Identifying parameters of host cell vulnerability during *Salmonella* infection by quantitative image analysis and modeling

Jakub Voznica[1,2,3], Christophe Gardella[4,5], Ilia Belotserkovsky[6], Alexandre Dufour[2], Jost Enninga[1], Virginie Stévenin[1*]

[1]Institut Pasteur, Dynamics of Host-Pathogen Interactions Unit, 25 rue du Dr. Roux, 75724 Paris, France

[2]Institut Pasteur, BioImage Analysis Unit, 25 rue du Dr. Roux, 75724 Paris, France

[3]Ecole Normale Supérieure de Cachan, Université Paris-Saclay, 61 avenue du Président Wilson, 94235 Cachan, France

[4]Laboratoire de Physique Statistique, CNRS, UPMC and Ecole Normale Supérieure, 24, rue Lhomond, 75005 Paris, France

[5]Institut de la Vision, INSERM and UPMC, 17 rue Moreau, 75012 Paris, France

[6]Institut Pasteur, Molecular Microbial Pathogenesis Unit, 25 rue du Dr. Roux, 75724 Paris, France

*For correspondence:

    Virginie Stévenin   virginie.stevenin@ens-cachan.fr

Key words:

Salmonella enterica serovar Typhimurium, cooperative behavior, cell vulnerability, single-cell heterogeneity, mathematical modeling.

Running title:

Host cell vulnerability during *Salmonella* infection

1   **ABSTRACT**

2

3

4   *Salmonella* target and enter epithelial cells at permissive entry sites: some cells are more

5   likely to be infected than others. However the parameters that lead to host cell heterogeneity

6   are not known. Here, we quantitatively characterized host cell "vulnerability" towards

7   *Salmonella* infection based on imaged parameters. We performed successive infections of

8   the same host cell population followed by automated high-throuput microscopy and observed

9   that infected cells have higher probability of being re-infected. Establishing a predictive

10 model we identified two combined origins of host cell vulnerability: the pathogen-induced

11 cellular vulnerability emerging from *Salmonella* uptake and persisting at later stage of the

12 infection, and the host cell-inherent vulnerability. We linked the host cell inherent vulnerability

13 with its morphological attributes such as the local cell crowding, and with host cell cholesterol

14 content. This showed that the probability of *Salmonella* infection success can be forecast

15 from morphological or molecular host cell parameters.

1   *Salmonella enterica* serovar Typhimurium (*S*. Typhimurium) is a Gram-negative bacterium
2   that causes enteric diseases in many vertebrates after ingestion of contaminated food or
3   water. Salmonellosis is one of the most common causes of food-borne diseases in humans
4   and is considered to be major public health and global economic problem (26). After oral
5   uptake, more than 99% of *S*. Typhimurium are killed in the stomach or in the gut (2). The
6   surviving bacteria reach the distal ileum where they invade non-phagocytic intestinal
7   epithelial cells (38). *In vitro* experiments have shown that *S*. Typhimurium invasion of host
8   cells occurs after a phase of bacterial "Near Surface Swimming" (NSS) on the epithelial
9   layer. The bacteria scan the surface and eventually stop and dock at a "selected" host cell
10   (28, 37). Docking is irreversible (29) and followed by injection of *Salmonella* effectors into the
11   host cell through a Type 3 Secretion System (T3SS), leading to the formation of ruffles that
12   engulf the incoming bacterium (12, 22). Upon internalization *S*. Typhimurium either develops
13   inside a *Salmonella*-Containing Vacuole (SCV) or it ruptures the SCV to escape into the
14   cytoplasm where the pathogen replicates at a high rate, a phenomenon called
15   hyperreplication (HR) (20, 33).
16
17   The mechanism by which *S*. Typhimurium targets specific host cellular sites for its entry
18   remains debated. Santos and colleagues suggested that mitotic cells are selected due to
19   increased cholesterol accumulation at the cell surface during metaphase (32). By contrast,
20   Misselwitz and colleagues proposed that physical obstacles and forces that occur during the
21   process of NSS lead to the targeting of topologically prominent sites, such as dividing cells or
22   membrane ruffles (28). Finally, Lorkowski and colleagues have reported that the invasion of
23   *S*. Typhimurium at the ruffle site is a highly cooperative effort (25, 29). Indeed, co-infection of
24   WT and non-invasive *S*. Typhimurium mutants result in the entry of both strains inside the
25   host cells: non-invasive *S*. Typhimurium mutants are trapped at ruffle sites and concomitantly
26   internalized within the host cell, following active invasion by WT *S*. Typhimurium. However,
27   the cooperative effect between intracellular and entering bacteria remains poorly understood
28   at latter stage of the infection.
29
30   An increasing number of studies have highlighted the relevance of intrinsic cellular
31   heterogeneity within eukaryotic monocultures. After seeding, cells display a dynamic range of
32   variability in their morphology depending on their local microenvironment, including local
33   density, and peripheral or central position within cellular islets (35). This heterogeneity results
34   in differences of transcription (24, 9), lipid composition (9, 35) and sensitivity toward
35   infections (35). Such cell-to-cell variations have been studied during viral infection revealing
36   that simian virus 40 and mouse hepatitis virus present a population-determined pattern of
37   infection associated with differential cell local crowding (35). In the context of bacterial

1   infection, cell targeting has been related to bacterial cooperation at the entry site and

2   evaluated at the whole population level using Colony Forming Unit (CFU) counting or flow

3   cytometry analysis (25), but so far not *in situ* at the single cell level.

4

5   Here we investigated the susceptibility of epithelial host cells within the same cell population

6   to become infected by *S*. Typhimurium. Our analysis revealed that some cells are more likely

7   to be infected by *Salmonella* than others. We termed them "vulnerable cells". The cell

8   vulnerability was characterized in a quantitative manner by automated high-content imaging

9   through double sequential infections with a delay of 1 to 3 h between the bacterial

10   challenges. The number of intracellular bacteria per cell as well as the corresponding host

11   cell parameters were assessed, such as cell perimeter, local density, and number of infected

12   neighboring cells. Using a mathematical model, we showed that host cell vulnerability can be

13   induced by a first bacterial uptake but also emerged from its intrinsic morphological and

14   micro-environmental characteristics.

1   **RESULTS**
2
3   **Sequential infections allow studies of *Salmonella* cooperation at the single cell level**
4
5   We carried out a microscopy-based double infection assay to explore possible links between
6   host cell vulnerability and successive bacterial infections of epithelial cells (**Fig.1**). HeLa cells
7   grown in 96-well plates were subjected to a first infection with green *S*. Typhimurium
8   expressing the fluorescent protein GFP (SL_{GFP}) for 30 min followed by elimination of the
9   extracellular bacteria via gentamicin treatment and washing. The cells were then incubated
10 for 1, 2 or 3 h before being subjected to a second wave of infection with red *S*. Typhimurium
11 expressing the fluorescent protein dsRed (SL_{dsRed}). Extracellular bacteria were again
12 eliminated in the same way, and the host cells were stained with CellMask and DAPI before
13 automated image acquisition of entire culture wells (**Fig.1A**). The obtained images were
14 analyzed with CellProfiler, a widely used image analysis software (1, 17) (**Fig.1B**). The
15 differently labeled bacteria and the stained host cells enabled us to distinguish and quantify
16 distinct cellular populations: those cells infected during the 1st infection (I_1) or not (noI_1), those
17 infected during the 2nd infection (I_2) or not (noI_2), as well as the associated subpopulations
18 ($I_1\&I_2$, $noI_1\&noI_2$, $I_1\&noI_2$ and $noI_1\&I_2$) (**Fig.1C**). We based our analysis on comparing
19 probabilities of infection in these subpopulations.
20
21 **Cooperation at the entry site during the presence of ruffles**
22
23 In order to test the reliability of our method, we analyzed first if we could reproduce
24 previously published results on the ruffle-dependent cooperation between individual
25 salmonellae during host cell entry (28, 25). To do this we determined first the time window
26 during which ruffle-associated cooperation could potentially occur performing time-lapse
27 microscopy of *Salmonella* infection of HeLa cells (**Fig. 1D**) and Caco-2 cells (data not shown)
28 transiently expressing GFP-tagged actin and labeled with the membrane dye FM 4-64
29 respectively. Time series of 90 min at 3 min intervals provided image sequences with forming
30 and disappearing ruffles. In most of the cases, we observed for both cell lines the uptake of
31 one to two bacteria per ruffle, and we saw ruffle disappearance in less than 15 min
32 (**Video.S1**). We noticed that the more bacteria were engulfed by the ruffles, the longer we
33 could detect the presence of these ruffles. Therefore, newly arriving bacteria prompted
34 additional growth of the ruffles (**Video.S2**). We quantified the ruffle lifetime by measuring the
35 delay of their disappearance after the entry of the last bacterium. The few cases of very high
36 infection (>5 bacteria/ruffle) that could not be properly analyzed were excluded.
37 Quantification revealed an average ruffle lifetime of 13 min and that 90% of the ruffles
38 completely disappeared after 24 min (**Fig.1D**). The results for Caco-2 cell infection were
39 similar to those of HeLa cells.

1

2   We then challenged HeLa cells with SL_{GFP} and SL_{dsRed} at the same time and compared the

3   probability for SL_{dsRed} to infect the same cell containing simultaneously SL_{GFP} with those that

4   did not contain SL_{GFP} (**Fig.1E** and **Fig.1F**); *see Materials and Methods for details*. The

5   repartition of the different populations of infected cells (**Fig.1E**) shows a much larger overlap

6   between the cells co-infected with SL_{GFP} and SL_{dsRed} than one would anticipate theoretically

7   for two independent infections. Thus, the efficiency of *Salmonella* invasion in an individual

8   epithelial cell depends on the concomitant invasion of the same cell by other salmonellae.

9   The difference of the concomitant infection probability (**Fig.1F**) was striking as it was 8 times

10   more likely for SL_{dsRed} to infect a cell also infected by SL_{GFP} than a cell not infected by SL_{GFP}.

11   Interestingly, increasing the multiplicity of infection (MOI) resulted in a significant increase of

12   the SL_{dsRed} infection in cells infected by SL_{GFP}, but not in cells not infected by SL_{GFP}. This

13   result confirmed that the direct effect of an MOI increase is a higher number of bacteria that

14   infect certain cells rather than an increase of the overall number of cells that become

15   infected. It underlines the relevance of ruffle-associated cooperation between salmonellae at

16   the entry site. Taken together, these results validated that our system was operational.

17
18   **The probability of being re-infected by *Salmonella* is higher for already-infected cells,**
19   **even after the disappearance of the entry ruffles**
20

21   To study long-term and ruffle-unrelated cooperative events of *Salmonella* co-infections, we

22   set the sequential infections with a delay of 1 h between the two infection waves, killing

23   extracellular bacteria in between through gentamicin treatment. Scanning our time-lapse

24   movies, we were ensured that this time lag led to the complete absence of any remaining

25   entry ruffles from the first infection. In addition, we extended the delay between the two

26   sequential infections to 2 h and 3 h (*see* **Fig.1A**). We compared the different populations of

27   cells infected during the 2nd infection (population I_2), depending on whether they were already

28   infected during the first wave of infection (population $I_2 \mid I_1$) or not (population $I_2 \mid noI_1$) for

29   HeLa (**Fig.2A**) and Caco-2 (**Fig.2B**) cells. For both tested cell types, it was significantly more

30   probable for a cell infected the 1st time to be re-infected the 2nd time compared to a cell not

31   previously infected. We propose that such cells are somehow more vulnerable for future

32   infection.

33

34   During all sequential infection experiments we also controlled the overall infection efficiencies

35   of SL_{GFP} and SL_{dsRed} at all measured time points (1st: SL_{GFP} - 2nd: SL_{dsRed} or in the reverse

36   order) (**Fig.S1**). In all cases, the percentage of cells infected by each fluorescent *Salmonella*

37   was similar for cells subjected to single (control) or sequential infections, underlining that

38   sequential infections did not change the overall infection efficiency for the differently colored

1   salmonellae. Nevertheless, we noticed a decrease of the amount of infected cells between

2   the early infection and later time points. This effect is most likely due to the technically

3   unavoidable gentamicin treatment between infections. Besides, SL_{GFP} showed a higher

4   infectivity than SL_{dsRed} for each condition explained by general deleterious effects of the

5   heterologously overexpressed fluorescent proteins on *Salmonella* infectivity, and by the

6   partial loss of dsRed expression observed by us and others. Taking into account these

7   issues, we took advantage of the observed consistency of the differences of infection

8   efficiency between the initial and the successive infections, and between SL_{GFP} and SL_{dsRed}.

9   This consistency allows comparative analyses of the ratio of the different infection

10 probabilities, and it provided us with an analytical tool for precise quantification independently

11 of the variances of the differently colored bacteria and technical hurdles of sequential

12 infection.

13

14 We defined a "vulnerability score" as the conditional probability for a cell to be infected during

15 the 2nd infection after it had been already infected during the 1st one ($I_2 \mid I_1$), divided by the

16 conditional probability for a cell to be infected during the 2nd infection when it had not been

17 previously infected ($I_2 \mid noI_1$) (described in more detail in *Materials and Methods*). We also

18 analyzed changes of the vulnerability score in time comparing cells subjected to sequential

19 infection with 1, 2 and 3 h delays (**Fig.2B** and **Fig.S2** for detailed representation of the

20 conditional probability for each replicate). Surprisingly, the vulnerability score appeared un-

21 altered. We obtained similar results inverting the order of the used pathogen, infecting first

22 with SL_{dsRed} and then with SL_{GFP} (**Fig.S3**). It was not possible to shorten the delay between

23 infections to less than 1 h due to the ruffle influence, and we could not extend it beyond 3 h

24 due to potential release of hyper-replicative (HR) bacteria from the first infection into the

25 extracellular medium that could then re-infect new cells during the 2nd wave of infection.

26 Altogether, these results showed that, after ruffle disappearance, the infected cells remain

27 more vulnerable to a new infection than the non-infected ones, and this vulnerability is stable

28 in time.

