M, Bassaganya-Riera J, Tubau-Juni N, Kronsteiner B, Leber A, Philipson CW, et al. Cooperation of Gastric Mononuclear Phagocytes with *Helicobacter pylori* during Colonization. *Journal of immunology (Baltimore, Md : 1950)*. 2017;198(8):3195-204. Epub 2017/03/08. doi: 10.4049/jimmunol.1601902. PubMed PMID: 28264969; PubMed Central PMCID: PMCPMC5380565.

25. Collier N, North M. Repast HPC: A platform for large-scale agentbased modeling. *Large-Scale Computing Techniques for Complex System Simulations*. 2011:81-110.

26. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, et al. COPASI—a complex pathway simulator. *Bioinformatics (Oxford, England)*. 2006;22(24):3067-74.

27. Childs H, Brugger E, Whitlock B, Meredith J, Ahern S, Pugmire D, et al. VisIt: An End-User Tool For Visualizing and Analyzing Very Large Data. *High Performance Visualization-Enabling Extreme-Scale Scientific Insight*. *Insight*. 2012:357-72.

28. Saltelli A, Tarantola S, Campolongo F. Sensitivity analysis as an ingredient of modeling. *Statistical Science*. 2000;15(4):377-95.

1
2
3
4 1201 29. Thorne BC, Bailey AM, Peirce SM. Combining experiments with multi-cell agent-based
5 1202 modeling to study biological tissue patterning. *Briefings in Bioinformatics*. 2007;8(4):245-57. doi:
6 1203 10.1093/bib/bbm024.
7 1204 30. Moon H, Dean AM, Santner TJ. Two-stage sensitivity-based group screening in computer
8 1205 experiments. *Technometrics*. 2012;54(4):376-87.
9 1206 31. Marino S, Hogue IB, Ray CJ, Kirschner DE. A methodology for performing global
10 1207 uncertainty and sensitivity analysis in systems biology. *Journal of theoretical biology*.
11 1208 2008;254(1):178-96. Epub 2008/06/24. doi: 10.1016/j.jtbi.2008.04.011. PubMed PMID:
12 1209 18572196; PubMed Central PMCID: PMCPMC2570191.
13 1210 32. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. *Global sensitivity*
14 1211 *analysis: the primer*: John Wiley & Sons; 2008.
15 1212 33. Rasmussen CE, Williams CK. *Gaussian processes for machine learning*. 2006. The MIT
16 1213 Press, Cambridge, MA, USA. 2006;38:715-9.
17 1214 34. Santner TJ, Williams BJ, Notz WI. *The design and analysis of computer experiments*:
18 1215 Springer Science & Business Media; 2013.
19 1216 35. Thiele JC, Kurth W, Grimm V. Facilitating parameter estimation and sensitivity analysis of
20 1217 agent-based models: A cookbook using NetLogo and R. *Journal of Artificial Societies and Social*
21 1218 *Simulation*. 2014;17(3):11.
22 1219 36. Ankenman B, Nelson BL, Staum J. Stochastic kriging for simulation metamodeling.
23 1220 *Operations research*. 2010;58(2):371-82.
24 1221 37. Lamoureux B, Mechbal N, Massé J-R. A combined sensitivity analysis and kriging
25 1222 surrogate modeling for early validation of health indicators. *Reliability Engineering & System*
26 1223 *Safety*. 2014;130:12-26.
27 1224 38. Chen X, Kim K-K. Stochastic kriging with biased sample estimates. *ACM Trans Model*
28 1225 *Comput Simul*. 2014;24(2):1-23. doi: 10.1145/2567893.
29 1226 39. Dancik GM, Dorman KS. mlegp: statistical analysis for computer models of biological
30 1227 systems using R. *Bioinformatics (Oxford, England)*. 2008;24(17):1966-7.
31 1228 40. Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based
32 1229 sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer*
33 1230 *Physics Communications*. 2010;181(2):259-70.
34 1231 41. Sobol IM. Sensitivity estimates for nonlinear mathematical models. *Mathematical*
35 1232 *modelling and computational experiments*. 1993;1(4):407-14.
36 1233 42. Sobol' IM, Tarantola S, Gatelli D, Kucherenko SS, Mauntz W. Estimating the
37 1234 approximation error when fixing unessential factors in global sensitivity analysis. *Reliability*
38 1235 *Engineering & System Safety*. 2007;92(7):957-60. doi: <https://doi.org/10.1016/j.res.2006.07.001>.
39 1236 43. Jansen MJ. Analysis of variance designs for model output. *Computer Physics*
40 1237 *Communications*. 1999;117(1-2):35-43.
41 1238 44. Alam M, Deng X, Philipson C, Bassaganya-Riera J, Bisset K, Carbo A, et al. Sensitivity
42 1239 Analysis of an ENteric Immunity Simulator (ENISI)-Based Model of Immune Responses to
43 1240 *Helicobacter pylori* Infection. *PloS one*. 2015;10(9):e0136139. Epub 2015/09/04. doi:
44 1241 10.1371/journal.pone.0136139. PubMed PMID: 26327290; PubMed Central PMCID:
45 1242 PMCPMC4556515.
46 1243 45. Heiner M, Gilbert D. *BioModel engineering for multiscale Systems Biology*. *Progress in*
47 1244 *Biophysics and Molecular Biology*. 2013;111(2):119-28. doi:
48 1245 <https://doi.org/10.1016/j.pbiomolbio.2012.10.001>.
49 1246 46. Alzahrani S, Lina TT, Gonzalez J, Pinchuk IV, Beswick EJ, Reyes VE. Effect of
50 1247 *Helicobacter pylori* on gastric epithelial cells. *World Journal of Gastroenterology : WJG*.
51 1248 2014;20(36):12767-80. doi: 10.3748/wjg.v20.i36.12767. PubMed PMID: PMC4177462.
52 1249 47. Mimuro H, Suzuki T, Nagai S, Rieder G, Suzuki M, Nagai T, et al. *Helicobacter pylori*
53 1250 dampens gut epithelial self-renewal by inhibiting apoptosis, a bacterial strategy to enhance
54
55
56
57
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

colonization of the stomach. *Cell host & microbe*. 2007;2(4):250-63. Epub 2007/11/17. doi: 10.1016/j.chom.2007.09.005. PubMed PMID: 18005743.

48. Wroblewski LE, Peek RM. Orchestration of Dysregulated Epithelial Turnover by a Manipulative Pathogen. *Cell Host & Microbe*. 2007;2(4):209-11. doi: <https://doi.org/10.1016/j.chom.2007.09.011>.

49. Zhang M, Liu M, Luther J, Kao JY. Helicobacter pylori directs tolerogenic programming of dendritic cells. *Gut microbes*. 2010;1(5):325-9. Epub 2011/02/18. doi: 10.4161/gmic.1.5.13052. PubMed PMID: 21327041; PubMed Central PMCID: PMC3023617.

50. Rizzuti D, Ang M, Sokollik C, Wu T, Abdullah M, Greenfield L, et al. Helicobacter pylori inhibits dendritic cell maturation via interleukin-10-mediated activation of the signal transducer and activator of transcription 3 pathway. *Journal of innate immunity*. 2015;7(2):199-211. Epub 2014/11/22. doi: 10.1159/000368232. PubMed PMID: 25412627.

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

52. Fagiolo G, Moneta A, Windrum P. A critical guide to empirical validation of agent-based models in economics: Methodologies, procedures, and open problems. *Computational Economics*. 2007;30(3):195-226.

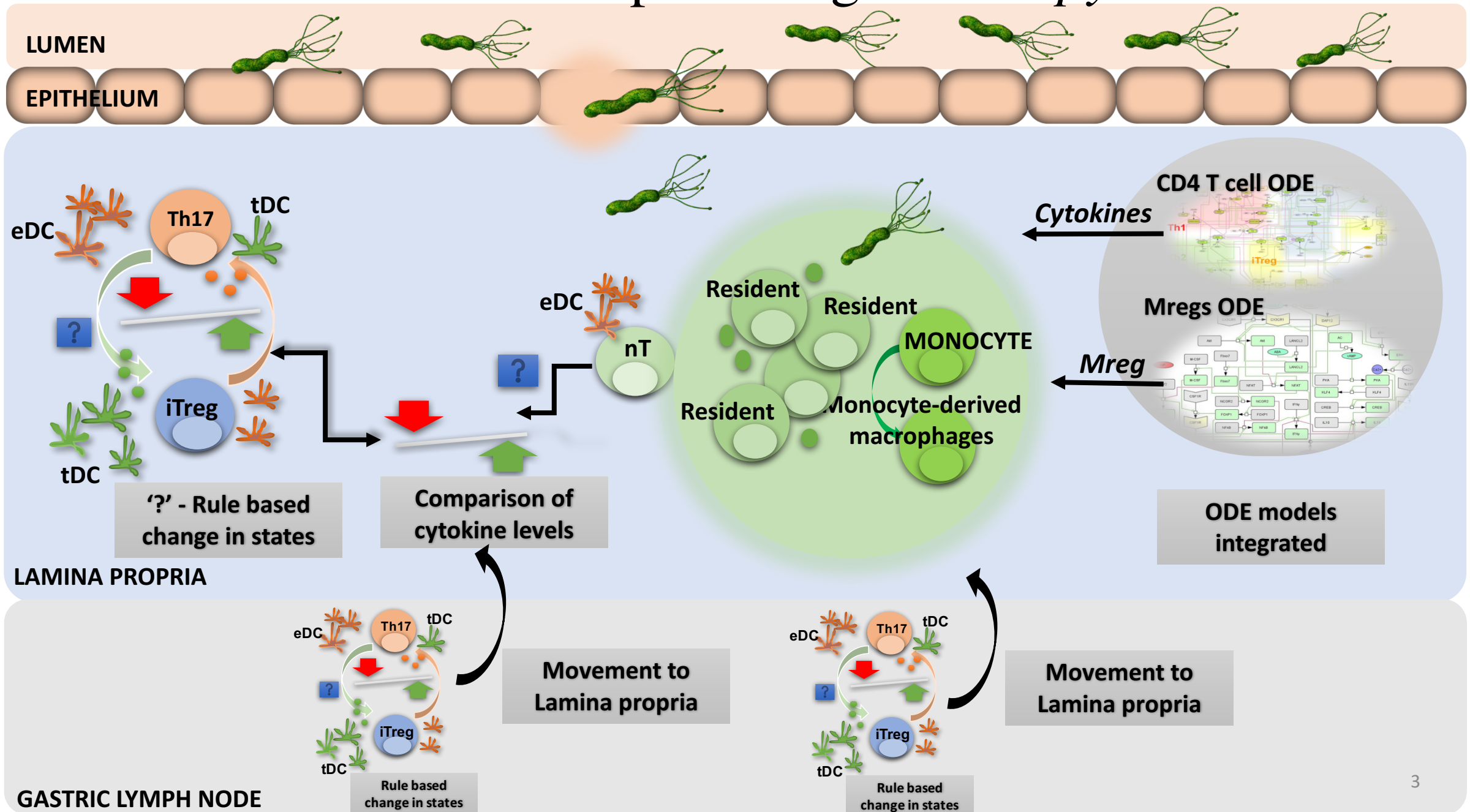
53. Crooks A, Castle C, Batty M. Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems*. 2008;32(6):417-30.

