Supplemental Data.

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

Running title:
Cell vulnerability during *Salmonella* infection

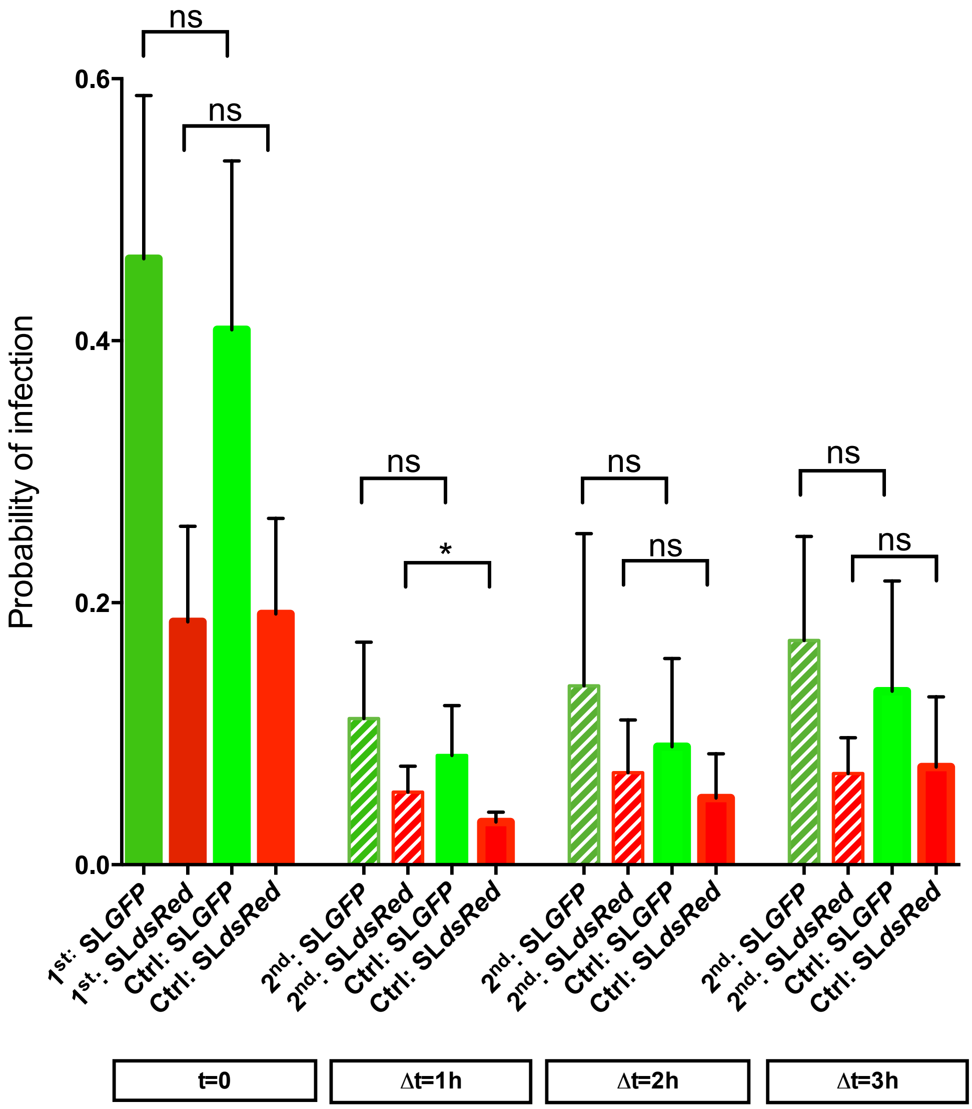


Fig.S1. Probability of SL$_{GFP}$ and SL$_{dsRed}$ infections at different time-points after the beginning of cell challenge (t=0) between single (control) or sequential infections.