29
30
31 **Cell vulnerability to secondary infection can be predicted from the number of**
32 **intracellular bacteria**
33
34 So far, we only considered the character "infected" or "non-infected" for each cell after SL_{GFP}

35 and SL_{dsRed} infections that provides global trends on their interaction. To further exploit our

36 data we quantified the number of bacteria per host cell and related the obtained numbers

37 with the previously extracted vulnerability scores. The distribution of intracellular bacteria

38 inside infected cells at 2.5 h post-infection (pi) showed that most of the cells contained few

1   bacteria, and the proportion of cells with higher number of intracellular bacteria decreases

2   drastically. Overall, we were able to distinguish three groups of infected cells: the ones

3   containing one to two intracellular bacteria (35% of the global population), the ones

4   containing three to eight intracellular bacteria (39% of the global population) and the ones

5   containing more than nine intracellular bacteria (26% of the global population), corresponding

6   respectively to low, medium and high infections (**Fig.3A)**.

7

8   We compared the vulnerability score of these three infection groups during sequential double

9   infections (**Fig.3B**). This analysis revealed that the more bacteria had entered in a given host

10 cell during the first infection, the more it was likely that this cell became re-infected. Such

11 tendencies still emerged when the bacteria were not grouped, but analyzed individually,

12 underlining the robustness of this result (**Fig.S4**).

13

14 Then, we investigated how the level of bacterial uptake during the second infection depends

15 on the number of intracellular bacteria of the first infection. For this we quantified the

16 probability for a cell to be highly infected during the second infection as a function of the

17 efficiency of the first uptake (**Fig.3C**). We found that the more intracellular bacteria had been

18 internalized during the first infection, the more likely they were to ingest a high amount of

19 new bacteria during the second infection. Therefore, we propose that cell vulnerability is

20 maintained from the first to the second infection.

21
22 **Cell vulnerability as intrinsic or induced property**
23
24 The results from the sequential infections (**Fig.3** and **Fig.4**) provided quantitative scores of

25 cell vulnerability towards *Salmonella* infection. We secondly investigated the origin of the

26 observed cell vulnerability. Two possibilities can be anticipated: (i) the cellular vulnerability

27 would be an intrinsic host cell attribute (hypothesis 1: "intrinsic vulnerability") or (ii) it would

28 be induced by bacterial uptake (hypothesis 2: "induced vulnerability") (**Fig.4A**). In theory,

29 these hypotheses can be distinguish by the observable different probability of the 2nd wave of

30 infection occurring in previously non-infected cells $P(I_2 \mid noI_1)$ as depicted in the two schemes

31 of **Fig.4B** and described as follows: In the case of vulnerability as intrinsic attribute, the

32 probability of infection $P(I_2 \mid noI_1)$ would be lower than $P(I_{2Ctr})$ as the pool of vulnerable cells

33 would be already partially consumed during the 1st sequential infection, whereas it would

34 remain conserved in the control (**Fig.4B-left**). In the case of induced vulnerability, the

35 probability of infection $P(I_2 \mid noI_1)$ would be similar to $P(I_{2Ctr})$, as the cells would be considered

36 with equivalent vulnerabilities before their first infection (**Fig.4B-right**). The experimental

37 data obtained did not show a significant difference between $P(I_2 \mid noI_1)$ and $P(I_{2Ctr})$ (t-test p-

38 value >0,05) (**Fig.4C**), suggesting that vulnerability may be induced by bacterial uptake

1 (**Fig.4B,** *hypothesis 2*). Taking into account the small percentage of cells belonging to the

2 studied subpopulations we caution that the absence of a statistically significant difference

3 between these populations did not allow to exclude the first hypothesis of host cell inherent

4 vulnerability.

5

6 **Single cell vulnerability to *Salmonella* infection is a combination of intrinsic and**

7 **induced vulnerability**

8 Considering that the subpopulation comparison could not exclude an involvement of inherent

9 vulnerability, we developed a mathematical model to evaluate the relative contribution of

10 induced and inherent vulnerability to the overall cell vulnerability towards *Salmonella*

11 infection. To investigate the contribution of cell parameters at the single-cell level, we

12 measured different intrinsic variables that could influence cellular vulnerability, namely the

13 cell morphology (cell perimeters, cell circularity), the local environment (local cell density,

14 number of infected and non-infected neighboring cells), and the above-analyzed features of

15 the *Salmonella* infection (delay between infections, load of intracellular bacteria per cell from

16 I_1) (**Fig.5A**). We extracted all these elements using Icy, an image analysis software (6)

17 recently used for *Salmonella* infection studies *in situ* (27) (*see* **Fig.S5A** for illustration of Icy

18 cell segmentation).

19

20 First, we analyzed the distribution of distinct cellular parameters in either infected or in non-

21 infected HeLa (upper panels) and Caco-2 (lower panels) cell populations (**Fig.5A**). For both

22 cell types, the infected cells displayed distinct cellular features in comparison to the non-

23 infected cells, such as a higher local crowding reflected by a higher number of neighboring

24 cells in direct contact. Comparing the relative correlations of the cellular parameters, we

25 highlight the presence of strong links between a number of them (**Fig.S5B-C, S6 and S7**). In

26 particular cell morphology is highly dependent on the local micro-environment, such as the

27 local cell density that negatively correlates with the cell perimeter in HeLa and Caco-2 cells.

28 Interestingly, cells infected during the second bacterial challenge are more likely to be nearby

29 cells infected during the first bacterial challenge ("infected neighbor cells") than by non-

30 infected neighbor cells. Thus *Salmonella* infection of one cell increases the probability of its

31 neighboring cells to be subsequently infected.

32

33 To quantify the direct involvement of each studied parameter on the overall cell vulnerability

34 we developed a statistic modeling approach adapted to our high-throughput microscopy

35 dataset on sequential *Salmonella* infection. This model is based on a logistic regression able

36 to predict the infection efficiency at the single cell level from cellular parameters. We

1   measured the contribution of each parameter for the prediction by estimating how well the

2   model predicts compared to a model that would ignore one parameter; as described in

3   *Materials and Methods* (**Fig.5B**). Taken separately, the load of intracellular bacteria resulting

4   from I_1 directly improved the prediction of cell vulnerability towards subsequent infection

5   (**Fig.5B**). Thus, host cell vulnerability is induced by bacterial uptake, which is in line with our

6   experimental data. In addition, the host cell parameters linked with cell morphology and local

7   environment also significantly improved the model prediction of infection for HeLa and for

8   Caco-2 cells (*see* **Table1** and **Table2** *for model details and the value of the coefficients*).

9   Together, our modeling approach revealed that single host cell vulnerability to *Salmonella*

10   infection is a combination of intrinsic and bacterial-induced vulnerability.

11   We quantified their relative involvement by calculating the model-based fold change of the

12   probability of infection for a cell not infected and having a low score of inherent vulnerability

13   with a cell infected and/or having a high score of inherent vulnerability (**Fig.5C**). This showed

14   that induced and intrinsic vulnerability have both a strong impact on the overall cell

15   vulnerability. Interestingly, the induced vulnerability is more prevalent for *Salmonella* infection

16   of HeLa cells (2,2 fold-increase) than infection of Caco-2 cells (1,3 fold-increase), whereas

17   the inherent vulnerability plays a more prominent role for Caco-2 cell infections (2,6 fold-

18   increase) than for HeLa cells (1,6 fold-increase). From these findings we conclude that the

19   analyzed host cell parameters are differentially involved in relation to cell vulnerability

20   towards *Salmonella* infection depending of the cell type. In particular, the local cell density

21   increases the cell vulnerability for HeLa cells but decreases it for Caco-2 cells (**Fig.5D**). This

22   could be explained by the polarization of the Caco-2 at high confluency and highlights the

23   specificity of each predicted model for a given cell-type.

24   We also investigated whether the first infection affects the inherent host cell parameters, we

25   compared the correlation between parameters that were identified as being either involved or

26   not involved in the inherent vulnerability of the cell. As their correlations were similar in

27   infected and non-infected cells (data not shown) we concluded that *Salmonella* infection did

28   not impact the implication of the studied inherent cell parameters.

29

30   **Reliability of the model-based prediction of infection**

31   To investigate the spatial distribution of the cell vulnerability among the cell population, we

32   created "vulnerability maps" from original images of the cell population after labeling each

33   cell nucleus with a color corresponding to its probability of infection (**Fig.6A**). Notably, we

34   could confirm that on average the infected cells were properly assigned with a higher

35   prediction score to be infected than the non-infected ones (*see* **Fig.S8** *for quantification*).

36   Based on our vulnerability maps, the predicted infected cells showed a very good overlap or

1 were in close vicinity with the experimentally infected cells (**Fig.6A**). This illustrates the

2 reliability of our approach in a qualitative way, and it also underlines the impact of local

3 micro-environment on cell vulnerability. We went on and quantified the veracity of the HeLa

4 and Caco-2 adapted models when confronted with 100 experimentally measured infected

5 and 100 experimentally measured non-infected cells. For both cell-types, models allowed a

6 good prediction in the majority of the cases, 62% for HeLa and 66% for Caco-2 respectively

7 (**Fig.6B**). Taken together, these results attest that the probability of *Salmonella* infection

8 success can be forecast at the near single-cell level based on host cell parameters.

9

10 **Involvement of cellular cholesterol rate as an inherent vulnerability factor**

11 To investigate molecular players that are linked with the inherent cell vulnerability to

12 *Salmonella* infection, we analyzed the plasma membrane composition as main feature

13 known to be relevant for *Salmonella* infection. We focused on cholesterol as cells at low

14 crowding present a higher amount of free cholesterol than cells at high crowding (9). We

15 monitored the relation between global cellular cholesterol levels and host cell targeting

16 performing *Salmonella* infection of HeLa cells for 30 min followed by cholesterol labeling via

17 filipin staining and flow cytometry analysis (**Fig.7**). For each experiment, we binned the total

18 cell population into five subpopulations corresponding to increasing cellular levels of

19 cholesterol that we classified as 1 to 5, with each subpopulation containing 20% of the total

20 cells (*see **Fig.S9** for FACS gating details*). Comparing the number of infected cells in these

21 different subpopulations with different amounts of cholesterol, we revealed that the

22 probability of infection decreased with increasing cholesterol levels. From this we conclude

23 that cells with a lower amount of cholesterol are preferentially targeted by *Salmonella*

24 compared to those with higher cholesterol levels.

1   **DISCUSSION**

2

3   Cellular heterogeneity describes cases in which genetically identical cells present different

4   behaviors and morphologies. This biological phenomenon is commonly present in an

5   epithelial layer of an individual as well as within a monolayer of cultured cells. Despite the

6   realization of the importance of cellular heterogeneity, its study has only become feasible

7   during recent years, mainly thanks to the implementation of novel technologies such as

8   imaging and computer-assisted analysis. In the context of pathogen infection, this

9   heterogeneity produces cells unequally vulnerable or resistant which impacts on the overall

10   infection.

11

12   We investigated the cell vulnerability of epithelial cells for *S*. Typhimurium infection.

13   According to our results, infected cells display a strikingly higher probability of being re-

14   infected with *Salmonella*, even after disappearance of membrane ruffles. We obtained similar

15   results in two relevant epithelial cell lines, HeLa and Caco-2, suggesting that this represents

16   a conserved propensity towards *Salmonella* infection. The measured cellular vulnerability

17   remained unaltered for all measured time-points ranging from a delay of 1 h to 3 h between

18   the infections. Attributing a "vulnerability score" to the challenged cells, we showed a higher

19   vulnerability score in cells previously infected, and we found that this score increases with

20   the amount of intracellular bacteria contained by a given cell. This result raises the issue of

21   the bacterial impact on the cell vulnerability. Therefore, we aimed at distinguishing inherent

22   cell vulnerability from the one induced by bacterial uptake (**Fig.4A,** *hypothesis 1 and 2*

23   *respectively*) exploiting the imaging data obtained via a high-content analytical pipeline. This

24   allowed visualization of the infection *in situ* and provided a large number of associated

25   cellular parameters. We quantified the implication of specific parameters associated with

26   individual cells on the cell vulnerability towards *Salmonella* infection. It appeared clearly that

27   the efficiency of early bacterial uptake during the first infection directly determines cell

28   vulnerability. Thus *Salmonella* induce an increase of the cell vulnerability toward subsequent

29   infections.

30   While long-term cooperation between bacteria has been intensively studied for communities

31   of bacteria living in a common extracellular environment (5), little is known about the

32   cooperation between intracellular and extracellular bacteria leading to increased bacterial

33   uptake. Nevertheless, this phenomenon has been investigated more extensively for many

34   viruses, including bacteriophages (4), influenza virus (14), poxviruses (7, 21), flaviviruses

35   (39, 34), alphaviruses (18), and alphaherpesviruses (3). Generally, those works have

36   demonstrated that the first virus to infect a cell has the capacity to prevent co-infection of

37   other viruses belonging either to the same strain, or to more distantly related or unrelated

1   strains. It is termed "superinfection exclusion" and may protect limited cellular resources and

2   promote the replication and dissemination of the originally infecting virus. By analogy, the

3   increased probability of cellular re-infection by *Salmonella* can be phrased as a

4   "superinfection promotion". It remains to be clarified if such process is relevant for all

5   intracellular bacteria. For instance and in contrast to *Salmonella* infection, Jorgensen *et al*

6   reported that the *Chlamydia* effector protein CPAF secreted from bacteria within mature

7   inclusions prevents those that are still extracellular to invade (16). Thus, CPAF could be a

8   factor mediating *Chlamydia* resistance towards superinfection.

9

10 Our approach also allowed the relative quantification of the impact of different host cell

11 parameters on the inherent vulnerability of host cell to *Salmonella* infection. In particular,

12 morphological attributes and local cell crowding are highly linked with this vulnerability. Cell

13 crowding as major determinant for the probability to become infected has been proposed by

14 Snijder and colleagues in the context of viral infection. They showed that during infections by

15 the simian virus SV40 or the mouse hepatitis virus (MHV), the targeted cells have different

16 localization within cell islets (35). SV40 and MHV infect preferentially either peripheral or

17 central cells, a phenomenon that is linked with the differential expression levels of focal

18 adhesion kinase and the presence of sphingolipid GM1 at the plasma membrane of

19 challenged host cells. Thus, similarly to several viral infections, the probability of infection of

20 a single cell by *Salmonella* is influenced by its local environment.