54. Filatova T, Verburg PH, Parker DC, Stannard CA. Spatial agent-based models for socio-ecological systems: Challenges and prospects. *Environmental modelling & software*. 2013;45:1-7.

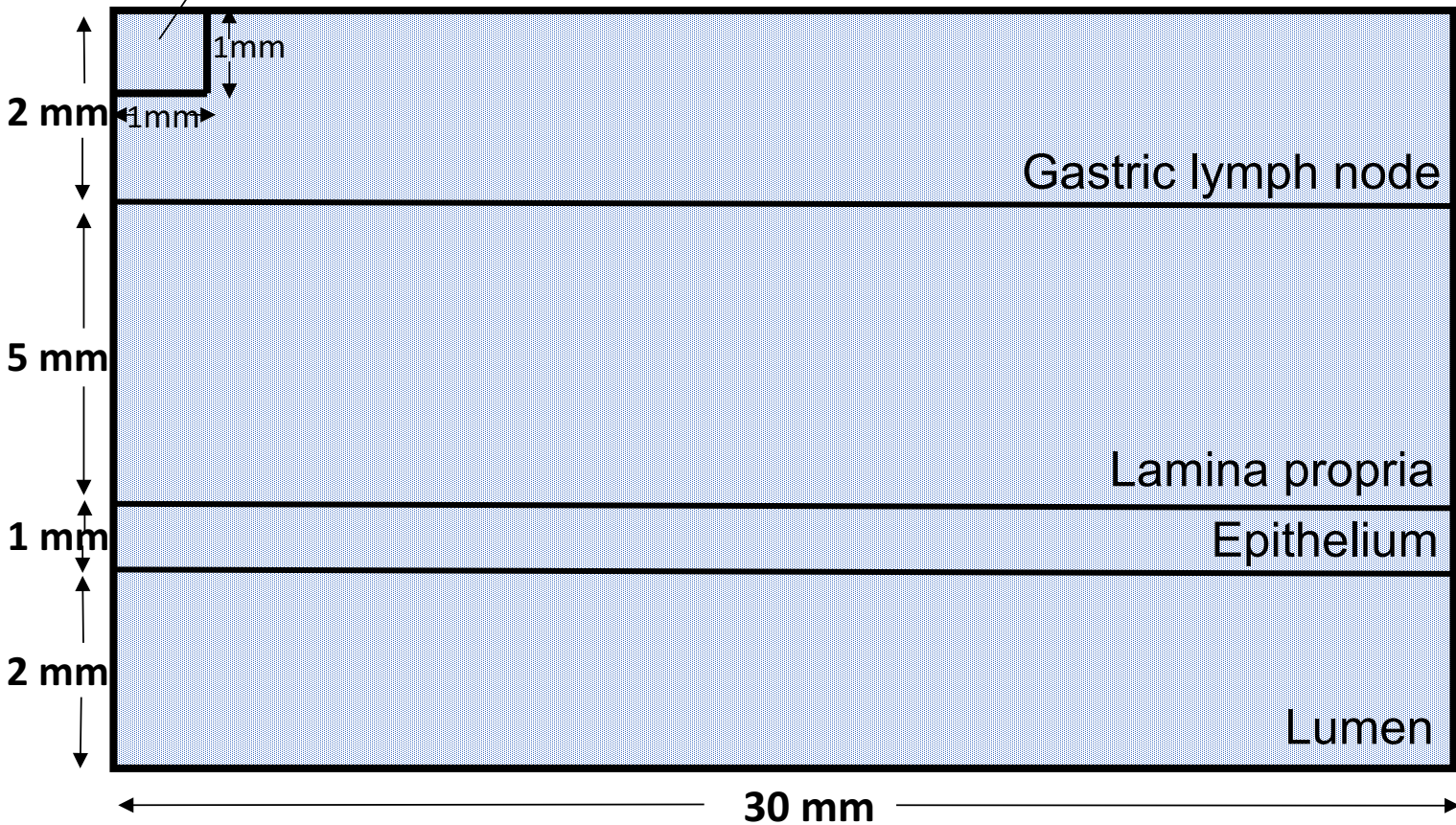
Name of Agent	Agent Type	Rules
<i>Helicobacter pylori</i>	<i>H. pylori</i>	<ul style="list-style-type: none"> - Moves across the epithelial cell border if near damaged epithelial layer - Proliferates in the lumen and lamina propria - Dies (removed from the simulation) in lamina propria and in the lumen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