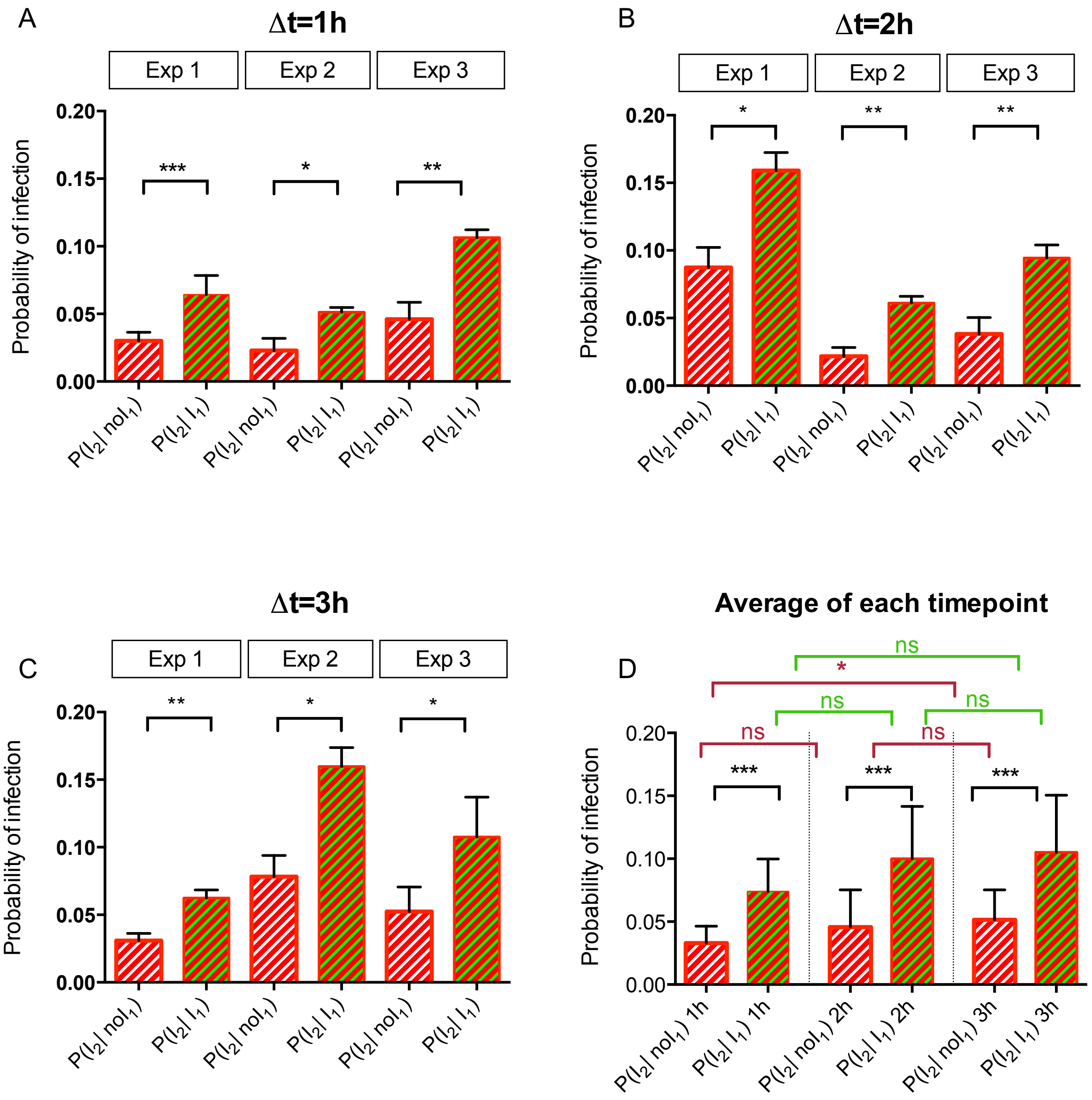


Fig.S2. A.B.C. Detailed depiction of the conditional probability of infection for two different populations during sequential infection with a delay of 1 h (**A**), 2 h (**B**) and 3 h (**C**) for each independent experiment with 3 replicates per experiment. P-values were obtained after paired t-test. **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.

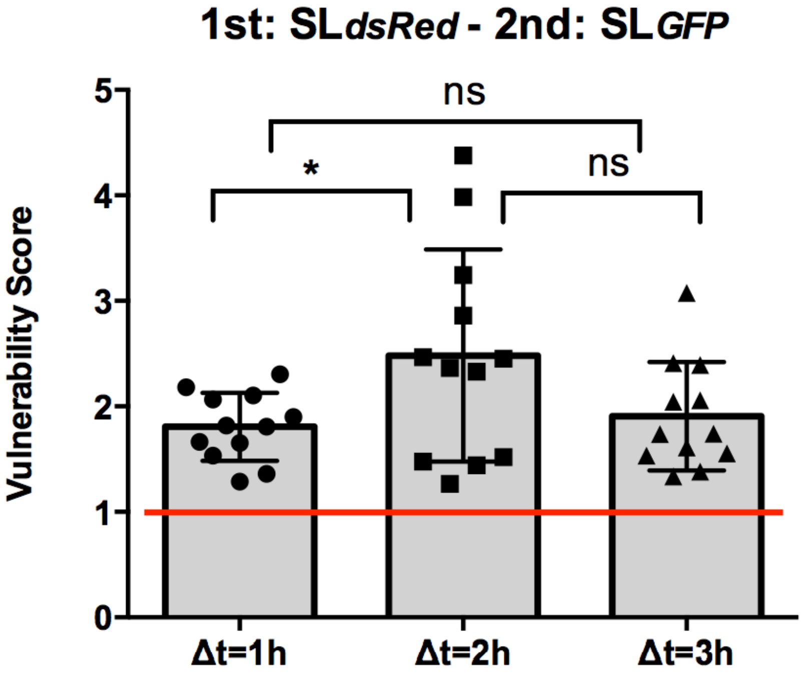


Fig.S3. Vulnerability scores for the inverted infections compared to **Fig.2C** (SL_{dsRed} before SL_{GFP}) with a delay of 1, 2 and 3 h between infections. The red line corresponds to $P(I_2 \mid I_1)=P(I_2 \mid noI_1)=1$ indicating the independence of the infections I_2 and I_1. Values above the red 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.

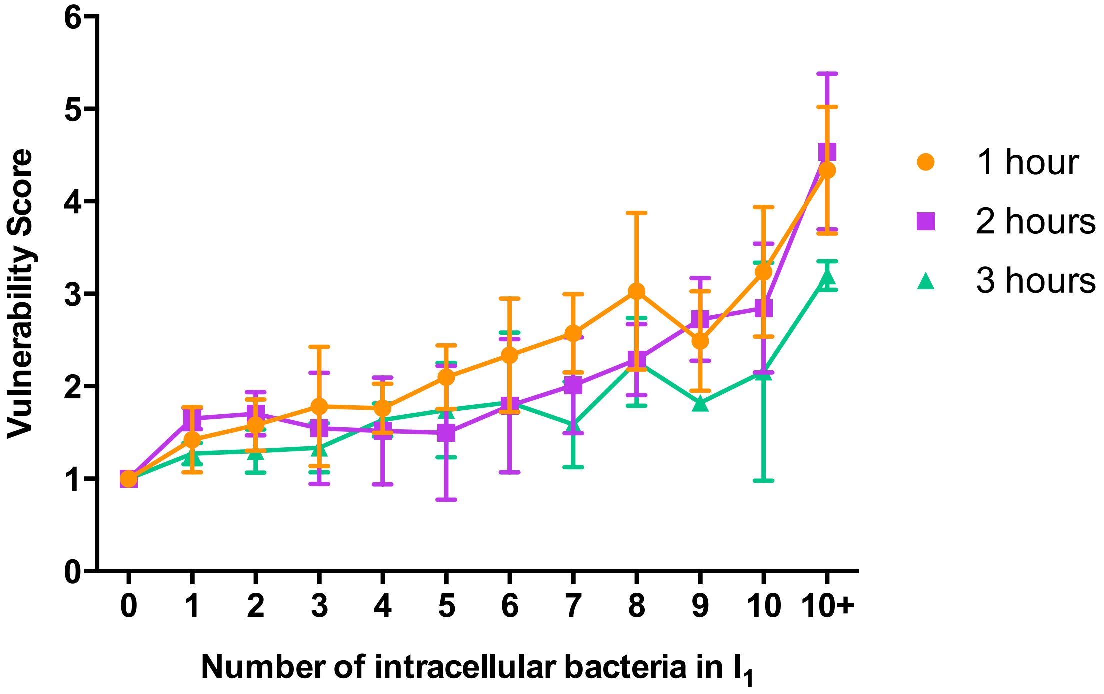
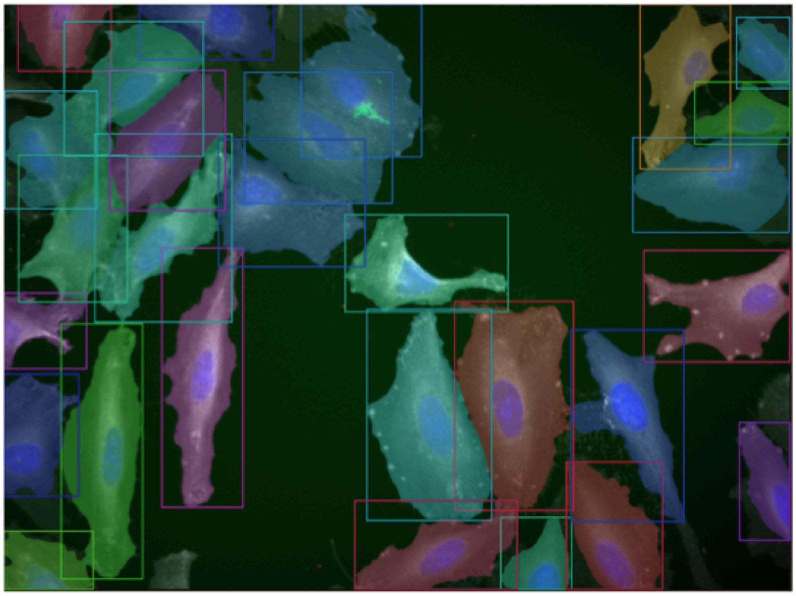