21

22 Our analytical tools will be useful for further studies on *Salmonella*, and for other researchers

23 working on different intracellular bacteria pathogens, such as *Chlamydia*, *Listeria* or *Shigella*

24 (*see Materials and Methods*). We revealed that some cells are indeed intrinsically more

25 vulnerable to *Salmonella* and will be targeted by the bacteria first. Most of the tested

26 parameters appeared relevant for model-based infection prediction but are differentially

27 involved in the cell vulnerability depending on the cell-type studied. Developing an adapted

28 model based on host cell parameters we could forecast the probability of *Salmonella*

29 infection success at the near single-cell level. Interestingly, the number of infected

30 neighboring cells is highly increased in the population of infected cells. Cases of bacterial

31 uptake impacting on the cells neighboring the infection (called bystander cells) have been

32 previously reported for *Shigella* that induces an IL-8 immune response after NFκB activation

33 detectable from 2 h pi in 70% of the bystander cells (19). However, it is not known whether

34 the neighboring cells are also more sensitive towards *Shigella* entry.

35

36 Because of our incomplete knowledge of host factors that are involved in the early

37 attachment, such as potential entry receptors, it remains difficult to identify the molecular

13

1 mechanisms that establish the differential vulnerability during *Salmonella* infection. Although

2 receptors for direct recognition of *Salmonella* have been proposed, such as the cystic fibrosis

3 transmembrane conductance regulator (CFTR) (31) and the epithelium growth factor

4 receptor (EGFR) (30), many cell types infected by *Salmonella* do not express them (15).

5 Therefore, it has been proposed that recognition mechanisms likely involve more ubiquitous

6 factors (10). To explore molecular cues involved in the inherent heterogeneity of host cell

7 vulnerability, we decided to investigate the membrane lipid composition, in particular cellular

8 cholesterol. This was based on previous studies showing that the amount of free cholesterol

9 per cell negatively correlates with the local cell crowding (9). We found that an increase of

10 cholesterol amount at single cell level correlates with a lower vulnerability of the cell, so that

11 *Salmonella* preferential target host cells with low amounts of cholesterol. Intriguingly, these

12 results on an implicated host molecule are in agreement with the morphological feature of

13 local density; HeLa cells at high density contain lower amounts of cholesterol and display an

14 increased inherent cell vulnerability. The role of cholesterol during *Salmonella* infection is still

15 debated. Several studies have demonstrated that the *Salmonella* SipB effector and

16 translocon component requires cholesterol for proper functioning (13, 10). In this context, it

17 should be noted that the translocons operate in small cholesterol-rich microdomains at the

18 plasma membrane and cannot be linked readily with the overall cholesterol levels.

19 Furthermore, those studies were based on sterol sequestering agents and biosynthesis

20 inhibitors. Contrastingly, Gilk and colleagues showed that cholesterol is not essential for

21 *Salmonella* invasion and intracellular replication inside host cells using an original mouse

22 model, (11). In our study we highlighted that non-treated cells with a low amount of global

23 cellular cholesterol are preferentially targeted by *Salmonella,* which does not exclude a

24 potential involvement of cholesterol at the subcellular level.

25

26 In conclusion, our study represents a first step in understanding *Salmonella* cell targeting and

27 provides a path for the identification of cellular and bacterial factors involved in host cell

28 vulnerability. Such factors could be targeted to render a cell more resistant to pathogen

29 infections, allowing potential new therapeutic strategies. Together, our study delineates in a

30 quantitative manner the importance of vulnerable cell recognition and bacterial cooperation

31 for cell targeting by *S*. Typhimurium.

1 **MATERIALS AND METHODS**

2

3 **Bacterial Strains**

4 The following *S.* Typhimurium were used: SL1344 (wild type), SL1344 pM965 (*Salmonella*-

5 GFP) described by Stecher *et al* (36), and SL1344 pGG2 (*Salmonella*-dsRed) obtained after

6 transformation of SL1344 with the pGG2 plasmid described by Lelouard *et al* (23). Bacteria

7 were grown in Lysogeny Broth (LB) medium supplemented with 0.3 M NaCl and ampicillin at

8 50 µg/ml at 37ºC in an orbital shaker.

9

10 **Cell Culture**

11 All cell culture reagents were purchased from Invitrogen unless otherwise stated. Human

12 epithelial HeLa cells (clone CCL-2 from ATCC) were cultured in Dulbecco's Modified Eagle's

13 Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS), at 37 ºC, 5% CO_2.

14 Intestinal epithelial Caco-2 TC7 cells (kindly provided by P. Sansonsetti) were grown in

15 DMEM supplemented with 10% FBS at 37°C, 10% CO_2. All infection assays were performed

16 in EM buffer (120 mM NaCl, 7 mM KCl, 1.8 mM $CaCl_2$, 0.8 mM $MgCl_2$, 5 mM glucose, 25 mM

17 HEPES, pH 7.4). HeLa cells were transfected with pEGFP-actin plasmid DNA (8) from a

18 maxiprep, using the X-tremeGENE 9 DNA transfection reagent (Roche) for 48 h.

19

20 **Double Infection Assays**

21 For invasion experiments, overnight bacterial cultures were sub-cultured 1/20 and grown until

22 late exponential/early stationary phase. Before infection, bacteria were gently washed and

23 resuspended in EM buffer. Bacteria were added to the cells at an MOI of 30 corresponding to

24 CFU, and incubated for 30 min at 37 ºC, 5% CO_2. Non-internalized bacteria were eliminated

25 by washing 3 times with warm EM buffer and incubated for 1, 2 or 3 h at 37 ºC, 5% CO_2.

26 Adding EM buffer containing 100 µg/ml gentamicin for 1 h killed extracellular bacteria. The

27 concentration of gentamicin was then decreased to 10 µg/ml and 10% FBS was added to the

28 medium. At the desired time points, the cells were washed again in EM buffer to eliminate

29 the remaining gentamicin and re-infected with a fresh batch of sub-cultured bacteria following

30 the same protocol. After killing the extracellular bacteria again by a 1 h of incubation with EM

31 buffer containing 100 µg/ml gentamicin, the cells were fixed with 4% paraformaldehyde at

32 room temperature for immunofluorescence analysis.

33

34 **Microscopy**

35 All image acquisitions were performed on a Nikon inverted widefield microscope using a

36 20x/0.5NA air objective, an automatic programmable XY-stage and the Nikon perfect focus

37 system. For sequential infections, 161 fields were imaged per well and four channels per field

15

1   were captured using a CoolSnap2 camera (Roeper Scientific). Nuclei and cells were stained

2   using DAPI (excitation/emission wavelengths: 350/470 nm) and the cell bodies with CellMask

3   DeepRed Plasma Membrane Stain (ThermoFisherScientific, excitation/emission

4   wavelengths: 640/670 nm) respectively. Caco-2 cells were stained with the FM® 4-64

5   membrane dye (Invitrogen) before time lapse imaging (excitation/emission wavelengths:

6   558/734 nm). Quantification of the ruffle timing was performed on the same microscope,

7   using a 20x/0.5NA air objective and time intervals of 3 min for 90 min. Time lapse imaging of

8   ruffles was performed on a DeltaVision widefield microscope using a 60x/1.42 NA oil

9   objective and z-stacks with a spacing of 500 nm. The images were subsequently de-

10   convolved using DeltaVision Elite integrated software.

11

12   **Cholesterol measurements**

13   HeLa cells were challenged with SL_{GFP} for 30 min before trypsinization, fixation with 4%

14   paraformaldehyde and incubation with 16ug/mL filipin complex from *Streptomyces filipinensis*

15   (Sigma-Aldrich). This treatment was directly followed by FACS measurement on *BD FACS*

16   *CANTO* cytometer using the excitation/emission wavelengths of 405/450 nm and 488/530

17   nm for filipin and GFP fluorescence detection respectively. Infected and non-infected cells

18   were distinguished using the green fluorescence emitted by SL_{GFP} (*see* **Fig.S9** *for gating*

19   *details*). Data were processed using FlowJo software.

20

21   **Image Analysis**

22   All images were analyzed with two open source software: CellProfiler (http://cellprofiler.org/)

23   and Icy (http://icy.bioimageanalysis.org/). CellProfiler was used to detect each single cell

24   and the number of its intracellular salmonellae expressing either GFP or dsRed. The

25   following modules were used during the analysis: *IdentifyPrimaryObjects* recognized nuclei

26   and bacteria; *IdentifySecondaryObjects* identified cells (here the secondary objects) by

27   extending the nuclear area previously recognized; *RelateObjects* assigned bacteria within

28   individual cells. Icy was used for accurate detection of cell borders and the cellular

29   microenvironment analysis. We used a graphical environment called *Protocols* for the

30   development of an analytical pipeline including the following plugins: *HK-Means* that identify

31   nuclei by pre-filtering the signal to identify objects within a size range; *Spot Detector* that

32   identify bacteria; *Active Contours* that identify the edges of the plasma membrane by

33   propagating the Region of Interest (ROI) detected for the nuclei; and *Javascript* that parent

34   the ROI of cells with bacteria, to measure local cell density and to distinguish which

35   neighboring cells are infected by which bacteria.

36

37

1 **Probability**

2 $P(I_2|I_1)$ means "Probability of the 2nd sequential infection, knowing that the cell has been

3 infected by the 1st one" and is calculated as follows:

$$P(I_2|I_1) = P(I_1\&I_2)/P(I_1)$$

4 Where $P(I_1)$ = [Number of cells in I_1 / Total number of cells], and $P(I_1\&I_2)$ = [Number of cells in

5 $I_1\&I_2$ / Total number of cells].

6

7 $P(I_2|noI_1)$ means "Probability of the 2nd sequential infection, knowing that the cell has not

8 been infected by the 1st one" and is calculated as follows:

$$P(I_2|noI_1) = P(I_2\&noI_1)/P(noI_1)$$

9 Where $P(noI_1)$ = [Number of cells in noI_1 / Total number of cells], and $P(I_2\&noI_1)$ = [Number of

10 cells in $noI_1\&I_2$ / Total number of cells].

11

12 **Model**

13 We modeled the influence of multiple parameters on the probability of a second infection. A

14 Boolean variable Y represents the second infection: It is equal to 1 for infected cells and 0

15 otherwise. Its probability is predicted by the following seven parameters: *Load of infection*

16 (LOI) represents the number of infecting bacterium during the first infection, separated in 4

17 groups corresponding to no (0 bacteria), low (1 or 2), medium (3 to 8) or high (9+)

18 infection. *Delay* is a categorical variable corresponding to the delay between the 1st and the

19 2nd infections (1, 2 or 3 h). *Infected neighbor cells* (X1) refers to the number of cells in

20 contact that had been infected during the first infection. *Non-Infected neighbor cells*

21 (X2) refers to the number of cells in contact which had not been infected during the first

22 infection. *Local Cell Density* (X3) is the number of cells present in a vicinity of 100 µm. The

23 distance is calculated between the center of the nuclei. *Cell perimeter* (X4) is the length of

24 the perimeter of the cell (in µm) obtained after segmentation. *Circularity* (X5) refers to the cell

25 circularity defined as: "$4\pi*area/perimeter^2$". This parameter is higher for circular cells, and

26 lower for cells that are elongated or have complex shape, but does not depend *a priori* on the

27 cell size. In practice we used to its square root. The probability of Y during the second

28 infection is modeled as:

$$P(Y = 1 | X1,\ldots,Xn) = 1 / [1 + exp(-(aLOI + aDelay + a1\,X1 + \ldots + a5\,X5))]$$

29 where aLOI (resp. aDelay) has a different value for each of the LOI categories (resp. Delay

30 categories), and $a_1,\ldots,a_5$ are constants. All parameters were learned by maximizing the

31 likelihood of the model, e.g. the probability of the observed data as measured by the

32 model. We used 115 000 and 327 000 cells to train and test the model for HeLa and Caco-2

33 cells respectively. We divided the cell population into two random sets; the training set

34 ($9/10^{th}$ of the cells per replicate) and testing set ($1/10^{th}$ of the cells) and computed the

35 likelihood of infection observed in the testing set. The higher the likelihood, the better the

17

1 parameters of the model predicted infection. We repeated this procedure 100 times. To

2 measure the improvement of infection prediction by taking into account each parameter, the

3 likelihood of the complete model was compared (on a log scale) with the likelihood of seven

4 models ignoring each time one parameter. This difference of log-likelihood is reported

5 in **Fig.5B**.

6 Quantification of the impact of a parameter towards cell vulnerability was obtained by

7 applying our statistical model to the 1st and the 3rd quantile values of a given parameter, while

8 other parameters were kept equal at their median values. We obtained the probabilities of

9 the second infection for these two sets and reported their ratio. In **Fig.5D**, the arrows "↗" and

10 "↘" correspond to a ratio above and under 1 respectively. The parameters-values

11 corresponding to a low inherent vulnerability of HeLa and Caco-2 cells were the following:

12 local cell density (1st quantile and 3rd quantile respectively), cell perimeter (1st quantile),

13 infected neighboring cells (median), non-infected neighboring cells (median), circularity

14 (median and 3rd quantile respectively). The parameters-values corresponding to a high

15 inherent vulnerability of HeLa and Caco-2 cells were the following: local cell density (3rd

16 quantile and 1st quantile respectively), cell perimeter (3rd quantile), infected neighboring cells

17 (median), non-infected neighboring cells (median), circularity (median and 1st quantile

18 respectively).

19 Models reliability was evaluated using 100 infected and 100 non-infected cells and

20 quantifying the amount of "good predictions" among those cells. We repeated this procedure

21 100 times and showed the average. As a comparison, a random model would provide

22 approximately 50% of "good predictions".

23

24 **Statistical analysis**

25 The statistical analysis was performed using R and GraphPad Prism. T-tests were used to

26 evaluate the significance of the results, referred like *, **, *** for p-values <0.05, <0.01, and

27 <0.001, respectively.

28

29 **Data availability**

30 The pipeline used on CellProfiler and on ICY, as well as the R code used to generate the

31 model can be provided by the authors.

REFERENCE

1. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7,** R100 (2006).

2. Carter, P. B. & Collins, F. M. The route of enteric infection in normal mice. *J. Exp. Med.* **139,** 1189–203 (1974).

3. Criddle, A., Thornburg, T., Kochetkova, I., DePartee, M. & Taylor, M. P. gD-Independent Superinfection Exclusion of Alphaherpesviruses. *J. Virol.* **90,** 4049–4058 (2016).