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

Immune responses against *H. pylori*



lattice site



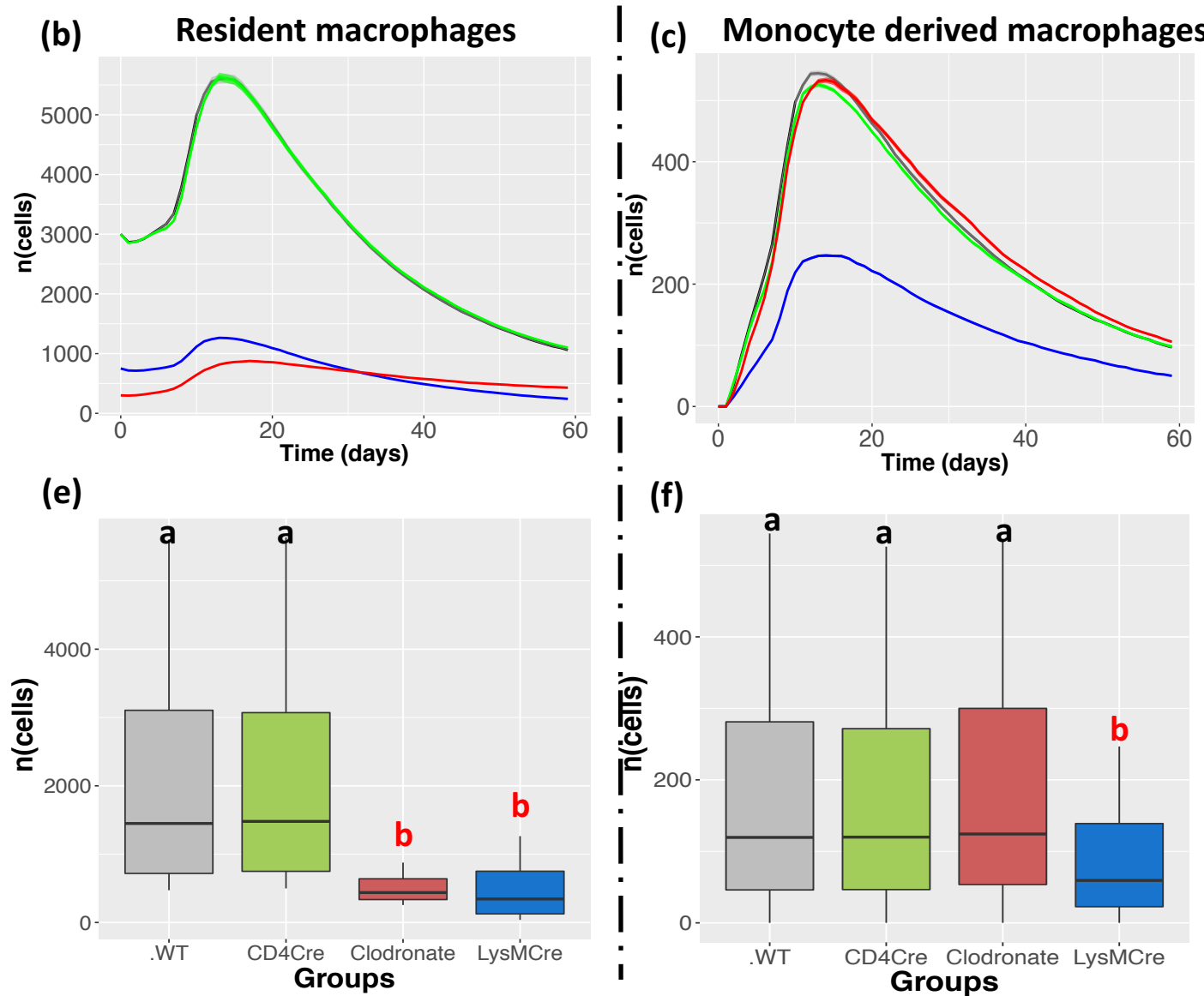
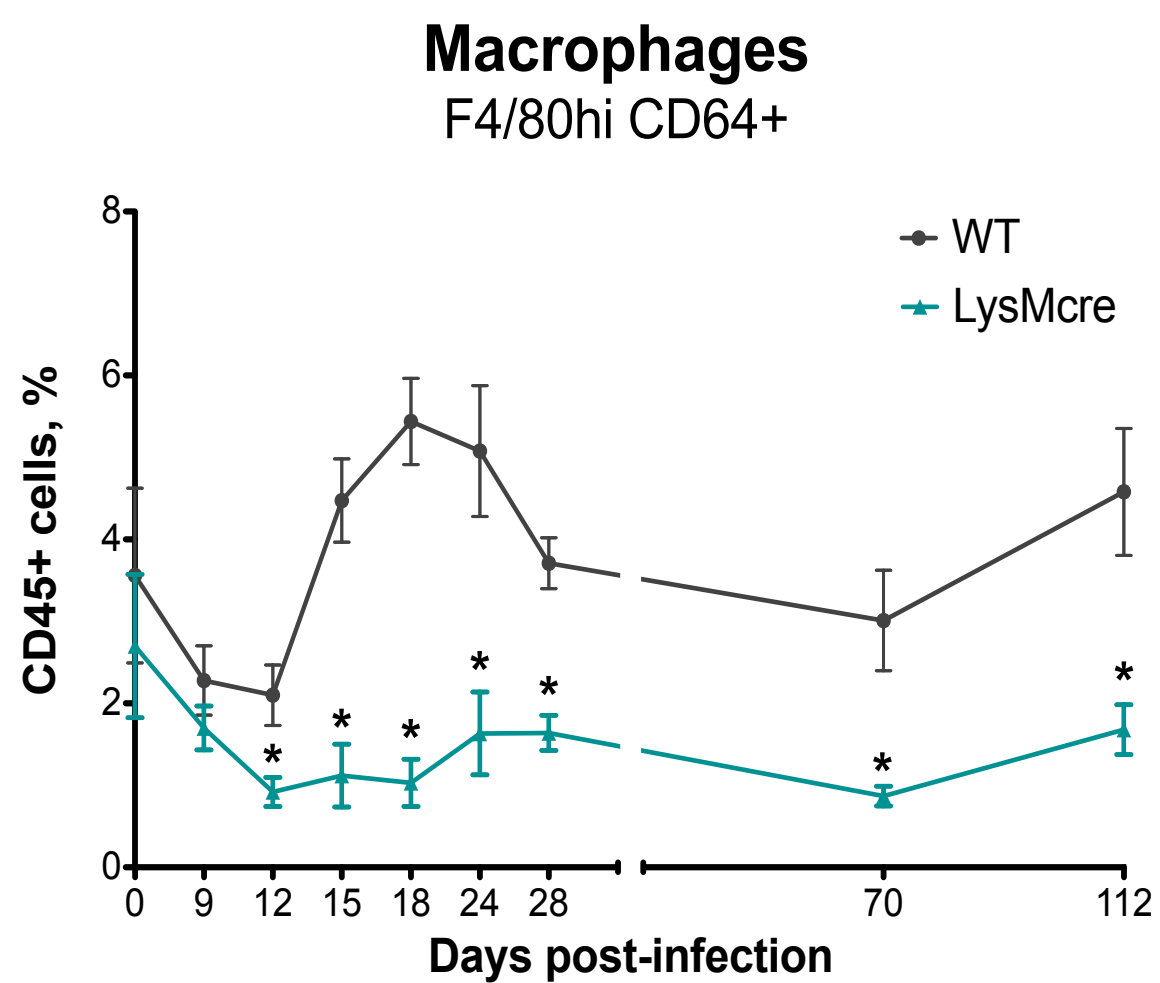