Fig.S4. Vulnerability score as a function of the number of intracellular bacteria resulting from the 1st infection with a delay of 1, 2 and 3 h between the infections. Results were obtained from 3 independent experiments per time-point.

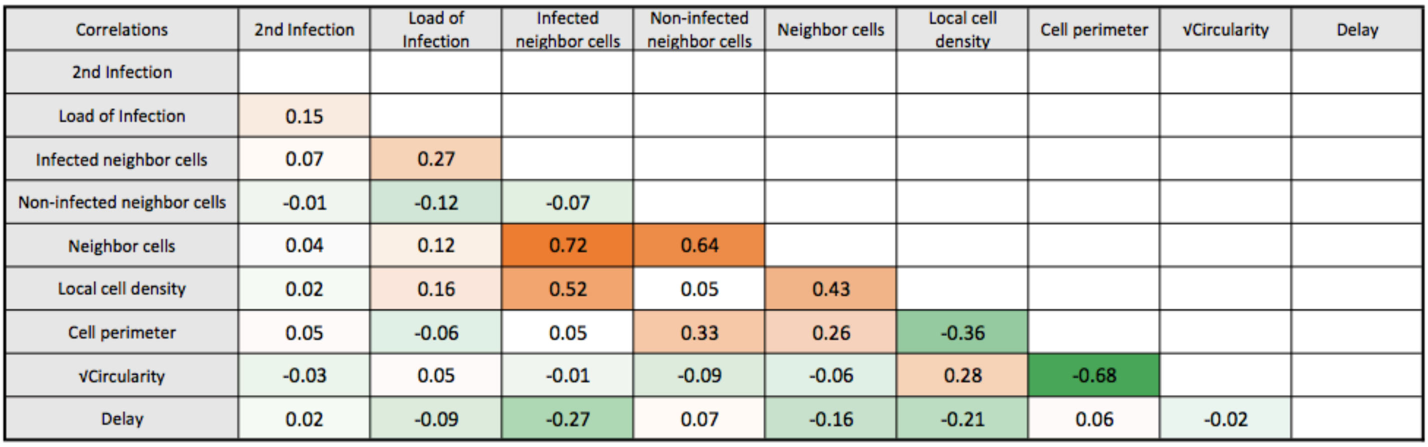
A.



Icy automated cell detection

B.

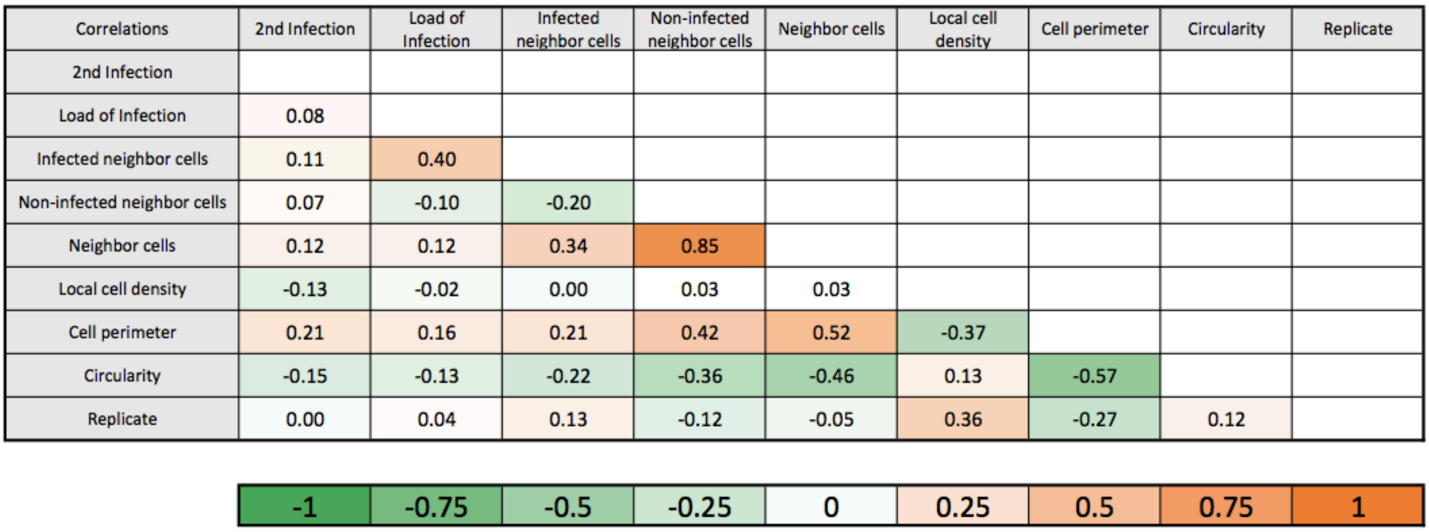
HeLa



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.