4. Cumby, N., Davidson, A. R. & Maxwell, K. L. The moron comes of age. *Bacteriophage* **2,** 225–228 (2012).

5. Davey, M. E. & O'toole, G. A. Microbial biofilms: from ecology to molecular genetics. *Microbiol. Mol. Biol. Rev.* **64,** 847–67 (2000).

6. de Chaumont, F. *et al.* Icy: an open bioimage informatics platform for extended reproducible research. *Nat. Methods* **9,** 690–696 (2012).

7. Doceul, V., Hollinshead, M., van der Linden, L. & Smith, G. L. Repulsion of superinfecting virions: a mechanism for rapid virus spread. *Science* **327,** 873–876 (2010).

8. Ehsani, S. *et al.* Hierarchies of host factor dynamics at the entry site of Shigella flexneri during host cell invasion. *Infect. Immun.* **80,** 2548–2557 (2012).

9. Frechin, M. *et al.* Cell-intrinsic adaptation of lipid composition to local crowding drives social behaviour. *Nature* **523,** 88–91 (2015).

10. Garner, M. J., Hayward, R. D. & Koronakis, V. The Salmonella pathogenicity island 1 secretion system directs cellular cholesterol redistribution during mammalian cell entry and intracellular trafficking. *Cell. Microbiol.* **4,** 153–165 (2002).

11. Gilk, S. D. *et al.* Bacterial Colonization of Host Cells in the Absence of Cholesterol. *PLoS Pathog.* **9,** (2013).

12. Haraga, A., Ohlson, M. B. & Miller, S. I. Salmonellae interplay with host cells. *Nat. Rev. Microbiol.* **6,** 53–66 (2008).

13. Hayward, R. D. *et al.* Cholesterol binding by the bacterial type III translocon is essential for virulence effector delivery into mammalian cells. *Mol. Microbiol.* **56,** 590–603 (2005).

14. Huang, I. C. *et al.* Influenza A virus neuraminidase limits viral superinfection. *J Virol* **82,** 4834–4843 (2008).

15. Jones, B. D., Paterson, H. F., Hall, a & Falkow, S. Salmonella typhimurium induces membrane ruffling by a growth factor-receptor-independent mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 10390–10394 (1993).

16. Jorgensen, I. *et al.* The chlamydia protease CPAF regulates host and bacterial proteins to maintain pathogen vacuole integrity and promote virulence. *Cell Host Microbe* **10,** 21–32 (2011).

17. Kamentsky, L. *et al.* Improved structure, function and compatibility for cellprofiler: Modular high-throughput image analysis software. *Bioinformatics* **27,** 1179–1180 (2011).

18. Karpf, a R., Lenches, E., Strauss, E. G., Strauss, J. H. & Brown, D. T. Superinfection exclusion of alphaviruses in three mosquito cell lines persistently infected with Sindbis virus. *J. Virol.* **71,** 7119–7123 (1997).

19. Kasper, C. A. *et al.* Cell-cell propagation of NF-??B transcription factor and MAP kinase activation amplifies innate immunity against bacterial infection. *Immunity* **33,** 804–816 (2010).

20. Knodler, L. A. Salmonella enterica: Living a double life in epithelial cells. *Curr. Opin. Microbiol.* **23,** 23–31 (2015).

21. Laliberte, J. P. & Moss, B. A novel mode of poxvirus superinfection exclusion that prevents fusion of the lipid bilayers of viral and cellular membranes. *J. Virol.* **88,** 9751–68 (2014).

22. LaRock, D. L., Chaudhary, A. & Miller, S. I. Salmonellae interactions with host processes. *Nat. Rev. Microbiol.* **13,** 191–205 (2015).

23. Lelouard, H. *et al.* Pathogenic Bacteria and Dead Cells Are Internalized by a Unique Subset of Peyer's Patch Dendritic Cells That Express Lysozyme. *Gastroenterology* **138,** 173–184.e3 (2010).

24. Liberali, P., Snijder, B. & Pelkmans, L. Single-cell and multivariate approaches in genetic perturbation screens. *Nat. Rev. Genet.* **16,** 18–32 (2014).

25. Lorkowski, M., Felipe-López, A., Danzer, C. A., Hansmeier, N. & Hensel, M. Salmonella enterica invasion of polarized epithelial cells is a highly cooperative effort. *Infect. Immun.* **82,** 2657–2667 (2014).

26. Majowicz, S. E. *et al.* The Global Burden of Nontyphoidal Salmonella Gastroenteritis. *Clin. Infect. Dis.* **50,** 882–889 (2010).

27. McQuate, S. E. *et al.* Long-term live-cell imaging reveals new roles for *Salmonella* effector proteins SseG and SteA. *Cell. Microbiol.* (2016). doi:10.1111/cmi.12641

28. Misselwitz, B. *et al.* Near surface swimming of salmonella Typhimurium explains target-site selection and cooperative invasion. *PLoS Pathog.* **8,** 9 (2012).

29. Misselwitz, B. *et al.* Salmonella enterica serovar typhimurium binds to hela cells via fim-mediated reversible adhesion and irreversible type three secretion system 1-mediated docking. *Infect. Immun.* **79,** 330–341 (2011).

30. Pace, J., Hayman, M. J. & Galán, J. E. Signal transduction and invasion of epithelial cells by S. typhimurium. *Cell* **72,** 505–514 (1993).

31. Pier, G. B. *et al.* Salmonella typhi uses CFTR to enter intestinal epithelial cells. *Nature* **393,** 79–82 (1998).

32. Santos, A. J. M., Meinecke, M., Fessler, M. B., Holden, D. W. & Boucrot, E. Preferential invasion of mitotic cells by Salmonella reveals that cell surface cholesterol is maximal during metaphase. *J. Cell Sci.* **126,** 2990–2996 (2013).

33. Santos, J. C. & Enninga, J. At the crossroads: Communication of bacteria-containing vacuoles with host organelles. *Cell. Microbiol.* **18,** 330–339 (2016).

34. Schaller, T. *et al.* Analysis of hepatitis C virus superinfection exclusion by using novel fluorochrome gene-tagged viral genomes. *J. Virol.* **81,** 4591–603 (2007).

35. Snijder, B. *et al.* Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* **461,** 520–523 (2009).

36. Stecher, B. *et al.* Flagella and Chemotaxis Are Required for Efficient Induction of Salmonella enterica Serovar Typhimurium Colitis in Streptomycin-Pretreated Mice Flagella and Chemotaxis Are Required for Efficient Induction of Salmonella enterica Serovar Typhimurium Colitis. *Infect Immun* **72,** 4138–4150 (2004).

37. Vonaesch, P. *et al.* Quantitative insights into actin rearrangements and bacterial target site selection from SalmonellaTyphimurium infection of micropatterned cells. *Cell. Microbiol.* **15,** 1851–1865 (2013).

38. Watson, K. G. & Holden, D. W. Dynamics of growth and dissemination of Salmonella in vivo. *Cell. Microbiol.* **12,** 1389–1397 (2010).

39. Zou, G. *et al.* Exclusion of West Nile virus superinfection through RNA replication. *J. Virol.* **83,** 11765–11776 (2009).

Acknowledgments

Author contributions

VS and JE designed the study and initiated the project. VS, JE and AD supervised the project. *Salmonella* infection were conducted by JV and VS. The automatic image analysis method was designed by VS, JV and AD. Flow cytometry measurements were achieved by IB and VS. The statistical analysis method was conceived by CG, JV and VS and performed by JV. VS and JE wrote the manuscript. JE provided the funding.

Conflict of interest

The authors declare that they have no conflict of interest.