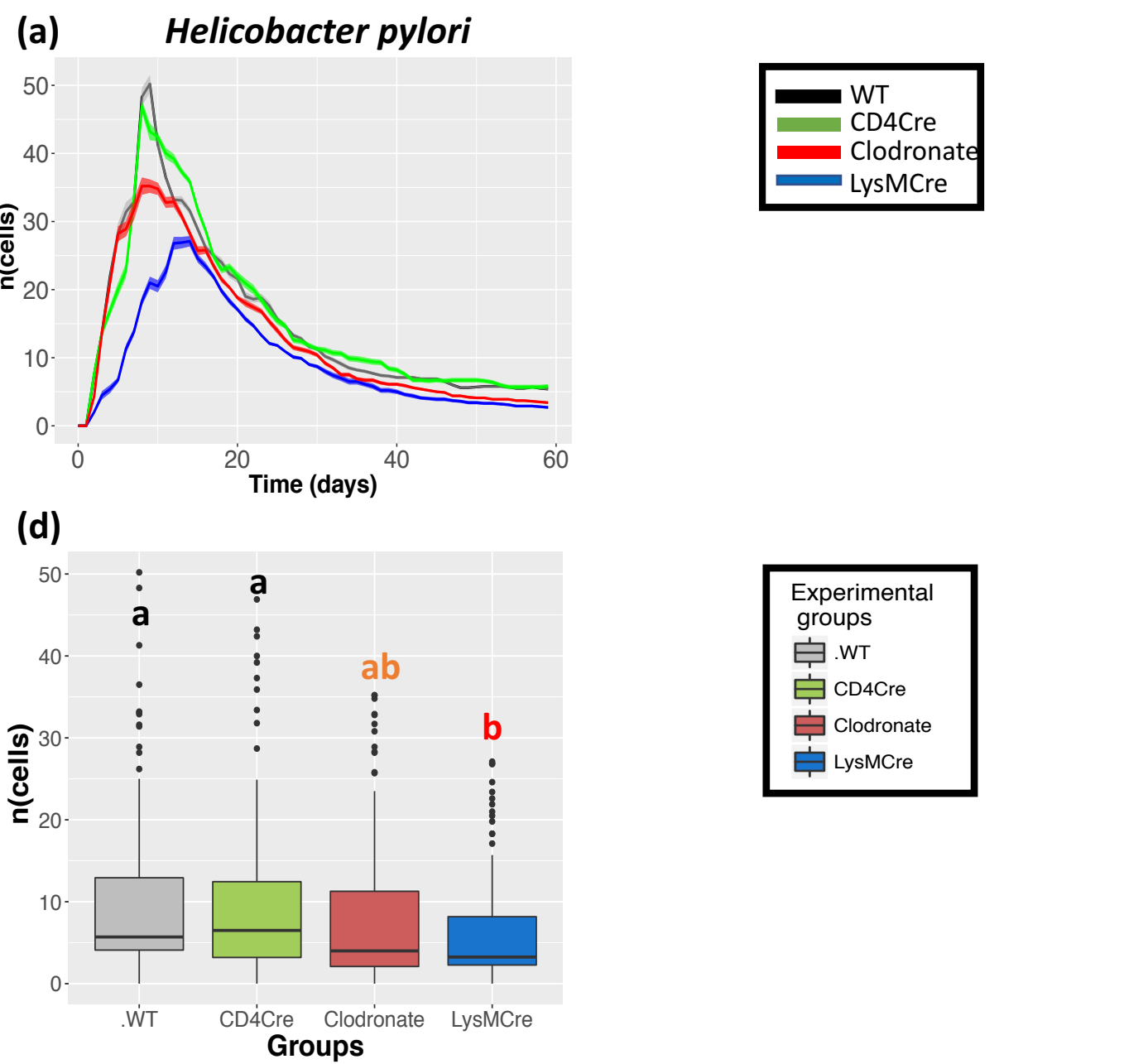
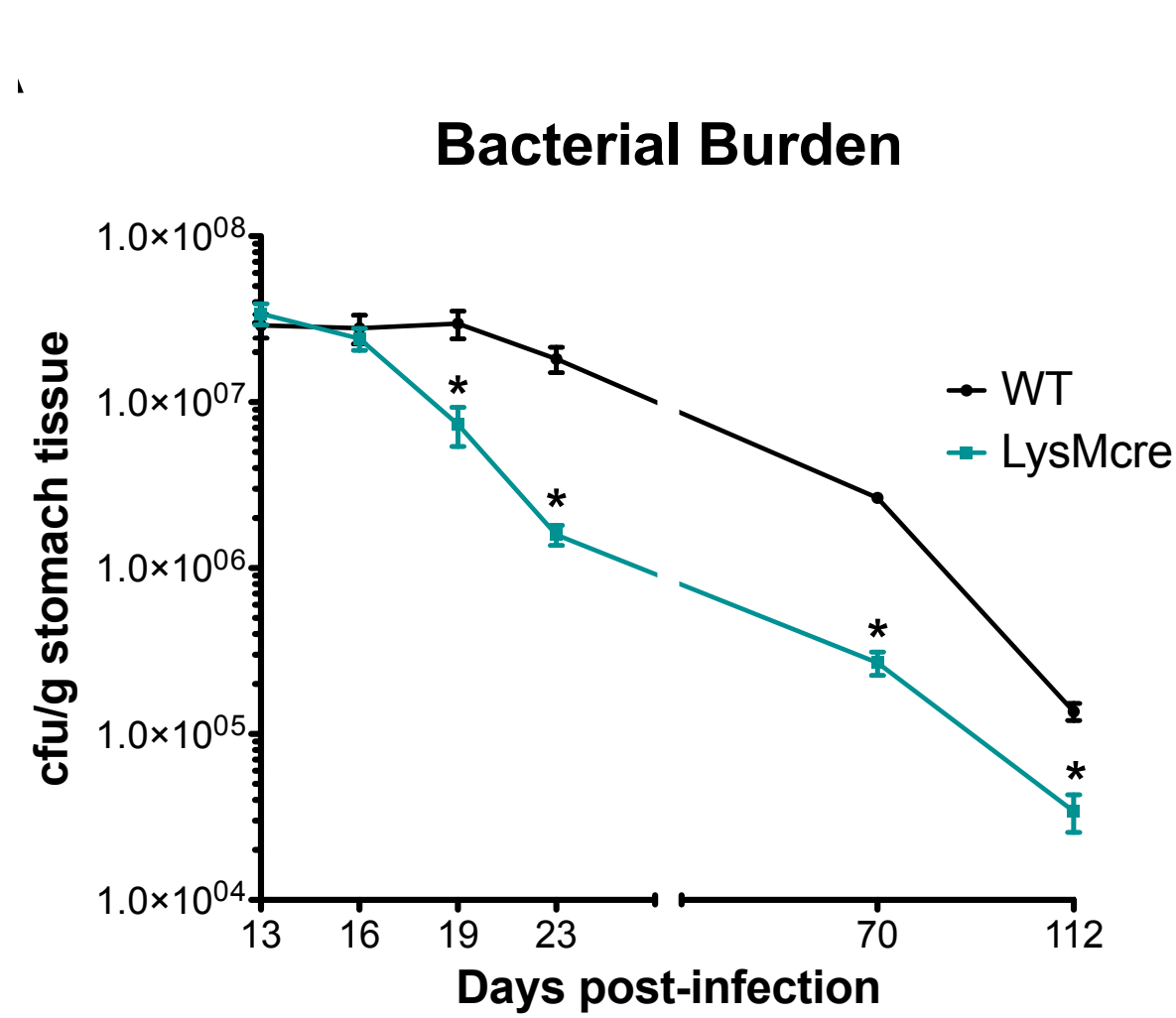
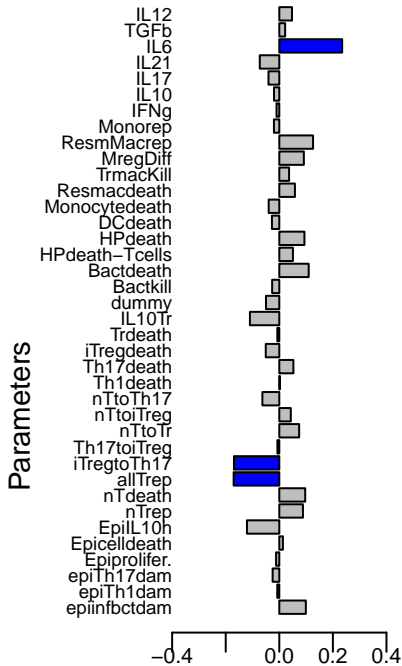
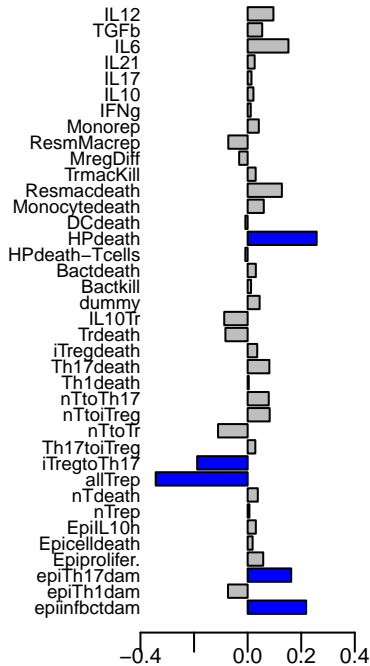