HeLa

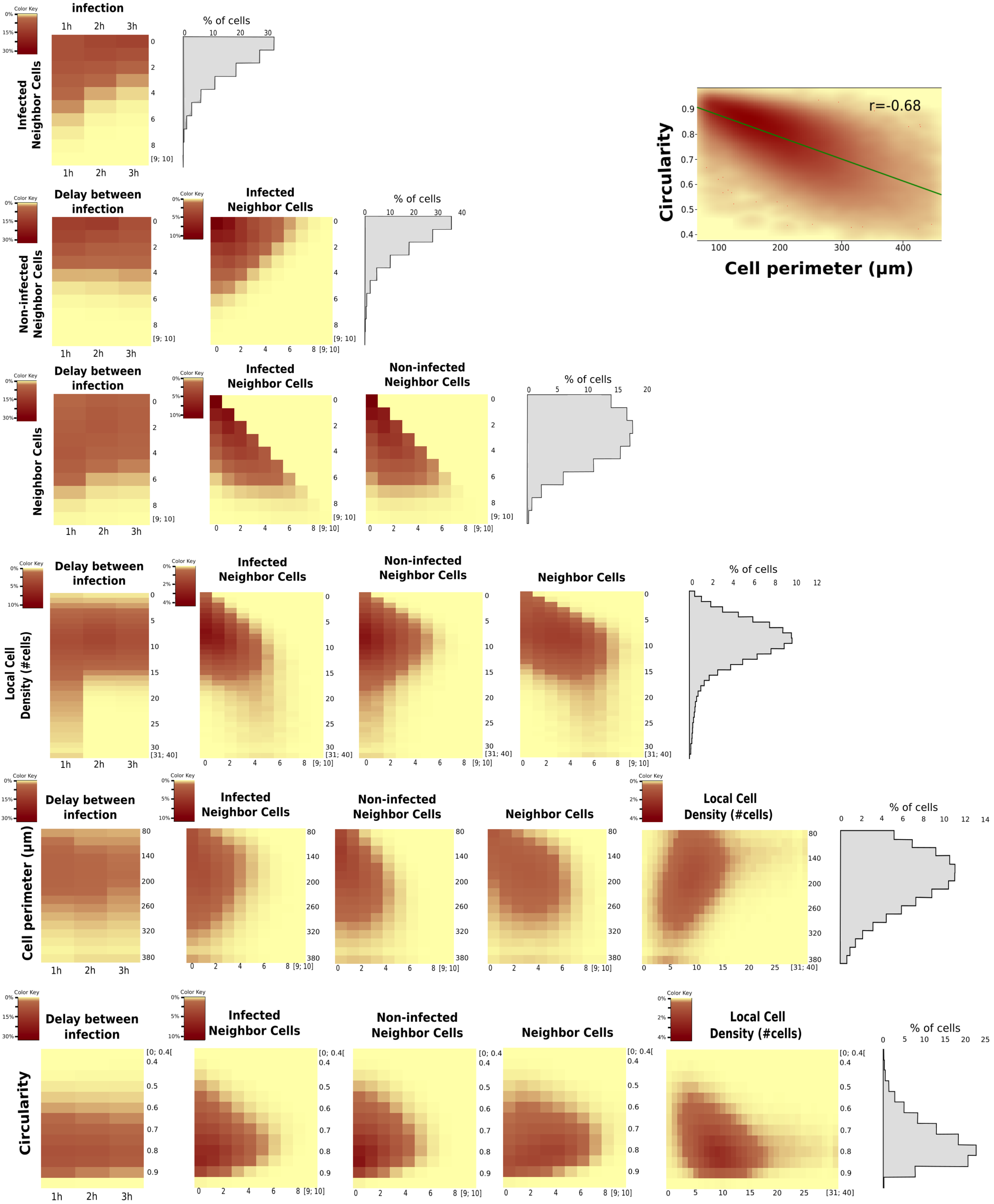


Fig.S6. Scatter plot and heat-map of the different cell parameters studied in HeLa cells model allowing to evaluate the relation between these parameters. Grey histograms represent the distribution of the vertical axe parameter in the entire cell population.