22
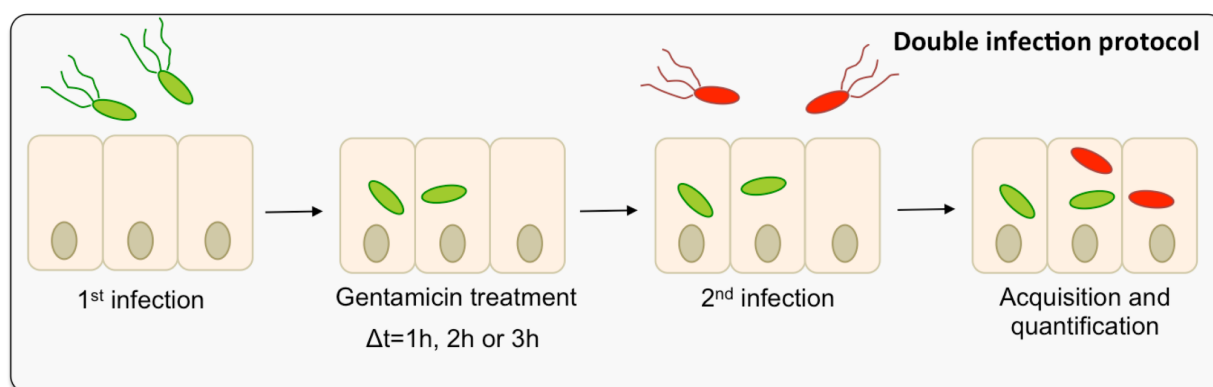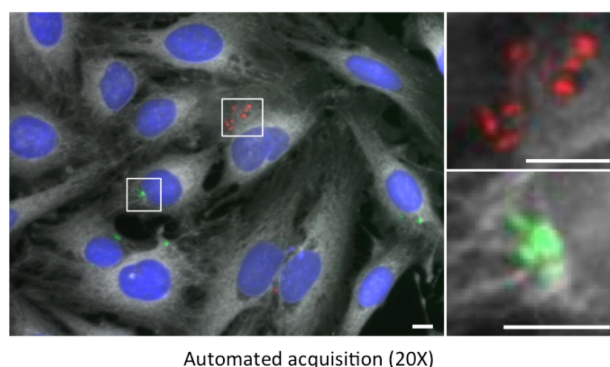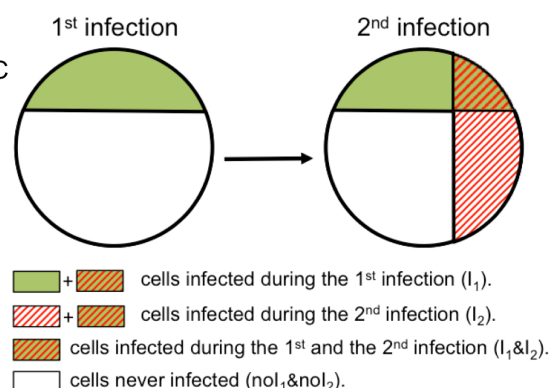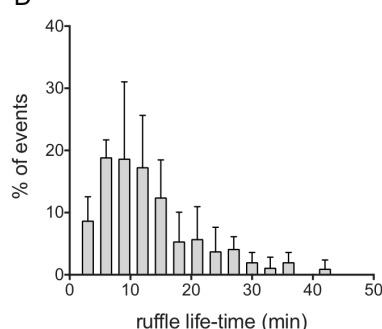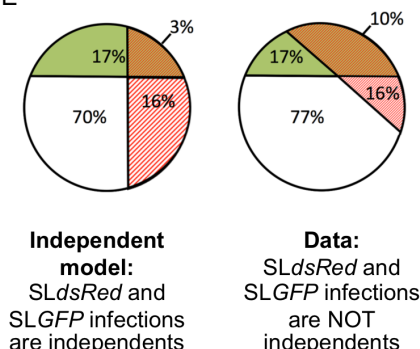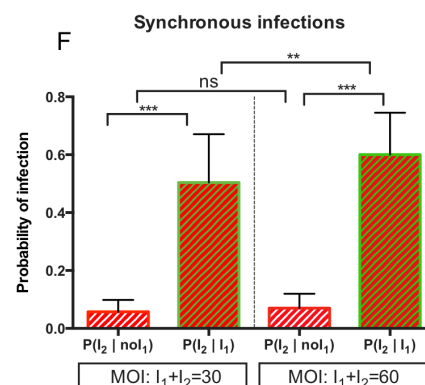
Fig.1. Double infections allow studies of *Salmonella* cooperation at the single cell level. A. B. C. Overview of the experimental workflow used in this study. **A.** Sequential infection protocol: HeLa cells grown in 96-wells plates since 24 h were subjected for 30 min to a first infection by SL_{GFP}. This was followed by elimination of extracellular bacteria by gentamicin and incubation of the cells for 1, 2 or 3 h. The cells were subsequently challenged by a second infection with SL_{dsRed} for 30 min. After removal of the extracellular bacteria, the samples were fixed. Nuclei were stained with DAPI and cell membranes were stained with CellMask before microscopic acquisition of the entire wells. **B.** Representative image of SL_{GFP} and SL_{dsRed} internalized in HeLa cells. Host cell nuclei are visible through DAPI (in blue), and cell membranes through CellMask (in grey). Scale bar correspond to 5μm. **C.** Scheme of our statistical analysis of different subpopulations. The following cellular populations can be distinguished: those cells infected during the 1st infection (I_1) or not (noI_1), those infected during the 2nd infection (I_2) or not (noI_2), along with the related subpopulations ($I_1\&I_2$, $noI_1\&noI_2$). This scheme maps the case of two independent infections. **D.** Time distribution of the ruffle disappearance during *Salmonella* infection followed in actin-GFP transfected cells by time-lapse microscopy. **E.** Comparison of an independent model (left) with the obtained data (right). The percentages are averaged from 6 independent experiments, represented in **C** with an MOI of 30. **F.** Comparison of the conditional probability of infection for two different populations during synchronous infection of SL_{GFP} and SL_{dsRed}. The MOIs were calculated after averaged CFU counting from 6 different experiments. P-values were obtained after paired t-test.
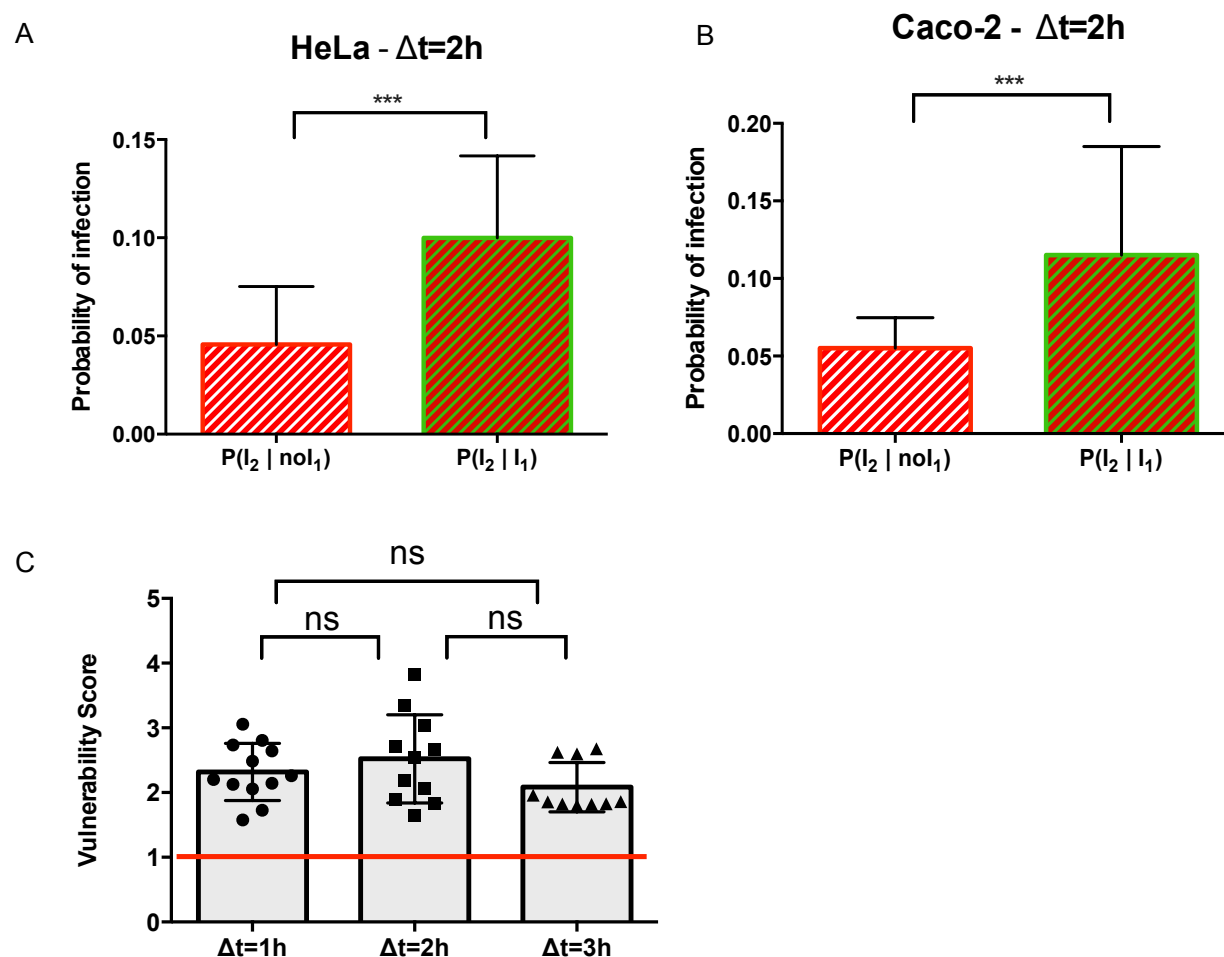
Fig.2. The probability of being re-infected by *Salmonella* remains higher for already infected cells after entry ruffle disappearance. A-B. Conditional probability of infection for two different populations during sequential infection with a delay of 2 h for HeLa cells (**A**) and Caco-2 cells (**B**). Results were obtained from 3 independent experiments and P-values were obtained after paired t-test. **C.** The vulnerability score was plotted for infection with a 1, 2 or 3 h delay before the second infection. The red line corresponds to $P(I_2 \mid I_1)=P(I_2 \mid noI_1)=1$ indicating the independence of the infections I_2 and I_1. Values above the red line correspond to $P(I_2 \mid I_1) > P(I_2 \mid noI_1)$ indicating a cooperation between infections. Values below the red line correspond to $P(I_2 \mid I_1) < P(I_2 \mid noI_1)$ indicating a competition between infections. Results were obtained from 3 independent experiments per time-point, and P-values were obtained after unpaired t-test.
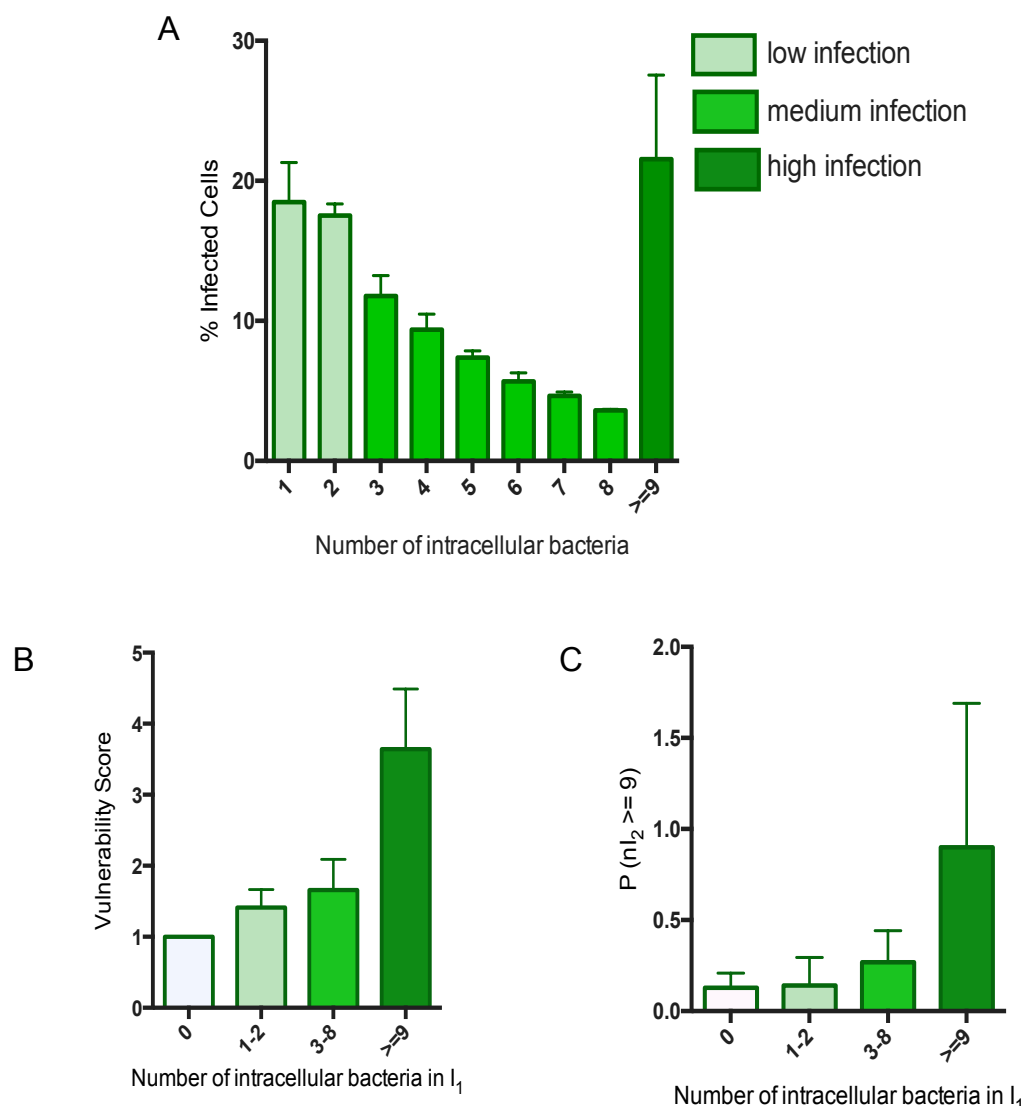
Fig.3. Cell vulnerability can be predicted from the number of bacteria previously internalized. **A.** Distribution of the number of intracellular bacteria detected at 1.5 h pi (average from 3 replicates). The infection efficiencies are clustered in 3 groups: low, medium and high infection, corresponding respectively to 1 to 2; 3 to 8 or more than 9 bacteria per cell. **B.** The vulnerability score is represented as a function of the number of intracellular bacteria resulting from the 1st infection. **C.** Probability of a cell to be highly infected during the 2nd infection (nI_2 ≥9) as a function of the number of intracellular bacteria being internalized during the 1st infection. **B** and **C** represent the data merged from all the experiments (delay of 1, 2 and 3 h before the second infection). Groups of infection efficiency are identical in **A**, **B** and **C**.
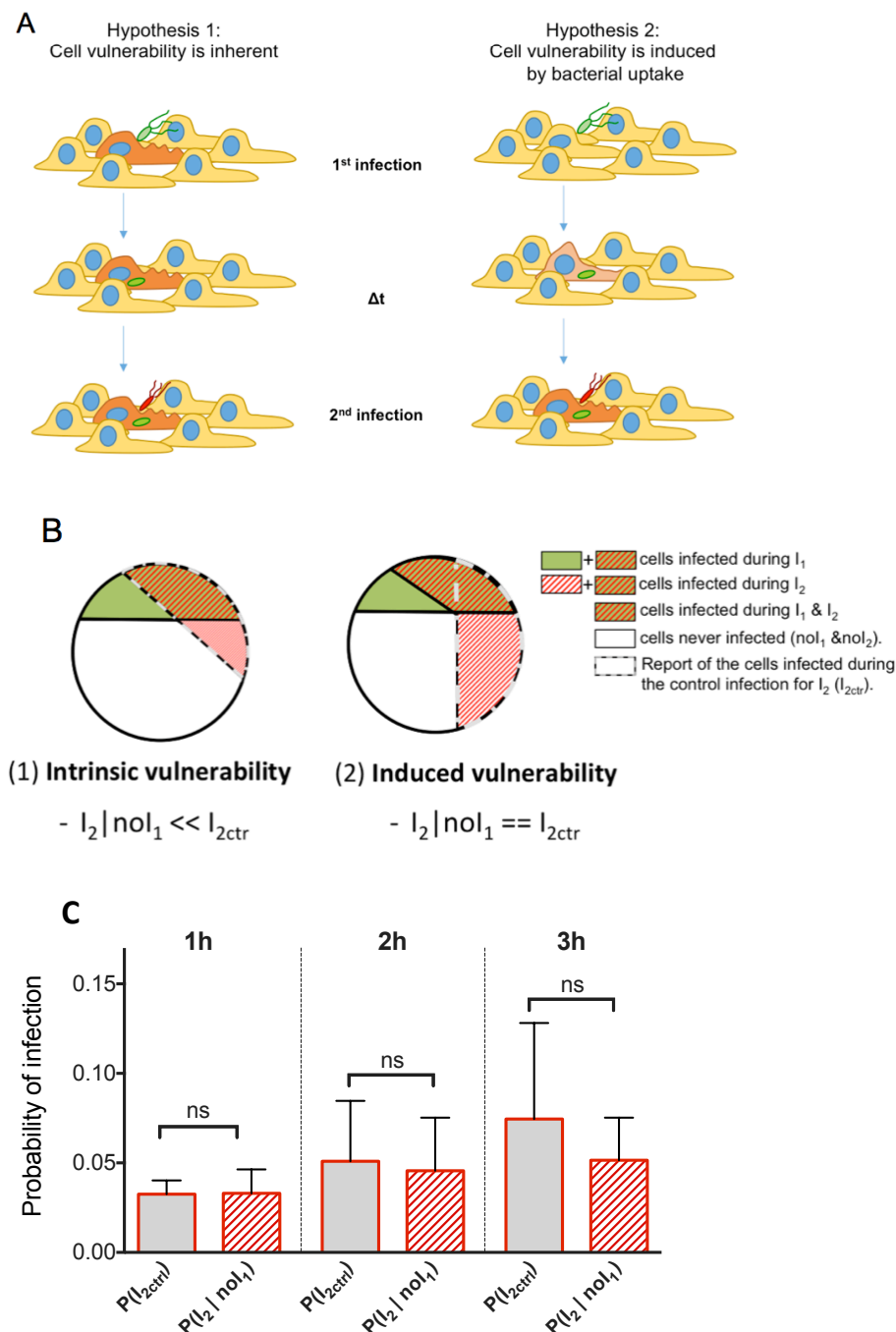
Fig.4. Cell vulnerability examined as an intrinsic or an induced property. A. Schemes of the two hypotheses for the origin of cell vulnerability. In the hypothesis 1, cell vulnerability is inherent: some cells (in orange) are more vulnerable towards infection than other cells (in yellow). In the hypothesis 2, cell vulnerability is induced by bacterial uptake: before infection cells are equal regarding their vulnerability (in yellow), but after infection the infected cells turn progressively more vulnerable (in orange). **B.** Graphic representation of the theoretical distribution of the different populations in the case of hypothesis 1 (left) or hypothesis 2 (right). **C.** Probability of infection during sequential infection with 1, 2 and 3 h delays for control cells (I_{2Ctr}) and cells non infected during the 1st infection (noI_1). P-values were obtained after unpaired t-test (P(I_{2ctr}) vs P($I_2 \mid noI_1$)).