Figure 4

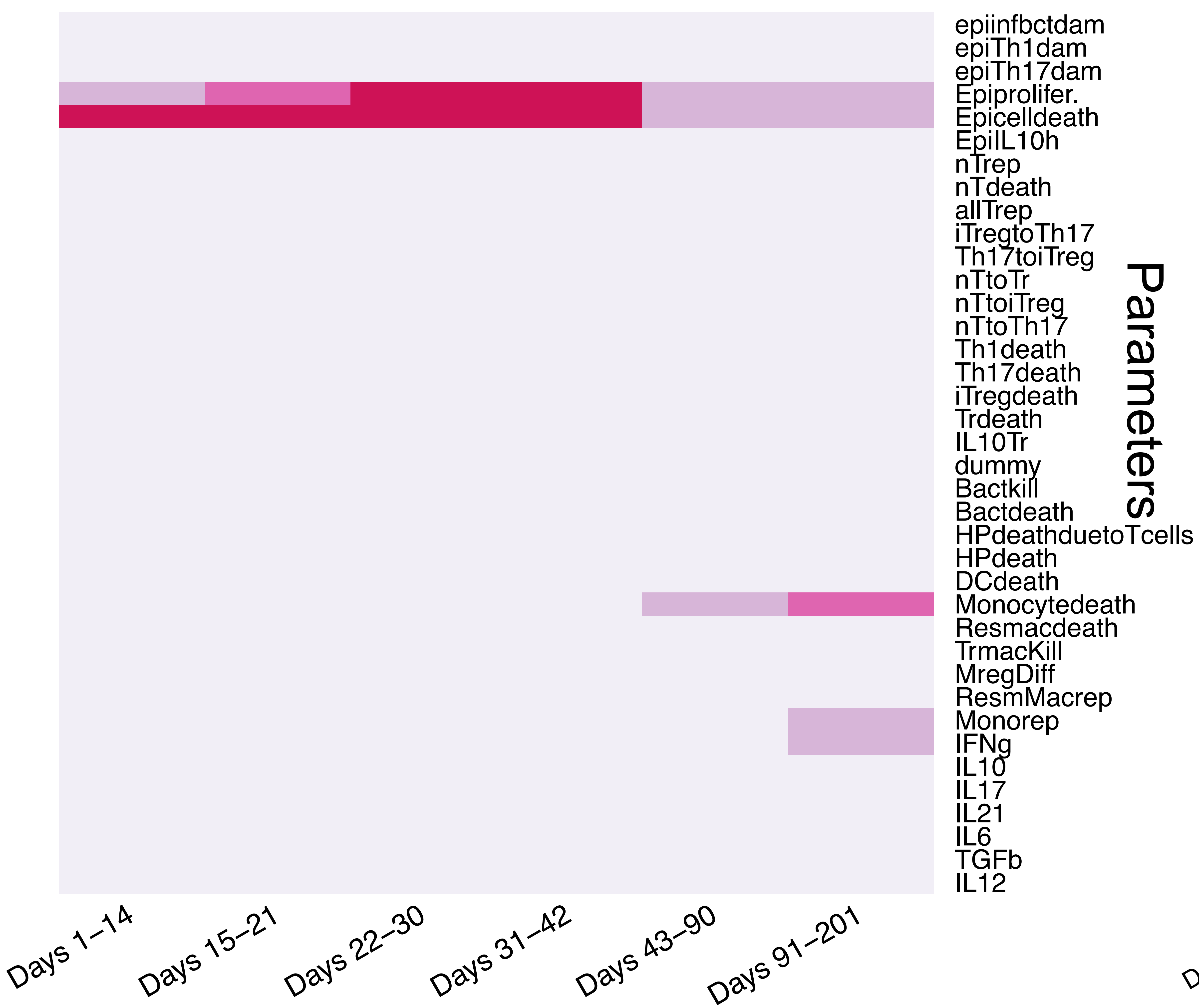
Helicobacter pylori in Lamina propria



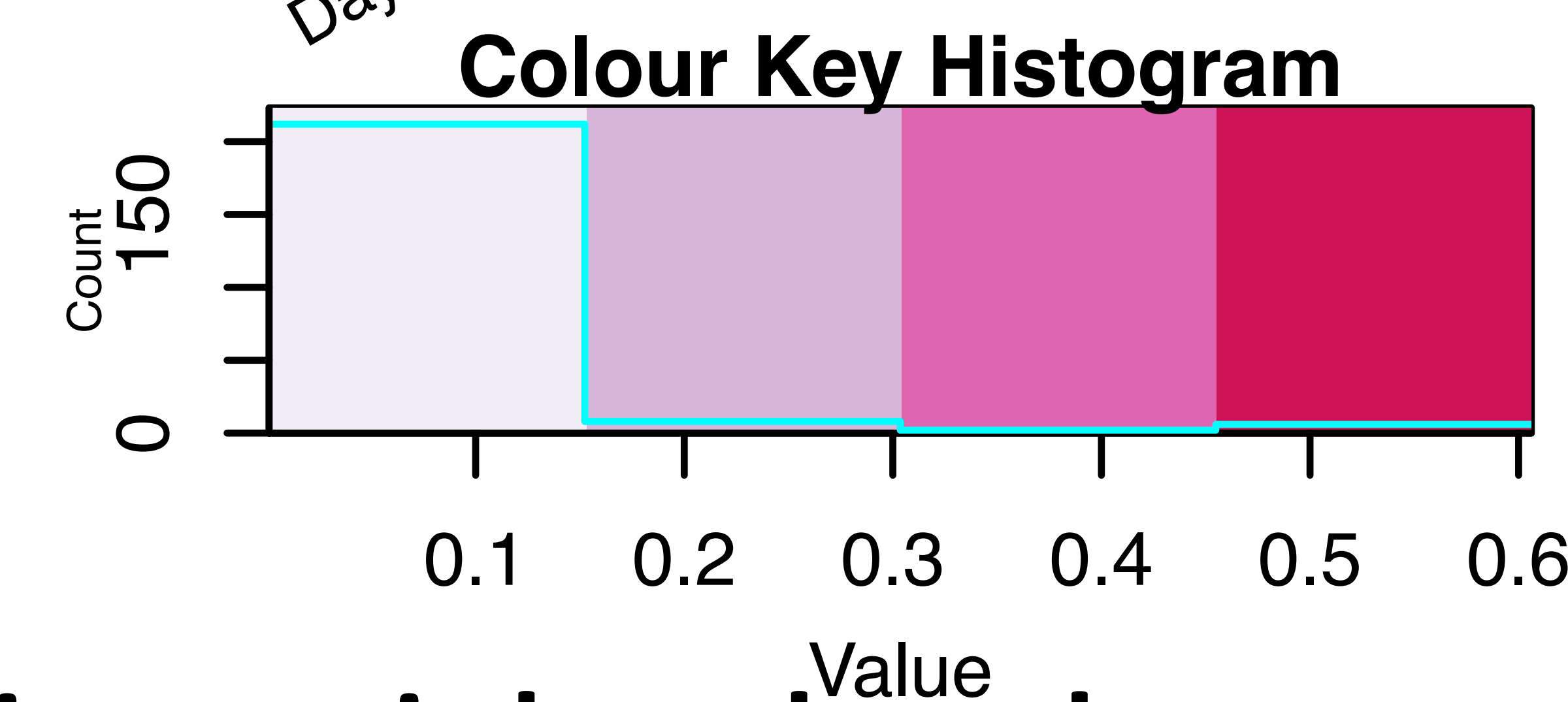
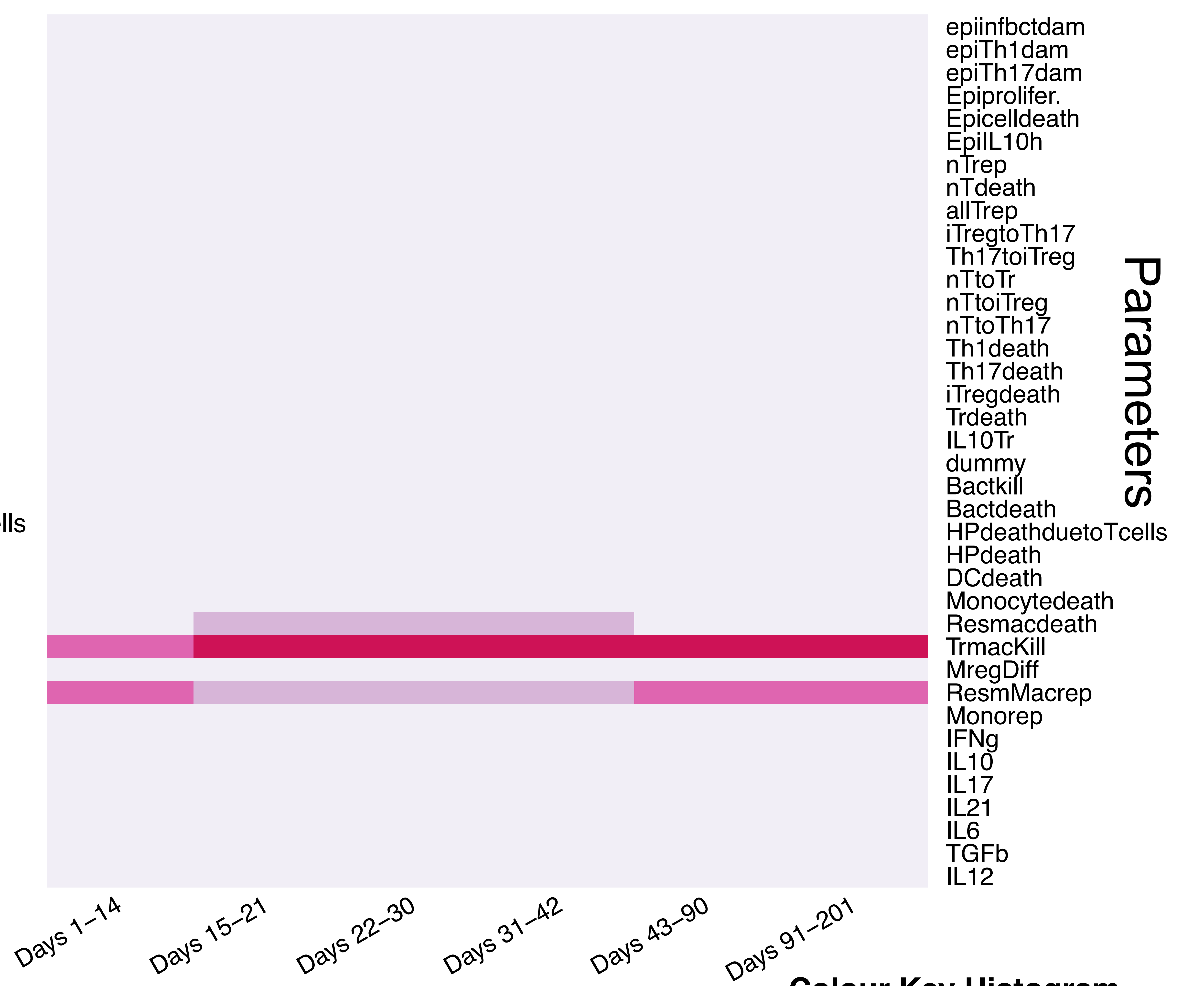
Click here to [access/download; Figure, Fig0.pdf](#) Resident macrophages in Lamina propria



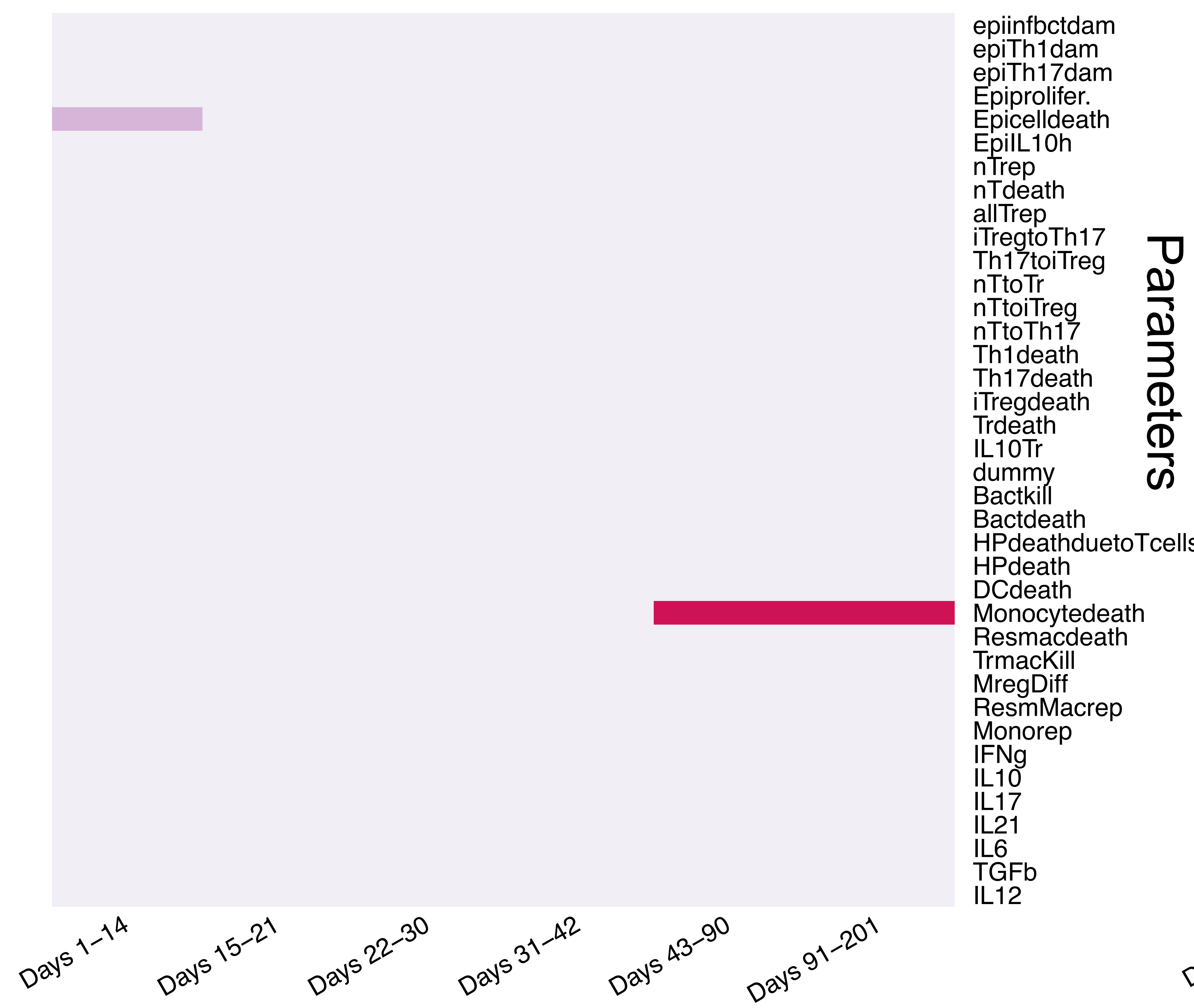
(a) *Helicobacter pylori* in lamina propria