Caco-2

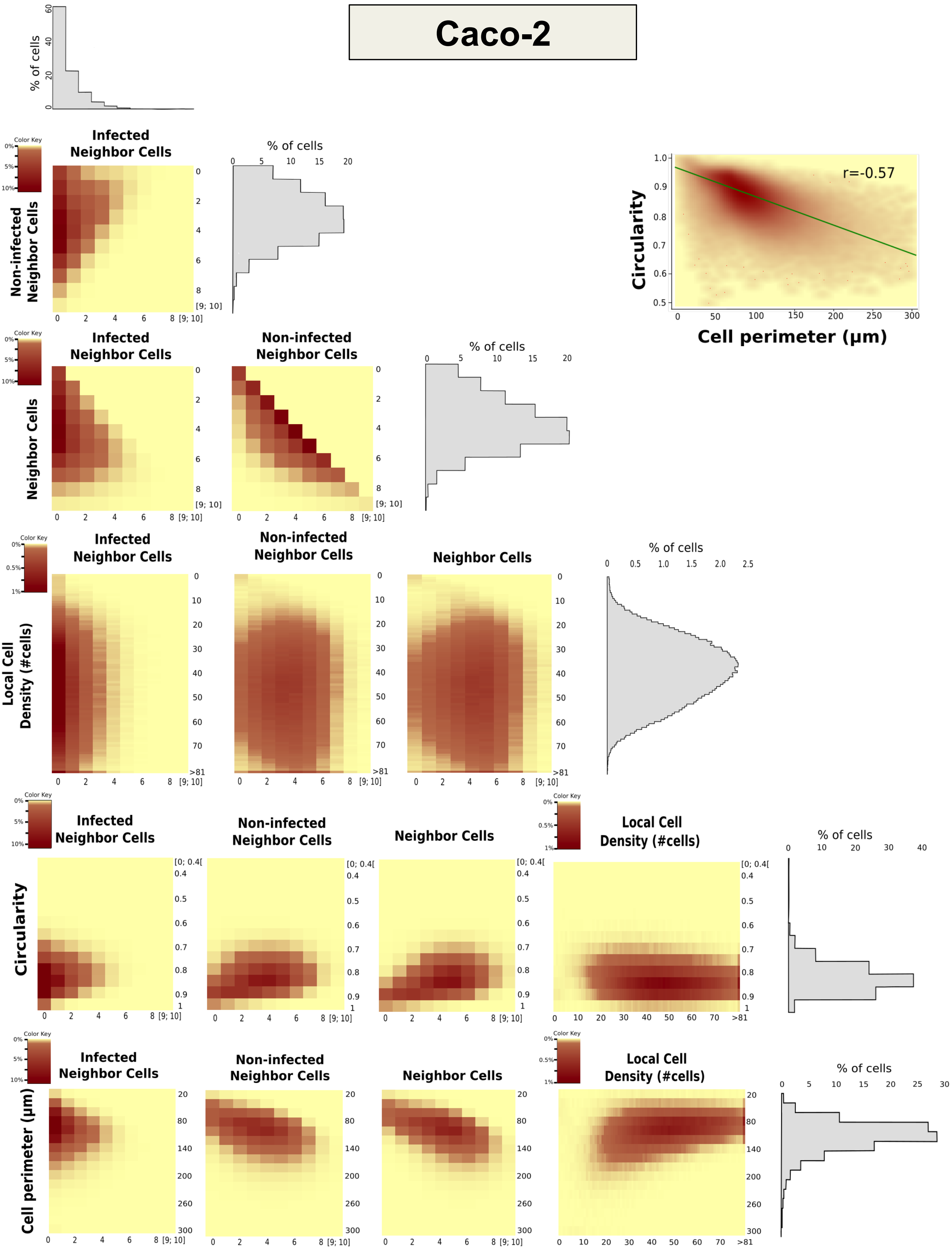


Fig.S7. Scatter plot and heat-map of the different cell parameters studied in Caco-2 cells model allowing to evaluate the relation between these parameters. Grey histograms represent the distribution of the parameter in the entire cell population.

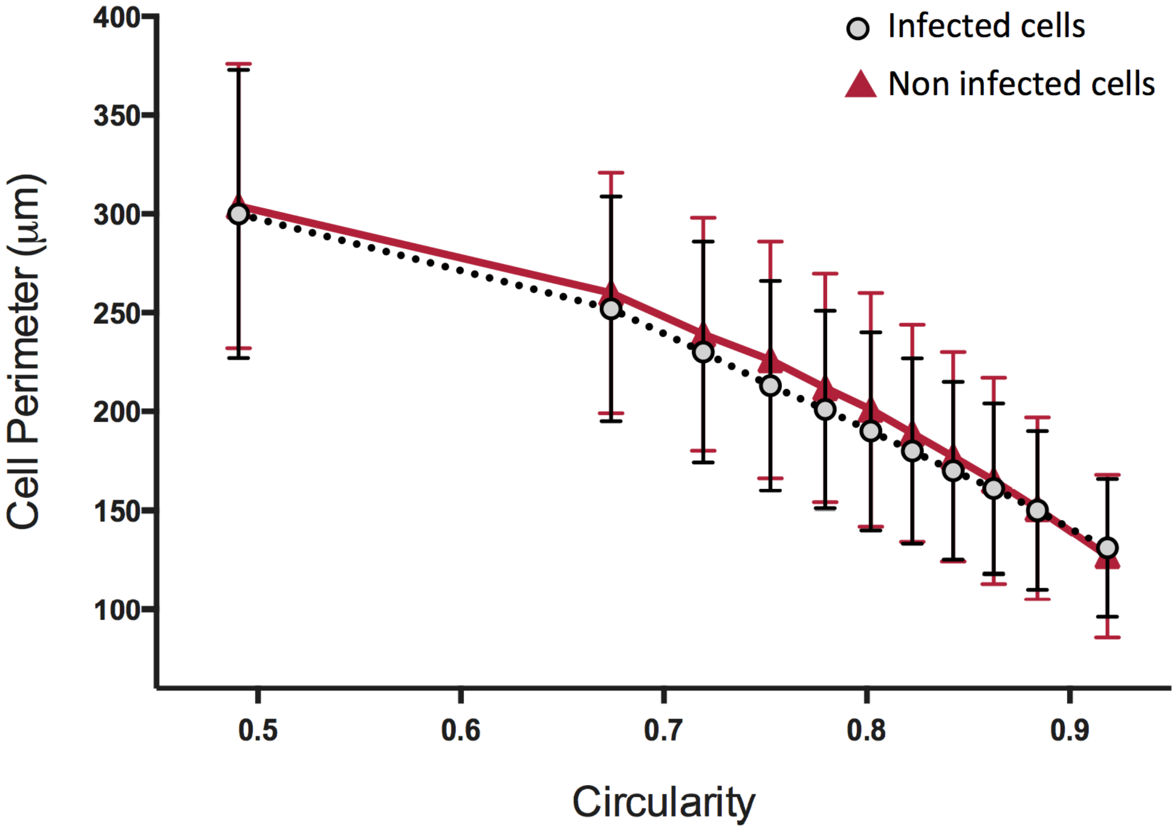


Fig.S8. Investigation of the impact of the first infection on the morphological and local cellular parameters. We selected the cell perimeter as a parameter associated with HeLa cell vulnerability, and the cell circularity as another that was not associated such that both parameters correlate together (correlation coefficient: -0.68). We show the distribution of perimeter values for given values of circularity in infected or non-infected cells. The values for circularity are divided in 11 bins containing the same number of cells, and the values for the perimeter is average for each circularity bin. This analysis was performed on the full dataset (more than 115 000 cells). The comparison of the perimeter of infected versus non-infected cells for groups of different circularity did not show any significant difference. The cell parameters associated with cell vulnerability are not induced by infection.