Fig.5. Single cell vulnerability to *Salmonella* infection is a combination of intrinsic and induced vulnerability. A. The depicted cellular parameters were determined for HeLa cells (upper panel) and Caco-2 cells (lower panel) as described in detail in *Materials and Methods*. An overlay of the distribution of some of these parameters in infected (red) or non-infected (blue) cells is shown. **B.** Quantification of the improvement of infection prediction by each cell parameters by subtracting the likelihood (in log) of the model including all parameters from a model ignoring one parameter. Results are averaged over 100 training/testing circles for each model. P-values were obtained after paired t-test. **C.** Fold change of the probability of infection as a function of the intrinsic vulnerability and of a previous infection. **D.** Increasing or decreasing of the probability of infection when the listed cell parameters increase their values.
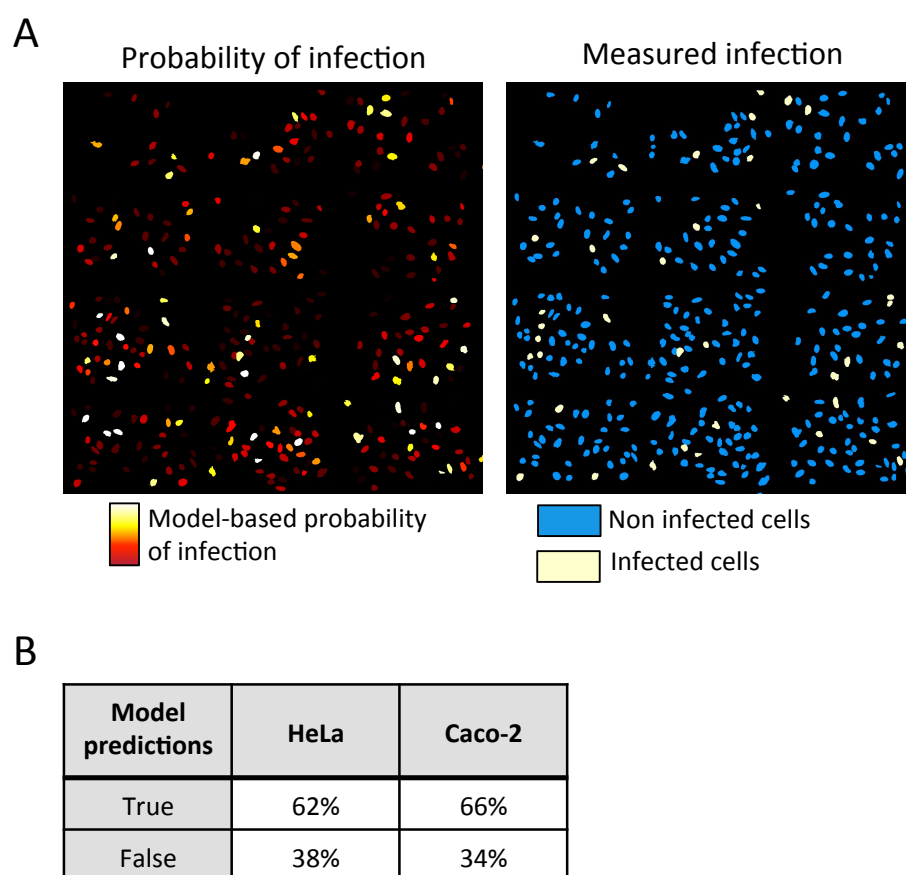
Model predictions	HeLa	Caco-2
True	62%	66%
False	38%	34%

Fig.6. Comparison of model-predicted vulnerability of single-cell with measured-infection. A. Model-predicted probability of infection displayed on reproduced original images of HeLa cells (left panel). Colors are adapted for maximum contrast between lowest (deep red) and highest (white) probability of infection. Measured infections from experiments are shown (top-right panel). **B.** Estimation of the reliability of the two (HeLa and Caco-2) models developed when tested on a total of 100 infected cells and 100 non-infected cells.
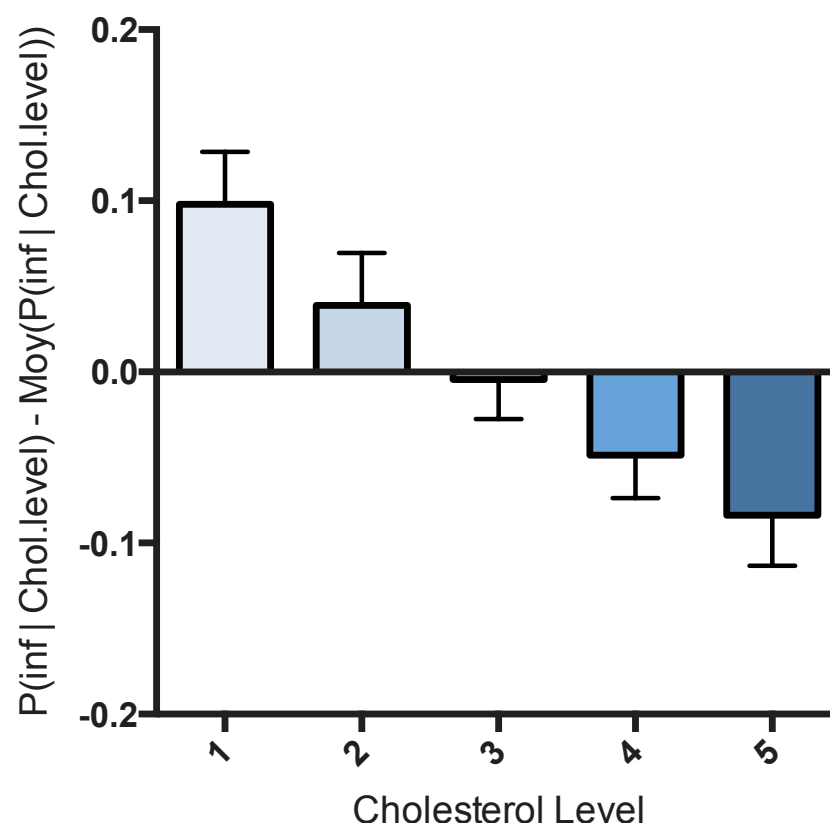
Fig.7. Probability of infection as a function of single cell cholesterol level. Variation of the probability of *Salmonella* infection at different levels of host cholesterol measured by FACS as described in detail in the text. Cholesterol levels were binned in five categories at 20% steps from lowest to highest levels over the total cell population, each category contains 20% of the whole cells. Results are averaged over 3 independent experiments.
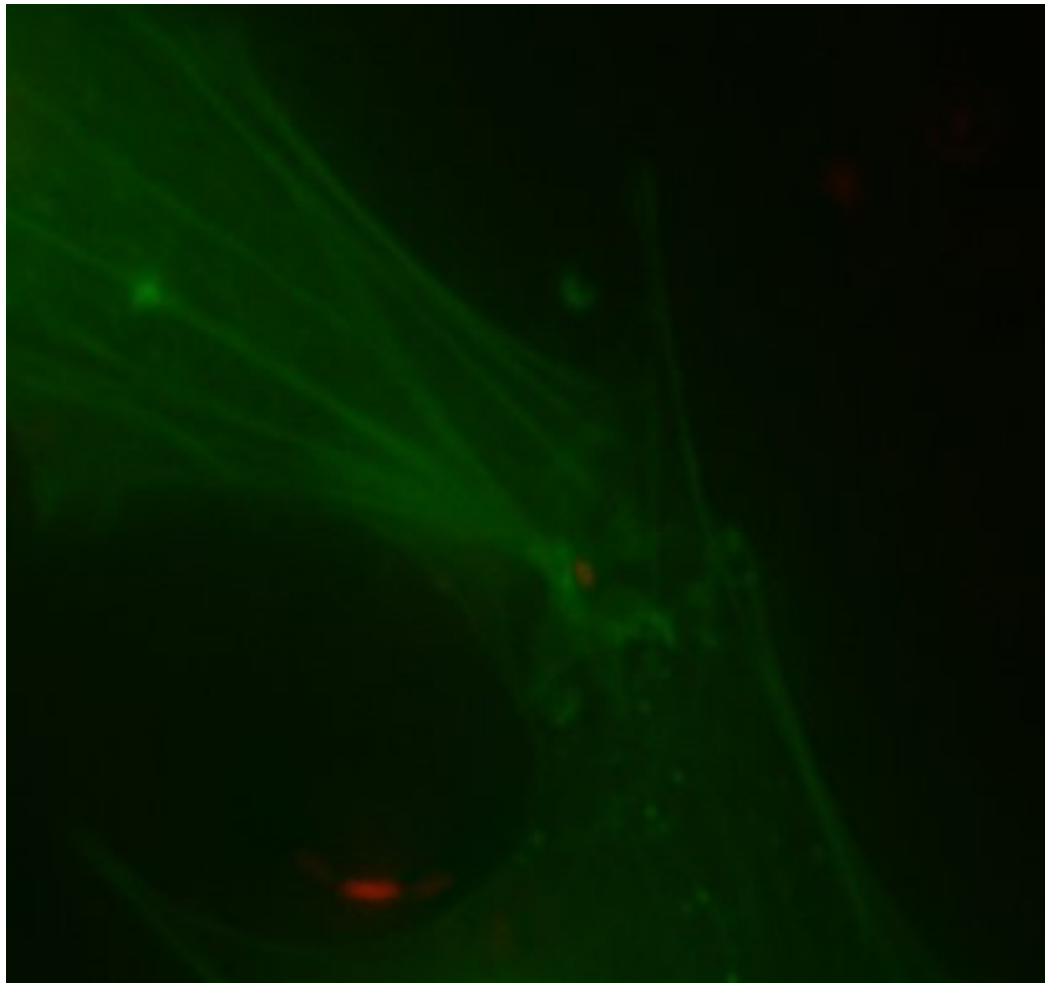
Video.S1. Ruffle appearance and disappearance after entry of single salmonellae in a host cell. Time intervals between the frames are 3 min. The green channel corresponds to actin-GFP transfected cells and shows the membrane ruffles. The red channel shows salmonellae SL$_{dsRed}$.
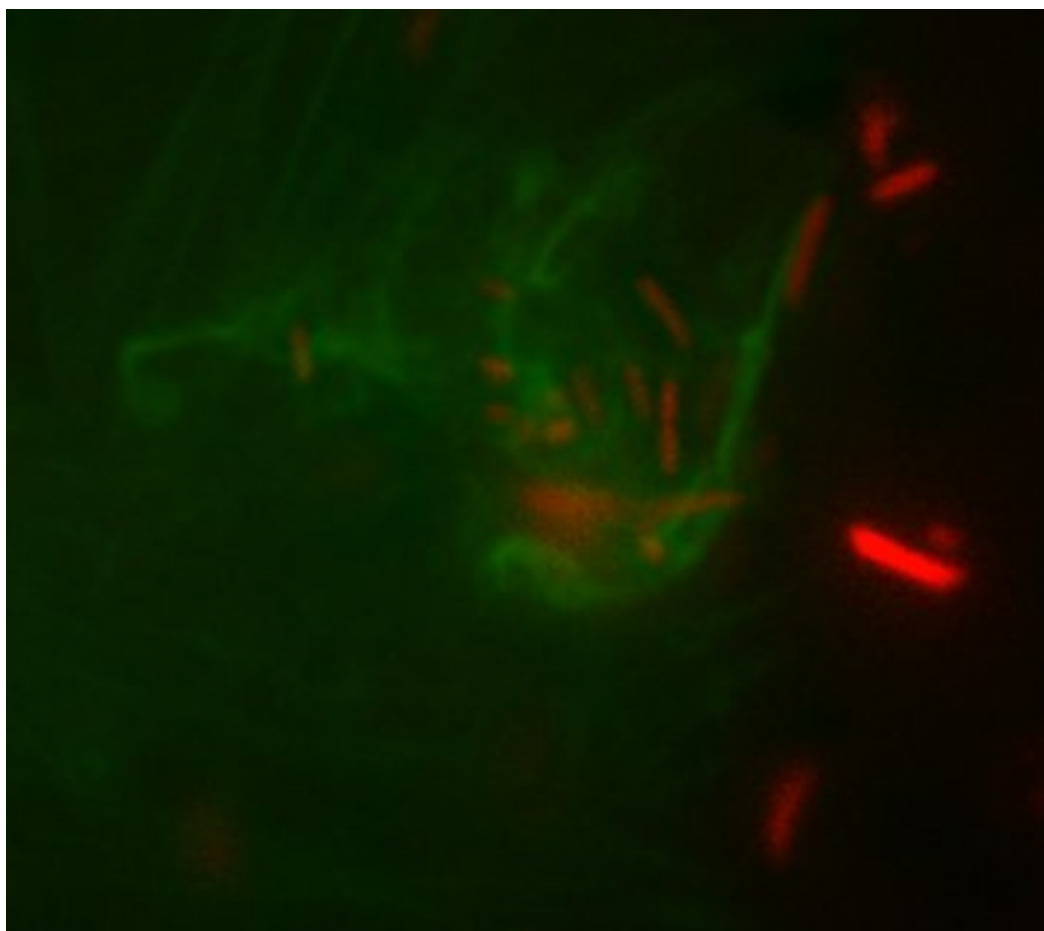
Video.S2. Ruffle appearance and disappearance after entry of multiple salmonellae in a host cell. Time intervals between the frames are 3 min. The green channel corresponds to actin-GFP transfected cells and shows the membrane ruffles. The red channel shows salmonellae SL_{dsRed}.

Cell line: HeLa	Coefficient	Standard Error	Difference of LogLikelihood	p-value on difference of LogLikelihood
Intercept	-4.231	0.169		
Low 1st Infection	0.385	0.033		
Medium 1st Infection	0.765	0.031	97.725	5.99E-52
High 1st Infection	1.461	0.0322		
Infected neighbor cells	0.055	0.009	1.941	2.01E-18
Non-infected neighbor cells	-0.053	0.01	1.25	1.10E-09
Local cell density	0.018	0.003	1.752	6.07E-16
Cell perimeter	0.004	0.000	11.409	9.27E-38
Circularity	0.198	0.167	-0.008	8.89E-01
Delay-2h	0.155	0.029	5.332	1.57E-33
Delay-3h	0.304	0.03		

Table.S1. Model coefficient values for HeLa cells with the corresponding standard error for each cell parameter. Difference of log-likelihood obtained after subtraction of the log-likelihood of the model including all parameters from the log-likelihood of a model ignoring one parameter (see graphic representation in **Fig.5B**). The presented values were averaged with the values obtained over 100 training/testing circles for each model. P-values were obtained after paired t-test.