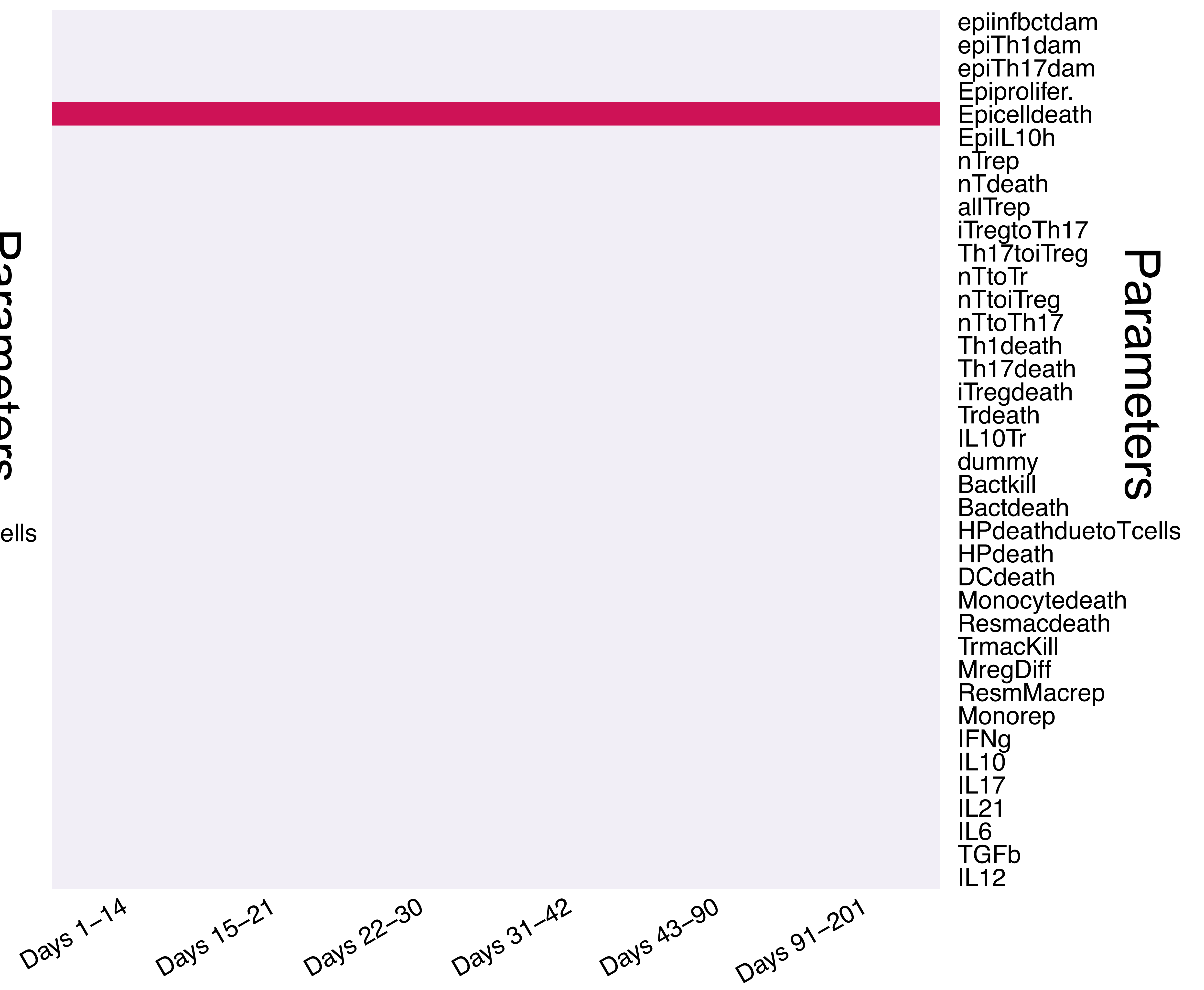
(b) Resident macrophages in lamina propria

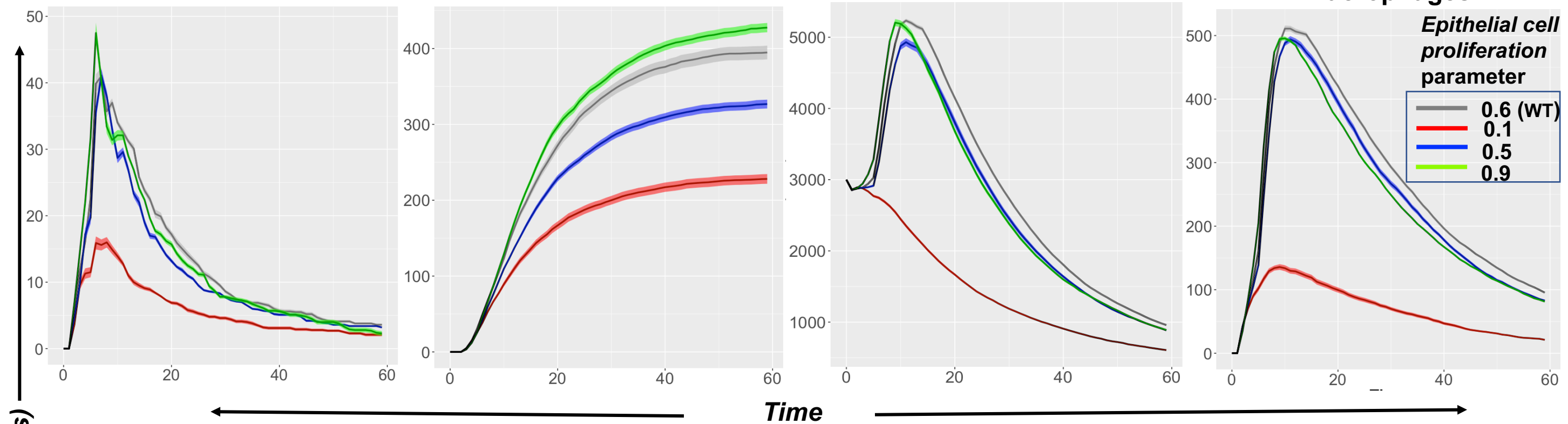
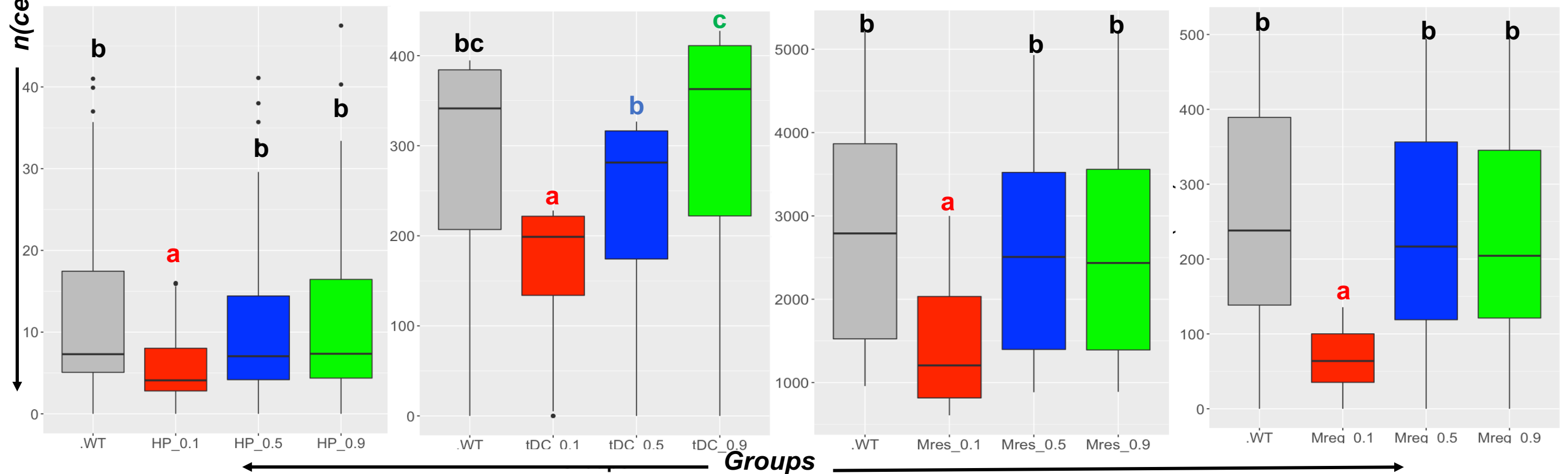



(c) Monocyte derived macrophages in lamina propria



(d) Tolerogenic dendritic cells in gastric lymph node




(a) *Helicobacter pylori***(b) Tolerogenic dendritic cells****(c) Resident macrophages****(d) Monocyte derived macrophages** $n(\text{cells})$ **Groups**




Click here to access/download
Supplementary Material
FileS1.pdf






Click here to access/download
Supplementary Material
FigS1.png



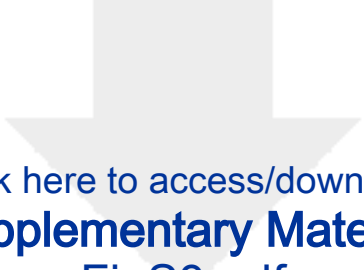
Click here to access/download
Supplementary Material
Table_S1_revised.docx



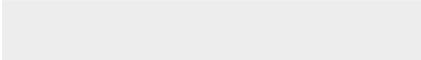


Click here to access/download
Supplementary Material
Fiig S2.pdf



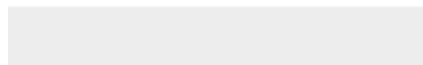



Click here to access/download
Supplementary Material
FigS3.pdf






Click here to access/download
Supplementary Material
FigS4.docx





Click here to access/download
Supplementary Material
FigS5.pdf



Once more, we would like to truly thank the reviewers and editor for the time and effort they are putting in reviewing our work. The fact that they've all brought up concerns about the grid dimensions and structure underlines the importance of computational biology as means of integrating biological concepts with biophysics and mathematics. Most biologists tend to focus on functions (i.e., gene expression or protein properties) and numbers, but they rarely consider the dimensions of the "biological space". The proposed revisions and clarifications better connect the biology, underlying mathematics and spatial/biophysical considerations. In response to the comments raised by the reviewers, we have revised the manuscript and provided explanations to further clarify the description of the grid.