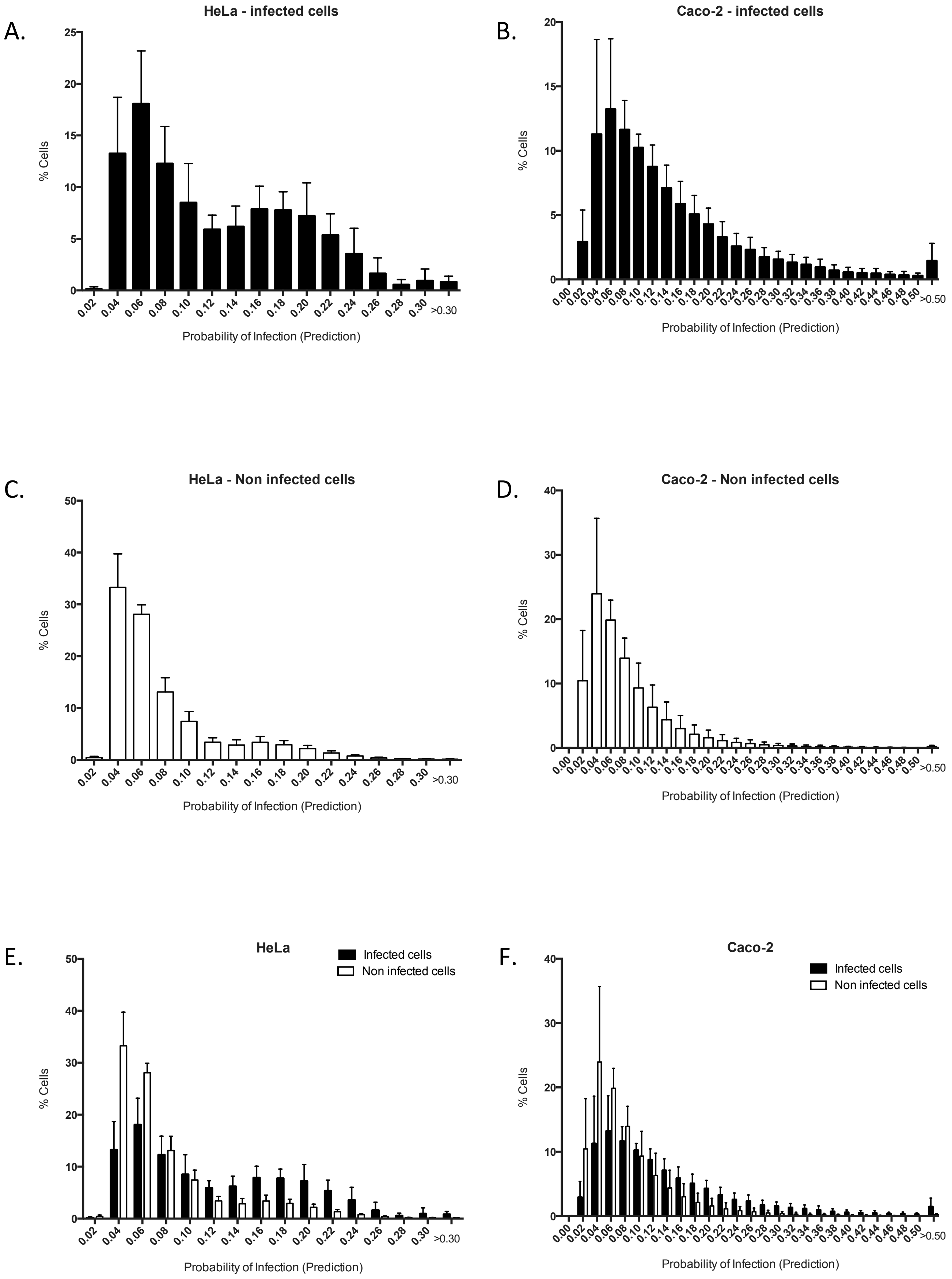
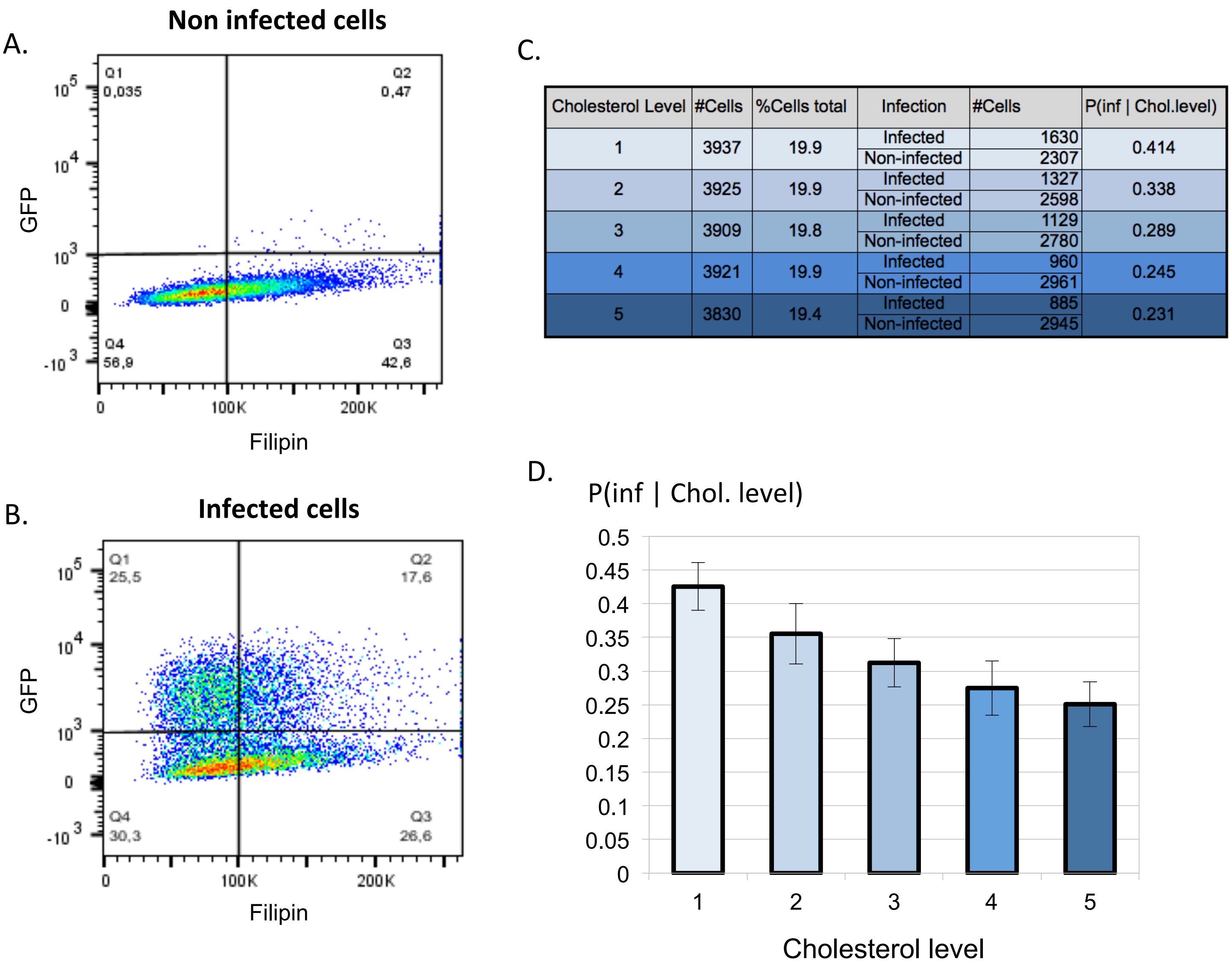


Fig.S9. Distribution of the predicted probability of infection at single-cell level for infected and non infected cells. **A-C-E.** HeLa cells. **B-D-F.** Caco-2 cells.