Cell line: Caco-2	Coefficient	Standard Error	Difference of LogLikelihood	p-value on difference of LogLikelihood
Intercept	-0.48	0.14		
Low 1st Infection	0.25	0.02		
Medium 1st Infection	0.23	0.03	7.07	2.29E-30
High 1st Infection	0.20	0.06		
Infected neighbor cells	0.23	0.01	55.69	3.28E-77
Non-infected neighbor cells	0.09	0.00	20.71	7.75E-58
Local cell density	-0.02	0.00	86.25	6.70E-82
Cell perimeter	0.01	0.00	70.56	2.52E-78
Circularity	-2.96	0.14	22.83	1.26E-53

Table.S2. Model coefficient values for Caco-2 cells with the corresponding standard error for each cell parameter. Difference of log-likelihood obtained after subtraction of the log-likelihood of the model including all parameters from the log-likelihood of a model ignoring one parameter (see graphic representation in **Fig.5B**). The presented values were averaged with the values obtained over 100 training/testing circles for each model. P-values were obtained after paired t-test.
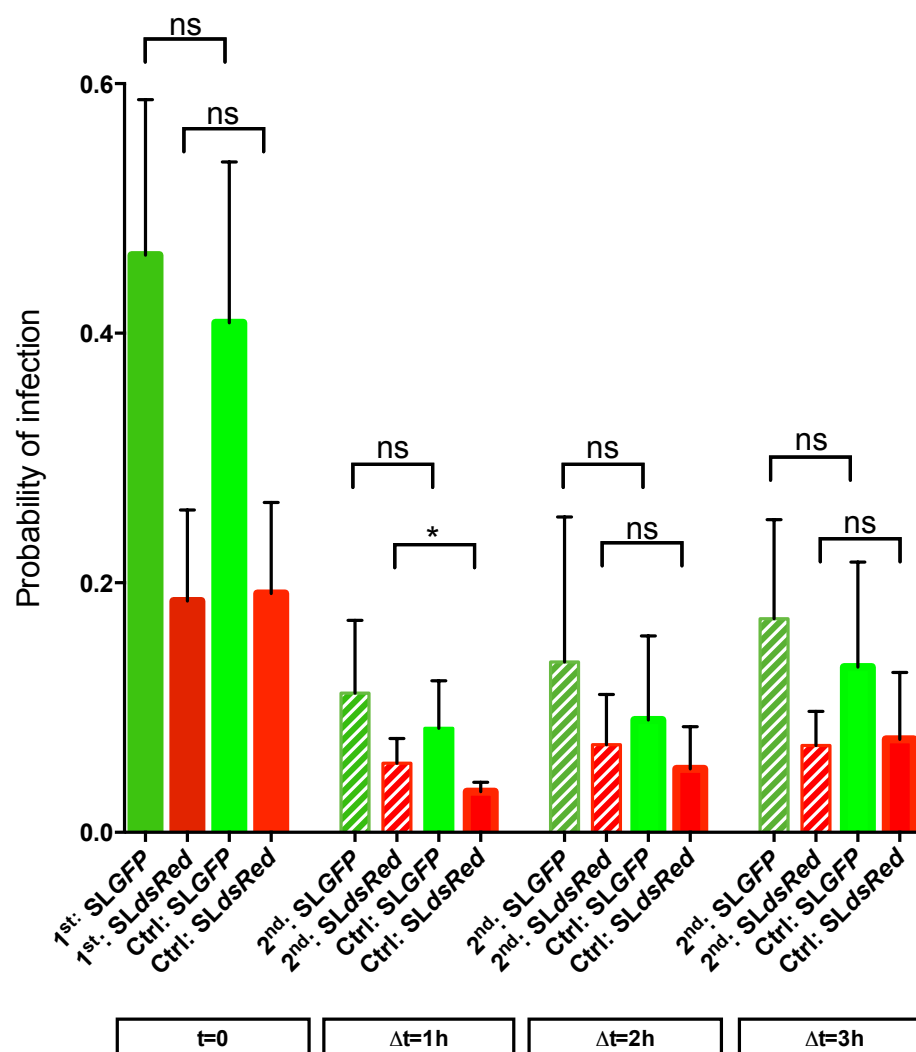
Fig.S1. Probability of SL_{GFP} and SL_{dsRed} infections at different time-points after the beginning of cell challenge (t=0) between single (control) or sequential infections.
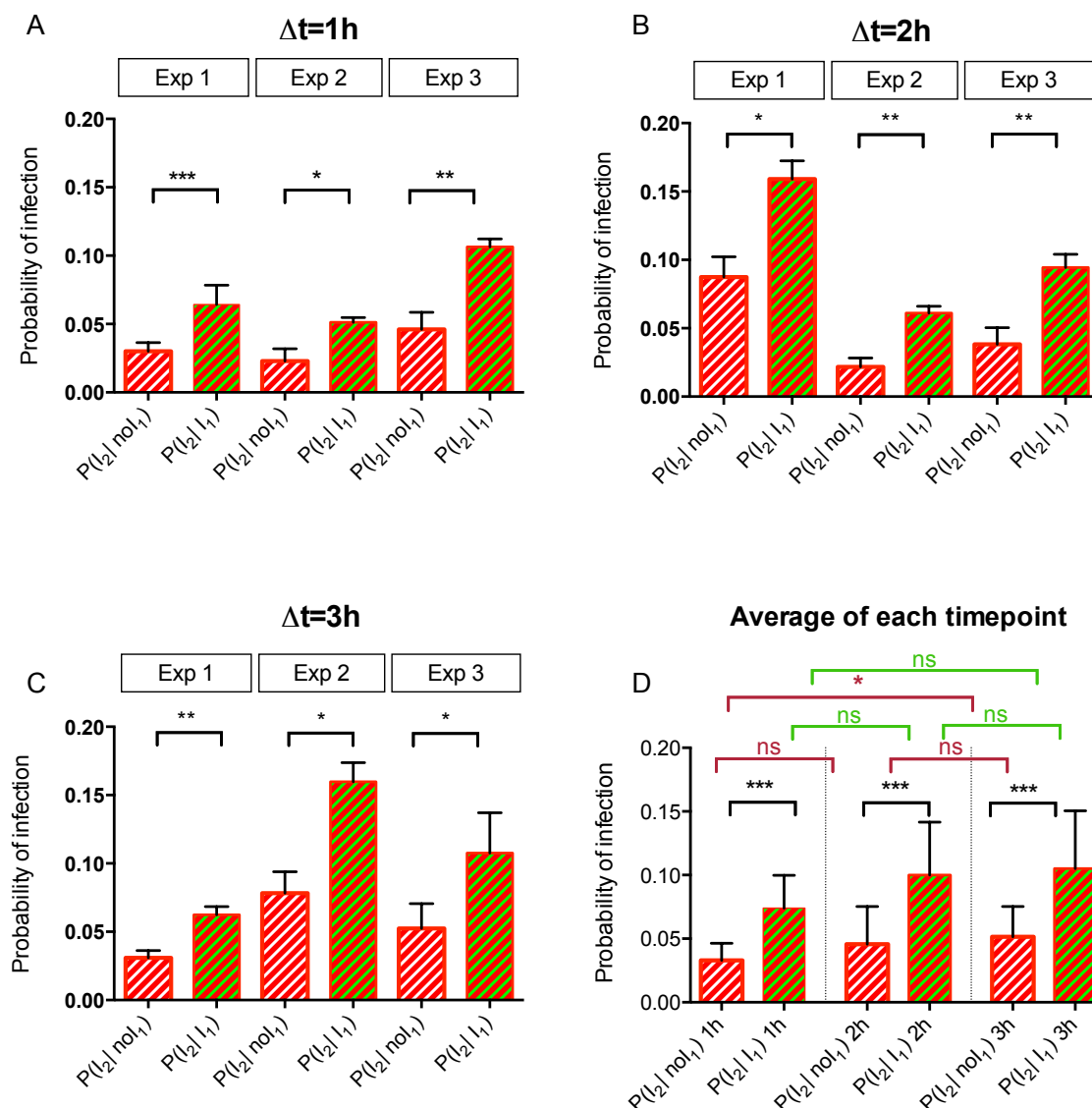
Fig.S2. A.B.C. Detailed depiction of the conditional probability of infection for two different populations during sequential infection with a delay of 1 h (**A**), 2 h (**B**) and 3 h (**C**) for each independent experiment with 3 replicates per experiment. P-values were obtained after paired t-test. **D.** Representation of the results from **A**, **B** and **C** after averaging them for each delay. P-values were obtained after paired t-test. The P-values in black resulted from a t-test comparing $P(I_2 \mid I_1)$ and $P(I_2 \mid noI_1)$. The P-values in red resulted from a t-test comparing $P(I_2 \mid noI_1)$ for 1 h versus 2 h and 2 h versus 3 h. The P-values in green resulted from a t-test comparing $P(I_2 \mid I_1)$ for 1 h versus 2 h and 2 h versus 3 h.
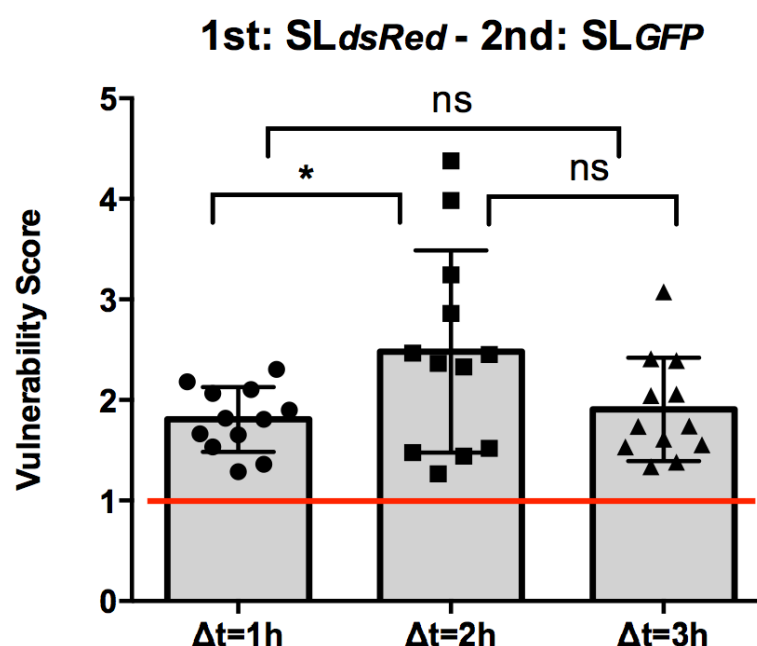
Fig.S3. Vulnerability scores for the inverted infections compared to **Fig.2C** (SL_{dsRed} before SL_{GFP}) with a delay of 1, 2 and 3 h between infections. The red line corresponds to $P(I_2 \mid I_1)=P(I_2 \mid noI_1)=1$ indicating the independence of the infections I_2 and I_1. Values above the red line correspond to $P(I_2 \mid I_1) >P(I_2 \mid noI_1)$ indicating a cooperation between infections. Values below the red line correspond to $P(I_2 \mid I_1) <P(I_2 \mid noI_1)$ indicating a competition between infections. Results were obtained from 3 independent experiments per time-point, and P-values were obtained after unpaired t-test.
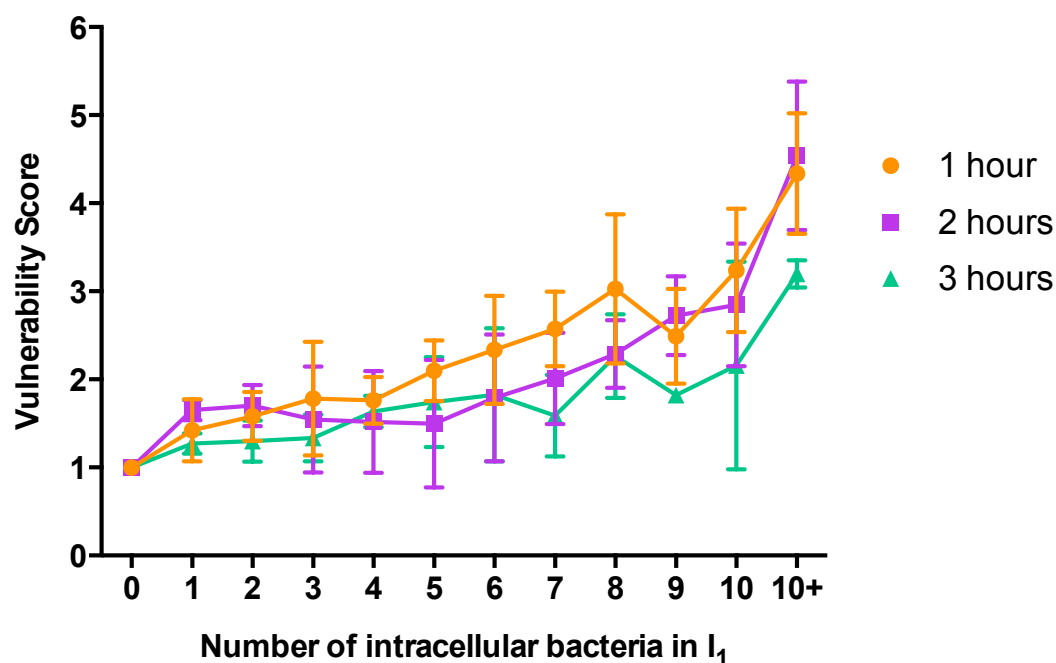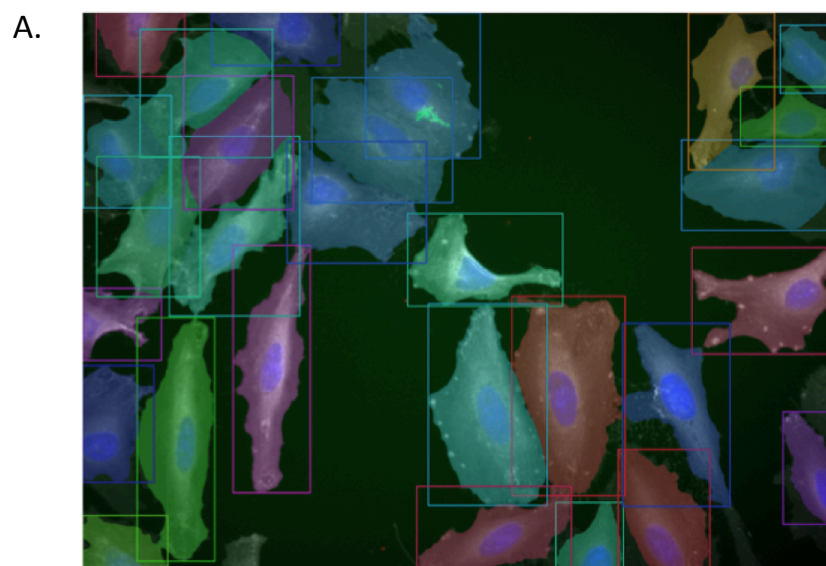
Fig.S4. Vulnerability score as a function of the number of intracellular bacteria resulting from the 1st infection with a delay of 1, 2 and 3 h between the infections. Results were obtained from 3 independent experiments per time-point.

A.

Icy automated cell detection

B.