Reviewer reports:

Reviewer #2 Comment 1: I have looked through the authors' responses and remain concerned about the grid dimensions. Redefining the grid in μm instead of nm still leaves the entire simulation environment at $30\mu\text{m}$ by $10\mu\text{m}$ which is about on the order of a single cell (e.g. macrophage diameter $\sim 20\mu\text{m}$ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1470168/>). I remain concerned that defining the grid at such a small scale creates a biologically unrealistic disconnect between the diffusing cytokines (on a very small scale) and the cells that they influence (on a much larger and abstract scale). I tried to understand the dimensions of the diffusion constants for the cytokines but the parameter table S1 contains no units to be able to compare diffusion scales to cell movement scales.

Response to Reviewer 2, Comment 1: *We thank the reviewer for their concern. We agree that defining the grid in μm leaves the simulation in a smaller scale and want to clarify that even though the units in the model are annotations, we understand that it is crucial to define the grid in a more biologically meaning way.*

*We define the area in the model being simulated as a simulation environment with $30\text{ mm} \times 10\text{ mm}$ two-dimensional grid. The size of an individual lattice site (previously referred to as grid cell in the paper) is $1\text{ mm} \times 1\text{ mm}$. The scales described as in the previous version of ENISI-MSM (Mei et al., 2015) were kept unchanged. The table describing the scales used in (Mei et al., 2015) are also shown here in **Table 2**.*

As described in table shown below (adapted from Table 1) of our previous work (Mei et al., 2015), the spatial properties for cytokine diffusion defined in the range of millimeters were unchanged in the version of ENISI-MSM used in this paper.

Scale	Example scenario	Spatial(m)	Time(s)	Technology	Tool
Intra-cellular	Signaling pathways	Nano (nm)	Nano	ODE	COPASI
Cellular	Cell movement and subtype	Milli (mm)	Tens	ABM	ENISI
Intra-cellular	Cytokine-diffusion	Milli (mm)	Tens	PDE	ValueLayer
Tissue	Inflammation and lesions	Centi (cm)	Thousands	Projection	ENISI

Table 2. The four scales of ENISI models, their spatial and temporal properties and modeling technologies and tools used for each scale. (Table 1 as adapted from (Mei et al., 2015))

We updated the manuscript accordingly, please refer to L213-L223.

We thank the reviewer for pointing this out. We updated Table S1 with units for clarity. Lastly, we want to clarify that the model deals with numbers and the units are annotations in the simulation

hence the corrections in the dimensions above do not affect the simulation results in any way. All the values used in the code were internally consistent with the model.

Reviewer #3, Comment 1: In this review, I am looking at the more limited questions on the matter of units and scale, which have been raised by both Reviewer #1 and Reviewer #2.

Generally, I don't think the units matter, so long as they have been handled correctly (i.e., with correct conversion and internal consistency within the computational model). The units should be properly labeled in any parameter tables.

I tend to use microns and minutes (or seconds for some problems) for this scale of problem, but there are others who just use cm and sec for everything. As long as the values are correct in the displayed units (and as long as the code used internally consistent values), I think it's purely aesthetic. That said, if something is being labeled on a multicellular level, then labeling 10 microns will be much more appropriate than 10,000 nm.

Response to Reviewer 3, Comment 1: We thank the reviewer for highlighting this point. We concur that the units don't matter for the simulation results, but want to best clarify the dimensions to make them relevant with the biology. We have labelled all the units in the parameter Table S1. The values used in the code were internally consistent with the model and handled correctly. For further clarity, we show the scales used in the previous ENISI models and the spatial and temporal properties as described in (Mei et al., 2015) and included the Table 1 from (Mei et al., 2015).

Scale	Example scenario	Spatial(m)	Time(s)	Technology	Tool
Intra-cellular	Signaling pathways	Nano (nm)	Nano	ODE	COPASI
Cellular	Cell movement and subtype	Milli (mm)	Tens	ABM	ENISI
Intra-cellular	Cytokine-diffusion	Milli (mm)	Tens	PDE	ValueLayer
Tissue	Inflammation and lesions	Centi (cm)	Thousands	Projection	ENISI

Table 2. The four scales of ENISI models, their spatial and temporal properties and modeling technologies and tools used for each scale. (Adapted from (Mei et al., 2015)).

We updated the manuscript accordingly, please refer to L213-L223.

Reviewer #3, Comment 2: However, I want to further dig into Reviewer #2's concerns on grid sizes and scales.

As I read this draft, I see at least one potential source of confusion: this team appears to be very focused on mathematical and numerical methods. As such, they are using the word "cell" for both a biological cell and a computational lattice site. This is a really bad idea, and the authors should pick a better nomenclature (e.g., computational mesh or lattice site) to avoid this confusion. They should never use "cell" to mean anything other than a biological cell once they enter computational biology.

Response to Reviewer 3, Comment 2: The focus of the paper was to utilize an already published tool (Mei et al., 2015) to study *Helicobacter pylori* infection. So, a significant focus of the work was on the biology. However, we agree that the nomenclature should be clarified to avoid confusion from using the term "cell" for both actual biological cells and space/grid units. In the revised version of the manuscript we have substituted "grid cell" for "lattice site" and kept "cell" to refer to biological cells.

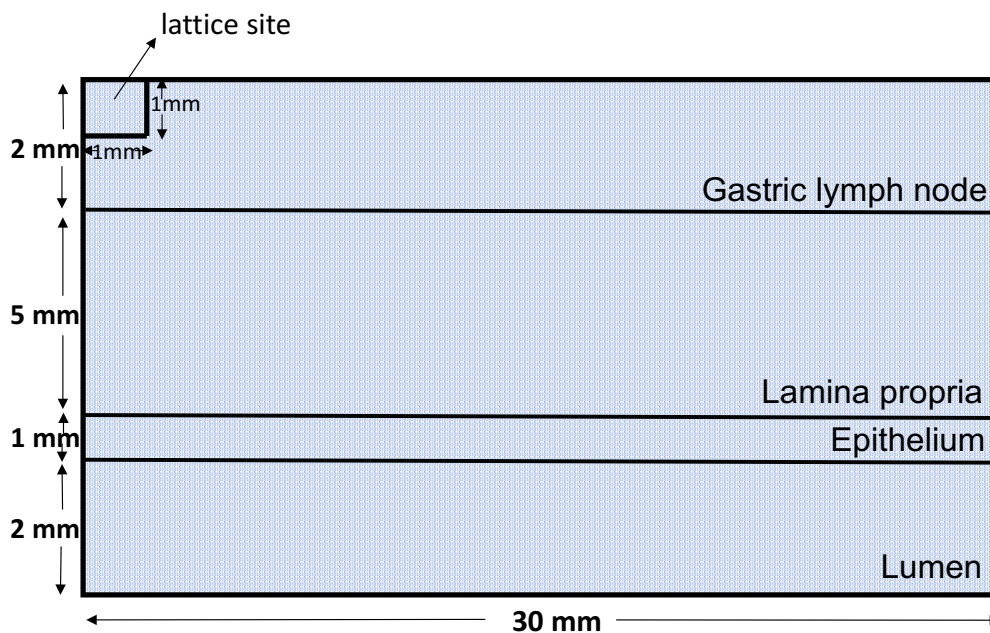
Reviewer #3, Comment 3: Next, they need to be clearer about what their grids represent. They should show a picture of the domain and meshing in their main text, and not just supplementary material. They should probably also clarify that they are simulating a cross-section of tissue in their model, rather than 3D or some top-down view. (At least from what I can tell.) (They seem to address this in text, but showing the mesh will provide better clarity.)

Response to Reviewer 3, Comment 3: We thank the reviewer for their valuable suggestion regarding this crucial point. We included a cartoon picture of the domain and mesh (as Fig. 2) in the main text as opposed to in the supplementary material (as shown below).

We highlighted that we are simulating a cross-section of tissue in the model and we redefined it as a simulation environment with 30 mm x 10 mm two-dimensional grid.

The size of an individual lattice site (previously referred to as grid cell in the paper) is 1 mm x 1mm.

We updated the manuscript accordingly, please refer to L231-L235.



Reviewer #3, Comment 4: Assuming the authors now have a 30 micron x 10 micron domain, they can simulate at most one epithelial cell, if it's all in plane. But if it's a cross-section, I suppose they could have more. Perhaps many h pylori (which they size at about 1 micron), but not many mammalian cells. So, their computational domain is still not very clear to me, and they should just show it, with appropriate labeling. They seem to skip straight to population dynamics in their figures, but it would be very helpful if they showed one actual spatiotemporal simulation. This would make the nature and performance of their model much clearer.

Response to Reviewer 3, Comment 4: We thank the reviewer for their concern.

With a redefined computational simulation environment of 30 mm x 10 mm, the epithelium is comprised of hundreds of epithelial cells.

For example, if the initial number of epithelial cells defined by the user is 12, the total number of epithelial cells amounts to = (30×1) dimension of epithelial compartment $\times 12_{\text{initial number}} = 360$.