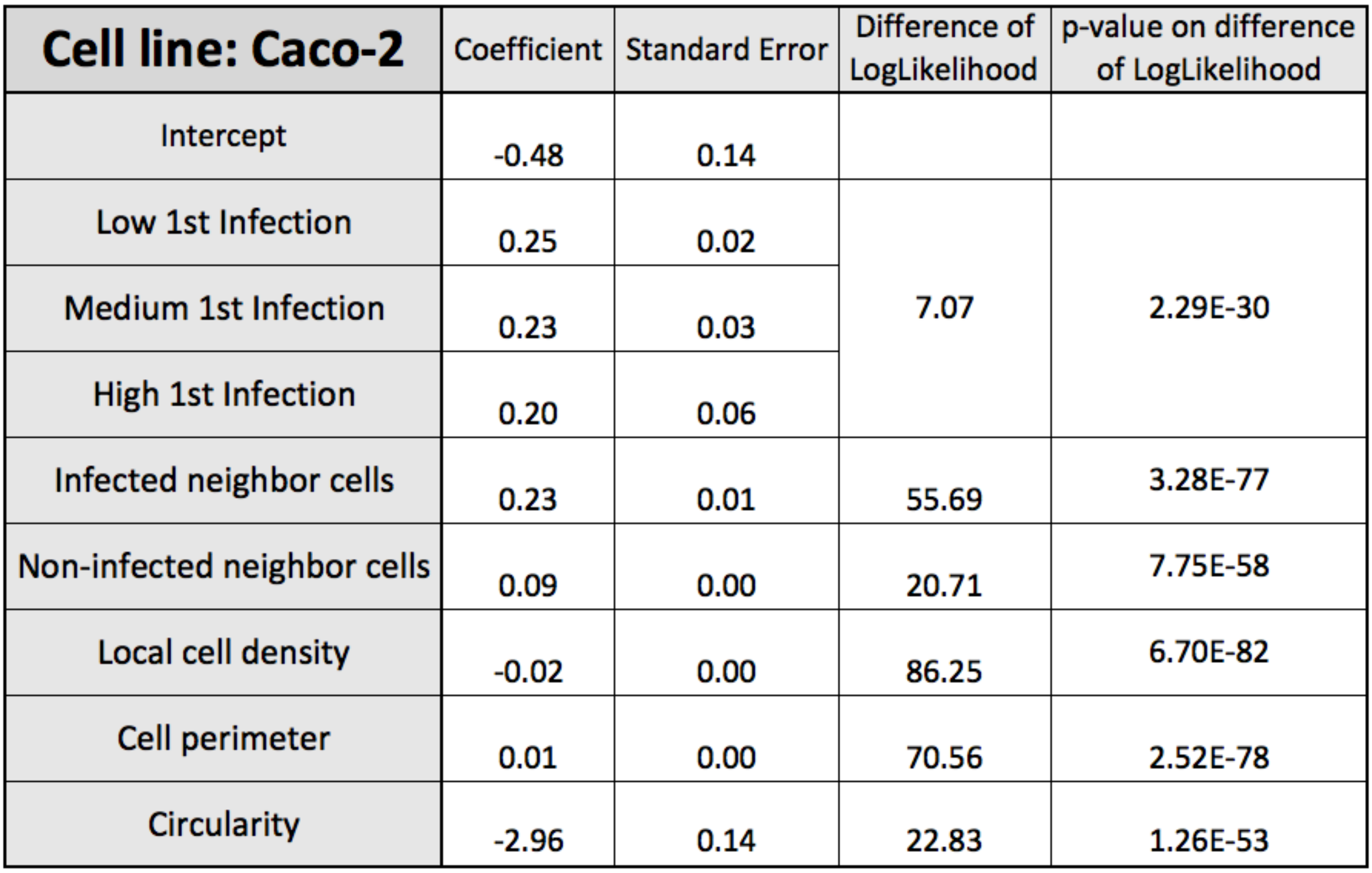


Cholesterol Level	#Cells	%Cells total	Infection	#Cells	P(inf | Chol.level)
1	3937	19.9	Infected	1630	0.414
			Non-infected	2307	
2	3925	19.9	Infected	1327	0.338
			Non-infected	2598	
3	3909	19.8	Infected	1129	0.289
			Non-infected	2780	
4	3921	19.9	Infected	960	0.245
			Non-infected	2961	
5	3830	19.4	Infected	885	0.231
			Non-infected	2945	

Fig.S10. 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.