HeLa

Correlations	2nd Infection	Load of Infection	Infected neighbor cells	Non-infected neighbor cells	Neighbor cells	Local cell density	Cell perimeter	√Circularity	Delay
2nd Infection									
Load of Infection	0.15								
Infected neighbor cells	0.07	0.27							
Non-infected neighbor cells	-0.01	-0.12	-0.07						
Neighbor cells	0.04	0.12	0.72	0.64					
Local cell density	0.02	0.16	0.52	0.05	0.43				
Cell perimeter	0.05	-0.06	0.05	0.33	0.26	-0.36			
√Circularity	-0.03	0.05	-0.01	-0.09	-0.06	0.28	-0.68		
Delay	0.02	-0.09	-0.27	0.07	-0.16	-0.21	0.06	-0.02	

C.

Caco-2

Correlations	2nd Infection	Load of Infection	Infected neighbor cells	Non-infected neighbor cells	Neighbor cells	Local cell density	Cell perimeter	Circularity	Replicate
2nd Infection									
Load of Infection	0.08								
Infected neighbor cells	0.11	0.40							
Non-infected neighbor cells	0.07	-0.10	-0.20						
Neighbor cells	0.12	0.12	0.34	0.85					
Local cell density	-0.13	-0.02	0.00	0.03	0.03				
Cell perimeter	0.21	0.16	0.21	0.42	0.52	-0.37			
Circularity	-0.15	-0.13	-0.22	-0.36	-0.46	0.13	-0.57		
Replicate	0.00	0.04	0.13	-0.12	-0.05	0.36	-0.27	0.12	

-1	-0.75	-0.5	-0.25	0	0.25	0.5	0.75	1

Fig.S5. Cell parameter correlations. **A.** Illustration of Icy cell segmentation using *Active Contours* (*see Materials and Methods for plugins detail*). **B-C.** Table of the correlations between the different cell parameters for HeLa (**B**) and Caco-2 (**C**) cells.
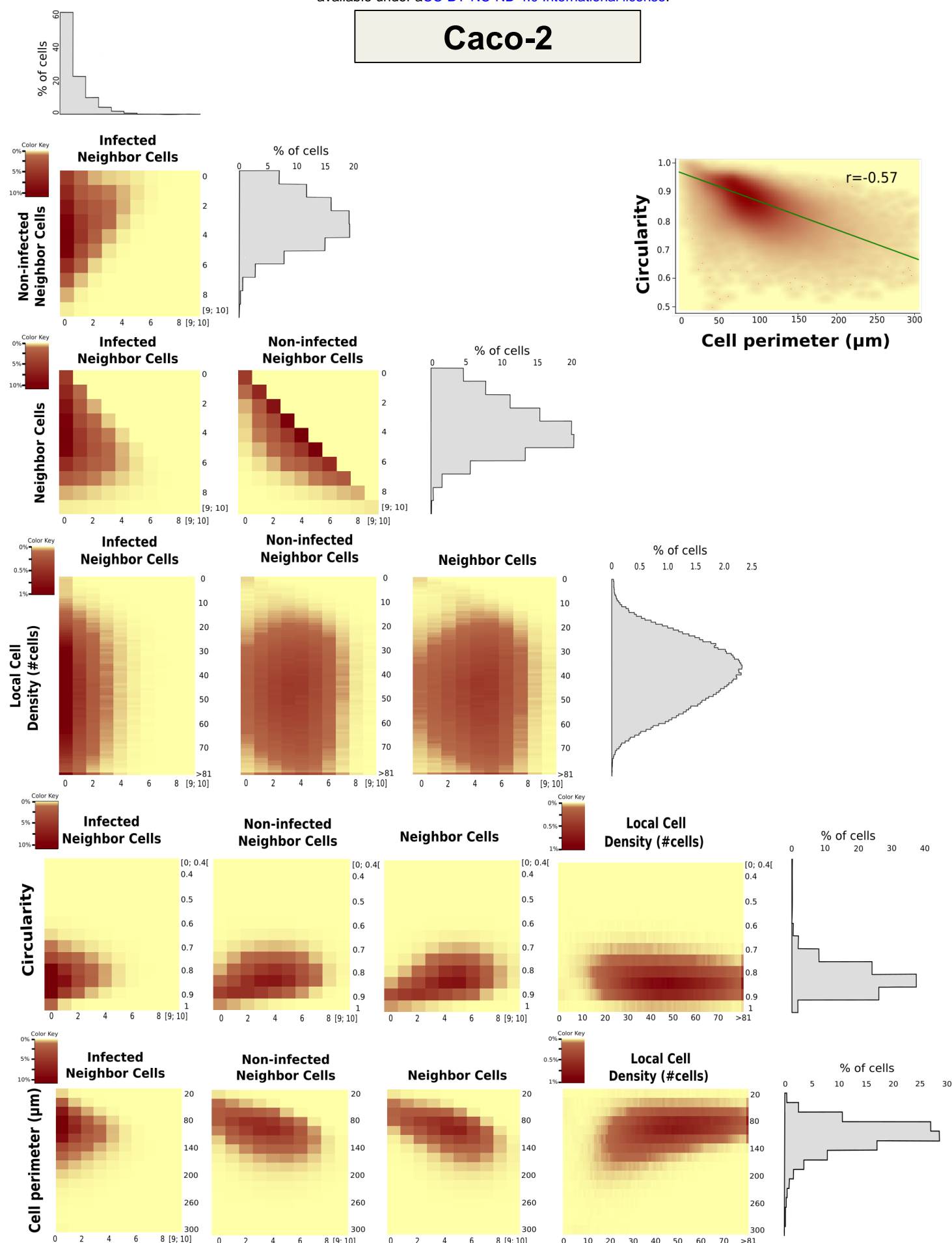
Fig.S6. Scatter plot and heat-map of the different cell parameters studied in HeLa cells model allowing to evaluate the relation between these parameters. Grey histograms represent the distribution of the vertical axe parameter in the entire cell population.

Fig.S7. Scatter plot and heat-map of the different cell parameters studied in Caco-2 cells model allowing to evaluate the relation between these parameters. Grey histograms represent the distribution of the parameter in the entire cell population.
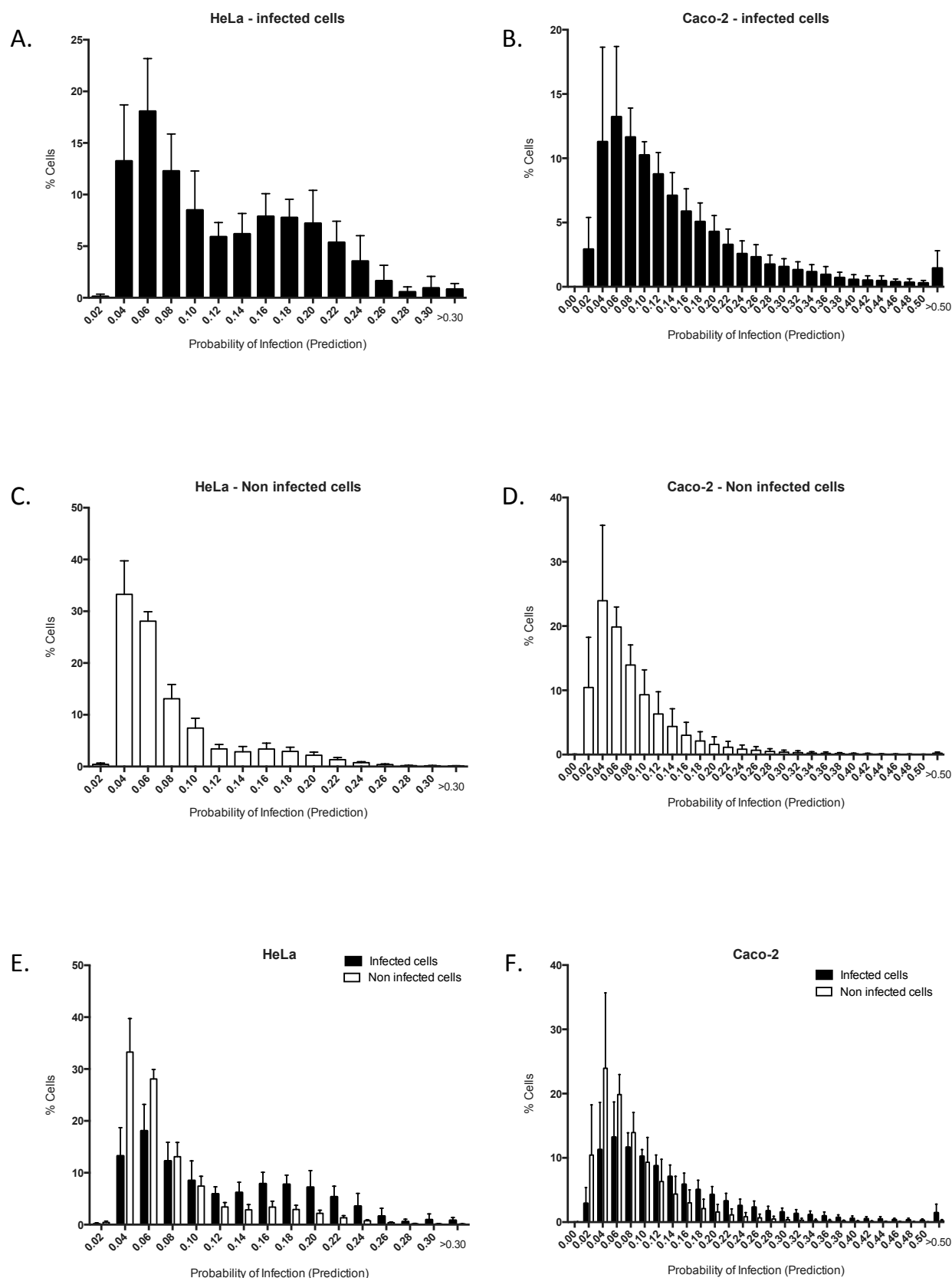
Fig.S8. Distribution of the predicted probability of infection at single-cell level for infected and non infected cells. **A-C-E.** HeLa cells. **B-D-F.** Caco-2 cells.
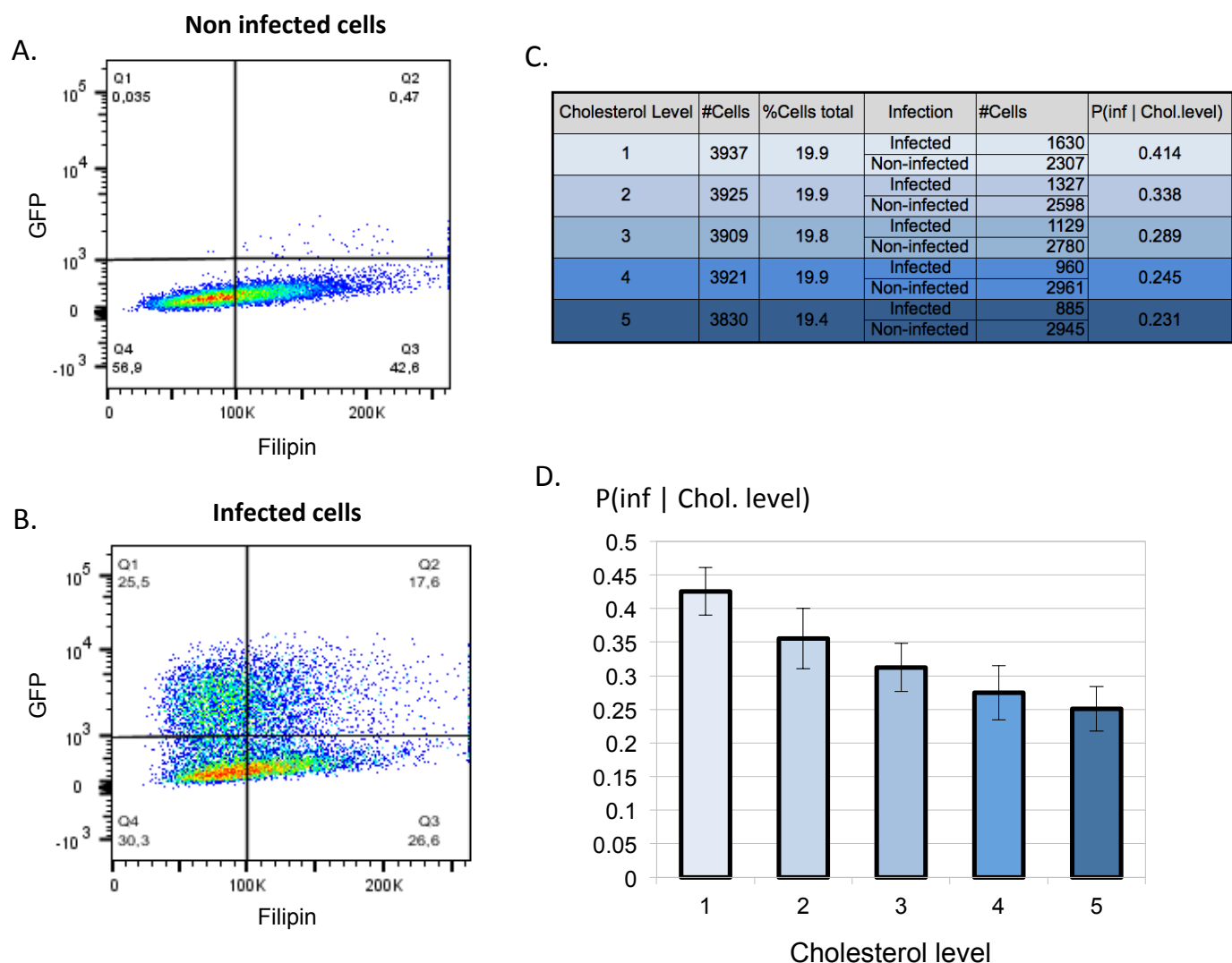
Fig.S9. Illustration of FACS acquisition and data processing. **A-B.** Heat map of Filipin fluorescence as measure of the host cholesterol level (horizontal axe) and GFP fluorescence representing Salmonella-GFP infection (vertical axe). The dials [Q1,Q2] and [Q3,Q4] correspond to the non-infected (**A**) and the infected cells respectively (**B**). **C.** Raw data obtained after binning of the total cell population in 5 categories of cholesterol level (from the lowest to the highest) containing approximately the same number of cell (column 1 to 4). Conditional probability of infection for each category of cholesterol level based on the raw data (5th column). **D.** Representation of the conditional probability of infection for each category of cholesterol level.