In addition to the figure of the grid environment, we included the screenshots of one actual in silico simulation of *H. pylori* infection to highlight the spatiotemporal aspects of the modeling outputs. The screenshots were created using VisIt version 2.12 (Childs et al., 2012), an interactive visualization and analysis tool. As shown in Additional file Fig. S2 the screenshots represent the spatial distribution of different agent cells over time points (2, 4, 5 and 6) distributed across the 2D grid. Further, we presented the insets in Fig. S2 showing a zoomed in portion of the respective grids across the time steps 2, 4, 5 and 6.

We also want to clarify that the agents represented in the screenshots below are only for visual representation and do not represent the actual size of the biological cells.

We updated the manuscript accordingly and added Fig 3, please refer to L333-L338.

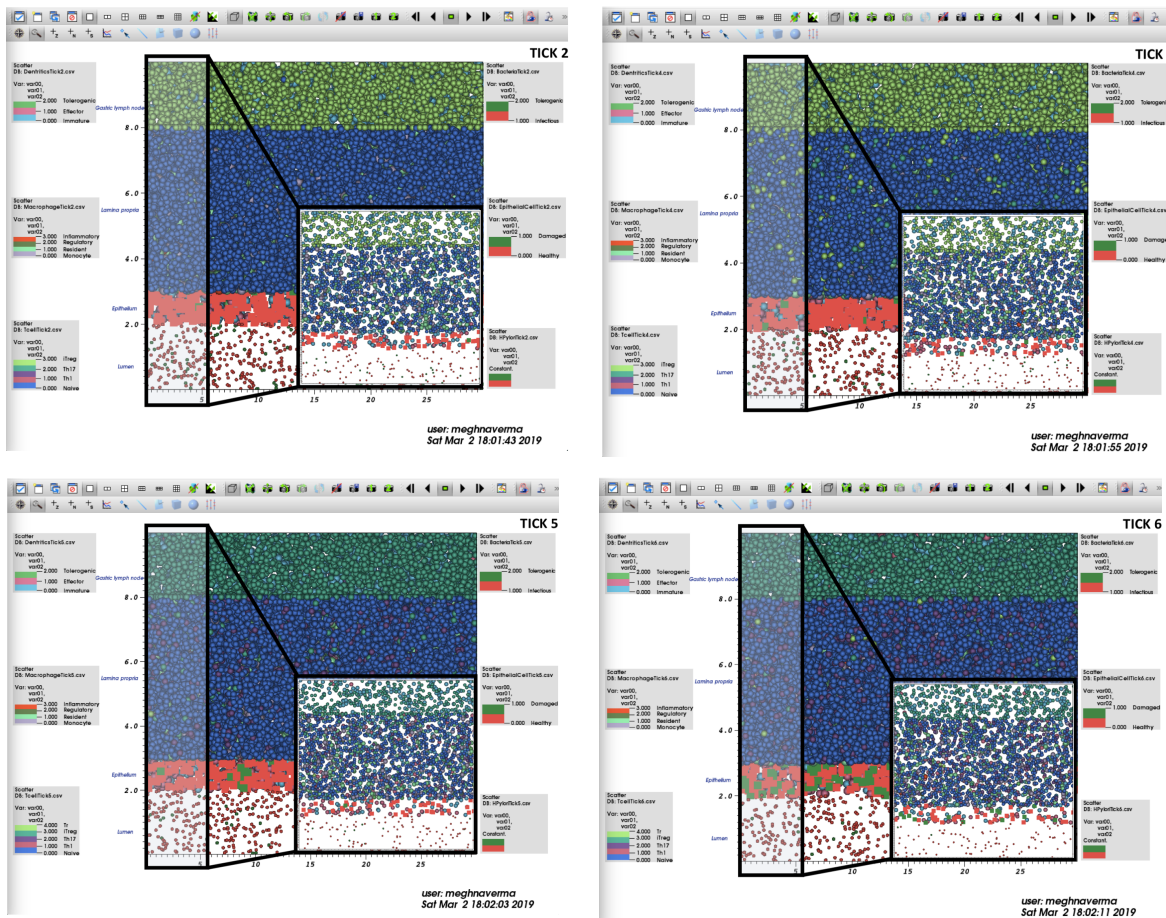


Fig S2. Time screenshots of a *Helicobacter pylori* infection modeled in a 30 mm (length) x 10 mm (width) two-dimensional grid. The thickness of the compartment is shown on the y-axis, such that: lumen spans (0 to 2) units, epithelium spans (2 to 3) units, lamina propria spans (3 to 8) units and gastric lymph node across (8-10) units on the scale. Two-dimensional distribution of different cell subsets over the time

steps (ticks) 2, 4 (top panels), 5 and 6 (bottom panels) are shown. The insets in each image shows a zoomed in portion of the respective grids across the time steps 2, 4, 5 and 6. The agents represented in the screenshots below are only for visual representation and do not represent the actual size of the biological cells. Future refinements of the model will create agents of the actual sizes of cells.

Reviewer #3, Comment 5: On sizes and scaling, I fully agree with Reviewer #2: if this is indeed a 30 micron x 10 micron cross section, there's no way there are more than a handful of mammalian cells at any time in any simulation. If they have made a scaling argument (and there are such arguments that could be made if formulated clearly and rigorously), they'd better be clear about it. Any results that show thousands of mammalian cells in a 30 x 10 micron² domain are simply beyond biophysical plausibility.

Response Reviewer 3, Comment 5: We defined the simulation environment as a 30 mm x 10 mm two-dimensional grid that represents a cross section area of stomach tissue modeled here.

Reviewer #3, Comment 6: Again, just actually showing a simulation (either a movie, or some time snapshots, but showing locations of all the cell agents and substrate distributions) would help clarify things much more. No limit to the number of cells in a mesh site, while mathematically possible, does not make sense on such a small simulated domain. Even though the authors treated the cell agents as no size (infinitesimal points), there are physical limits, and moreover if each cell is absorbing / secreting things at appropriate rates, then there should be ridiculous amounts of secretion of growth factors and ridiculous depletion of growth substrates, if there is a huge overcrowding of hundreds or thousands of mammalian cells in a 1 micron x 1 micron lattice site.

Response to Reviewer 3, Comment 6: We thank the reviewer for their valuable suggestion and have included time snapshots for the simulation (time points 2, 4, 5 and 6) created using VisIt version 2.12 (Childs et al., 2012), an interactive visualization and analysis tool. Please refer to Additional file Fig S2 also included in the above response to Comment 5. The hybrid multiscale modeling platform ENISI MSM is currently capable of scaling up to up to 10⁹ agents, at which the memory (on a 32 GB node) was exceeded due to the large number of agents.

Reviewer #3, Comment 7: If that's what's going on, the authors really do need to take a step back and consider domains sufficiently large to capture hundreds or thousands of mammalian cells. This is not simply a matter of relabeling axes: it's a matter of simulating a larger physical domain that is suited to the size of objects (mammalian cells) that they are considering, with biophysically reasonable parameters. The authors need to carefully review all their parameters (e.g., cell densities) to ensure they are correctly scaled and reasonable. If there is nobody on the team with sufficient domain expertise to review these parameters and results to check for reasonableness, it may be time to grow the team. Given that actual mammalian cells are much larger than the computational lattice sites, there must be constraints: if a lattice site is "occupied" by a mammalian cells, so are many (or most!) of the surrounding lattice sites. They would be no-fly zones for further mammalian cells. The common way to solve such problems is to use computational lattice sites that are of comparable size to the largest biological cells (e.g., 1 biological cell per lattice site, as a cellular automaton model), or use very large lattice sites (e.g, 100 micron x 100 micron) that can truly contain multiple cells.

Response to Reviewer 3, Comment 7: We thank the reviewer for their suggestion on the size of the lattice site.

In the current model, the size of the lattice site (referred as spatial grid previously) is 1mm x 1mm, capable of containing multiple cells.

The total number of agents in a compartment of size (length x width) is calculated as follows: number (agents) x size (compartment),

For example, if the user set the initial number of epithelial cell agents to be 12, the total number of epithelial agents within the epithelium compartment amounts to -

$$12_{\text{initial_number}} \times (30 \times 1)_{\text{size_epithelium_compartment}} = 360.$$

Reviewer #3, Comment 8: All is not lost, however. If for some reason that larger domain is computationally infeasible (hard to imagine), the authors really don't need a 1 micron mesh resolution for most of the effects here. The diffusion length scale of most chemokines and diffusing substrates would not require a 1 micron mesh resolution. (And numerical stability will improve if coarser mesh resolutions are used.) If the authors feel a 1 micron mesh is needed for the bacteria, they could easily use a separate mesh. This work looks interesting. I think the team has a great contribution to make, if they pay a bit closer attention to the biophysical limits of their system.

Response to Reviewer 3, Comment 8: The computational domain used here is of 30 x 10 units in size and the individual lattice site is 1 units for the simulation. The lattice site is a configurable run parameter and can be changed without modifying the model.

We thank the reviewer for their valuable suggestions, inputs and concerns and have tried to clarify the questions around the grid dimensions.

With these revisions and clarifications, we believe that the revised manuscript is acceptable for publication.

Thank you for considering this work.

References:

- Childs, H., Brugger, E., Whitlock, B., Meredith, J., Ahern, S., Pugmire, D., ...Weber, G. (2012). VisIt: An End-User Tool For Visualizing and Analyzing Very Large Data. High Performance Visualization-Enabling Extreme-Scale Scientific Insight. *Insight*, 357-372.
- Mei, Y., Abedi, V., Carbo, A., Zhang, X., Lu, P., Philipson, C., ..Bassaganya-Riera, J. (2015). Multiscale modeling of mucosal immune responses. *BMC Bioinformatics*, 16 Suppl 12, S2. doi:10.1186/1471-2105-16-s12-s2