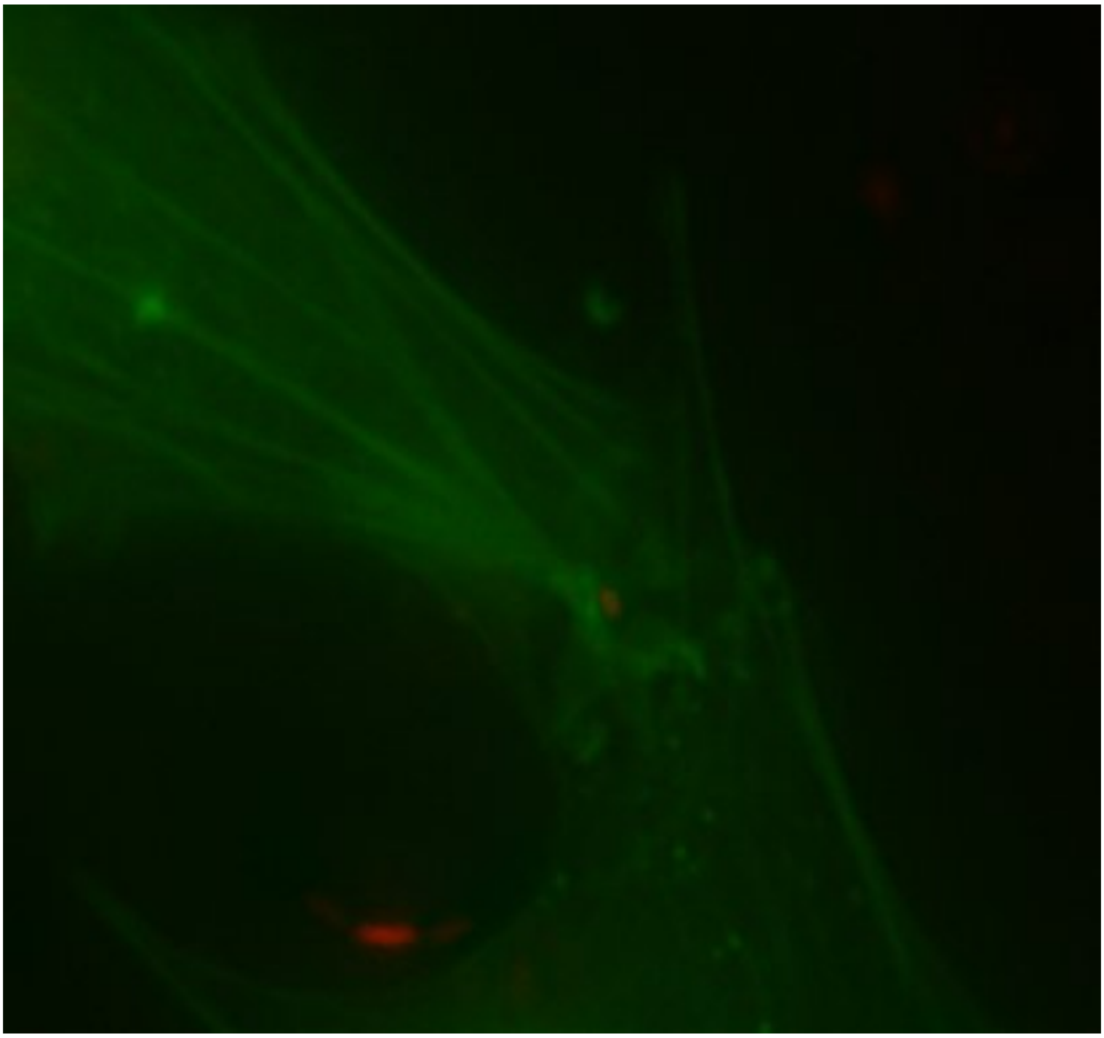
Cell line: HeLa	Coefficient	Standard Error	Difference of LogLikelihood	p-value on difference of LogLikelihood
Intercept	-4.231	0.169		
Low 1st Infection	0.385	0.033	97.725	5.99E-52
Medium 1st Infection	0.765	0.031		
High 1st Infection	1.461	0.0322		
Infected neighbor cells	0.055	0.009	1.941	2.01E-18
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.

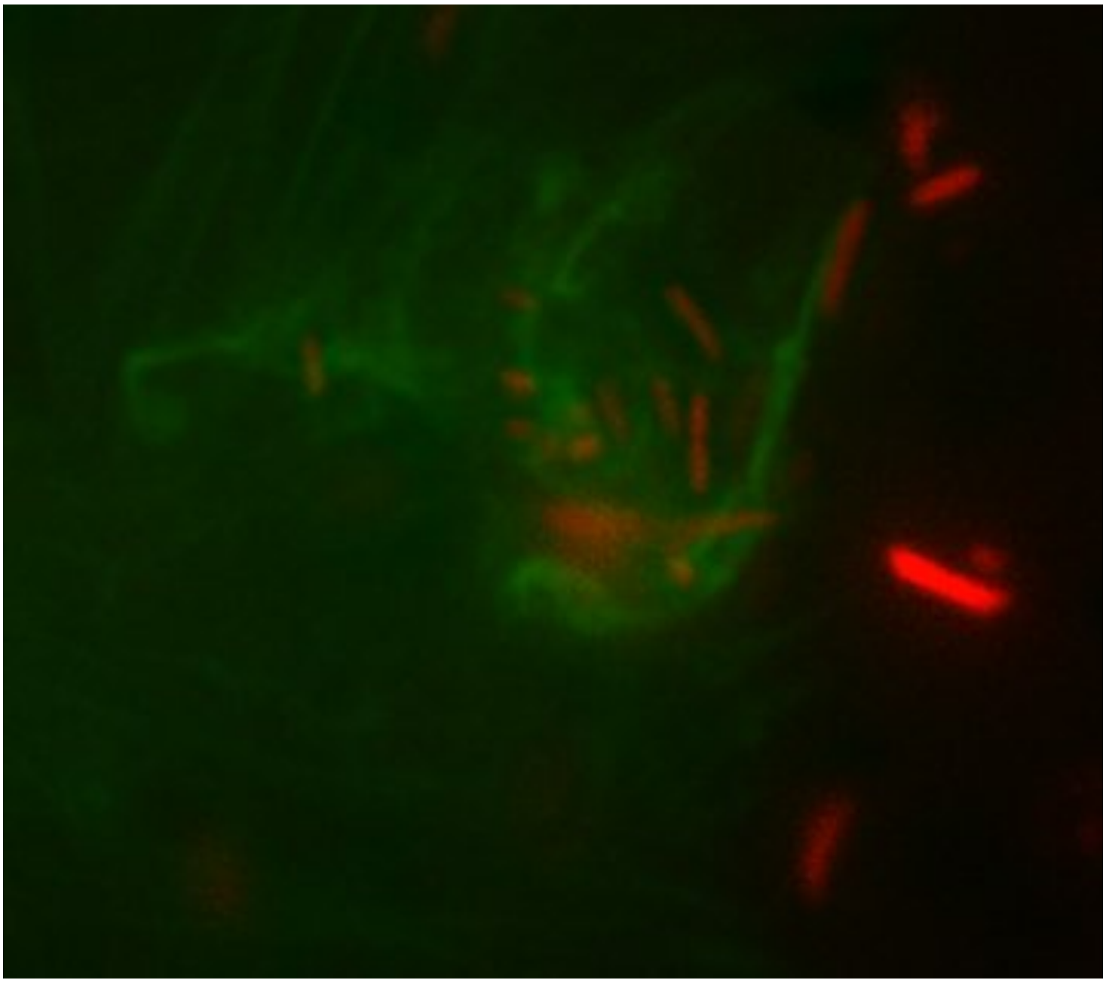


Cell line: Caco-2	Coefficient	Standard Error	Difference of LogLikelihood	p-value on difference of LogLikelihood
Intercept	-0.48	0.14		
Low 1st Infection	0.25	0.02	7.07	2.29E-30
Medium 1st Infection	0.23	0.03		
High 1st Infection	0.20	0.06		
Infected neighbor cells	0.23	0.01	55.69	3.28E-77
Non-infected 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.



Still image - Movie.S1. *(see MovieS1.avi file)* 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}.



Still image - Movie.S2. *(see MovieS2.avi file)* 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}.