Supplementary Information—Spatially correlated gene expression in bacterial groups: the role of lineage history, spatial gradients, and cell-cell interactions

Simon van Vliet, Alma Dal Co, Annina R. Winkler, Stefanie Spriewald, Bärbel Stecher, & Martin Ackermann

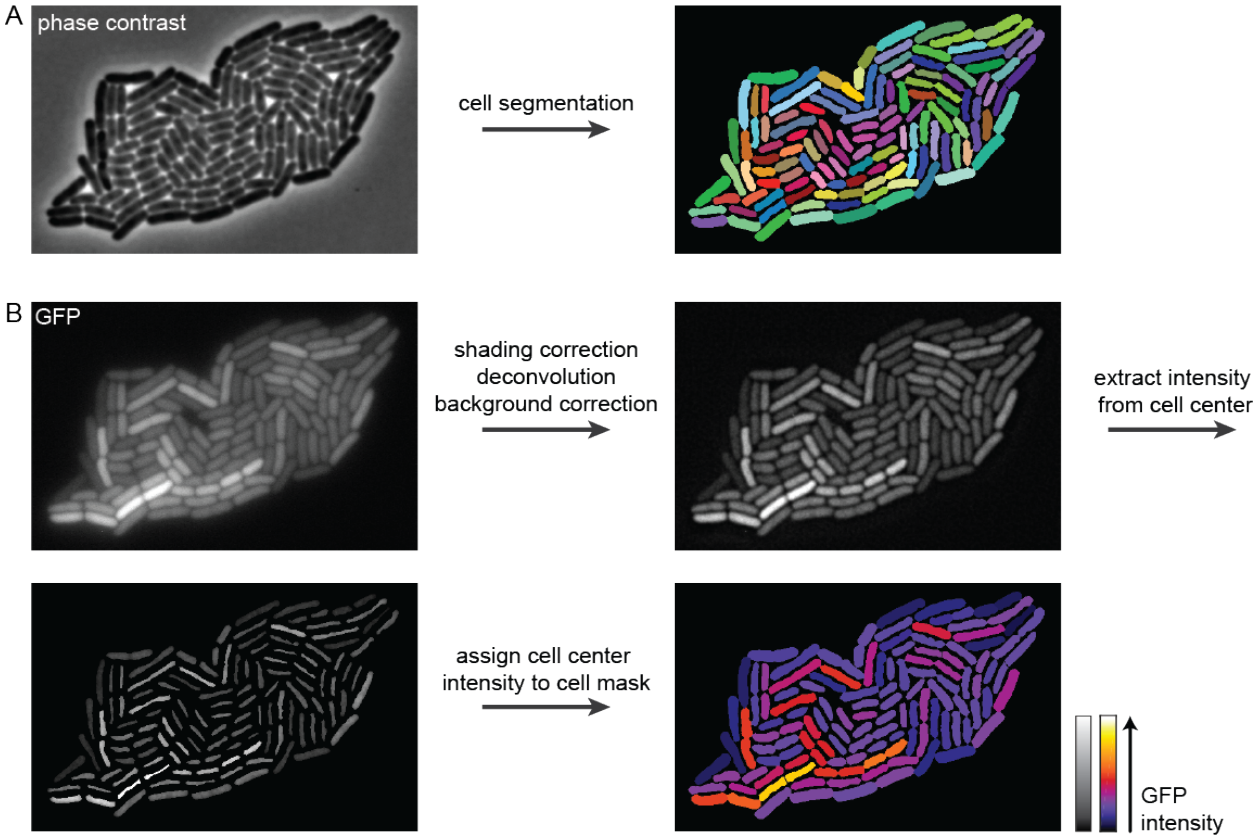

Figure S1. Related to Figure 1. Image processing pipeline. A) Phase contrast images were segmented using Schnitzcells (Young et al. 2011) to find the cell masks (right). For some reporters, segmentation was done on fluorescent images (see *Methods*). **B)** GFP images (upper left) were corrected for inhomogeneities in the illumination field (shading correction), diffraction (deconvolution), and background fluorescence (background correction), the resulting corrected image is shown (upper right). The mean fluorescence intensity was determined in the cell center only (lower left) and assigned to the entire cell mask (lower right). The figure shows an example of a microcolony with a transcriptional reporter for *cib*.

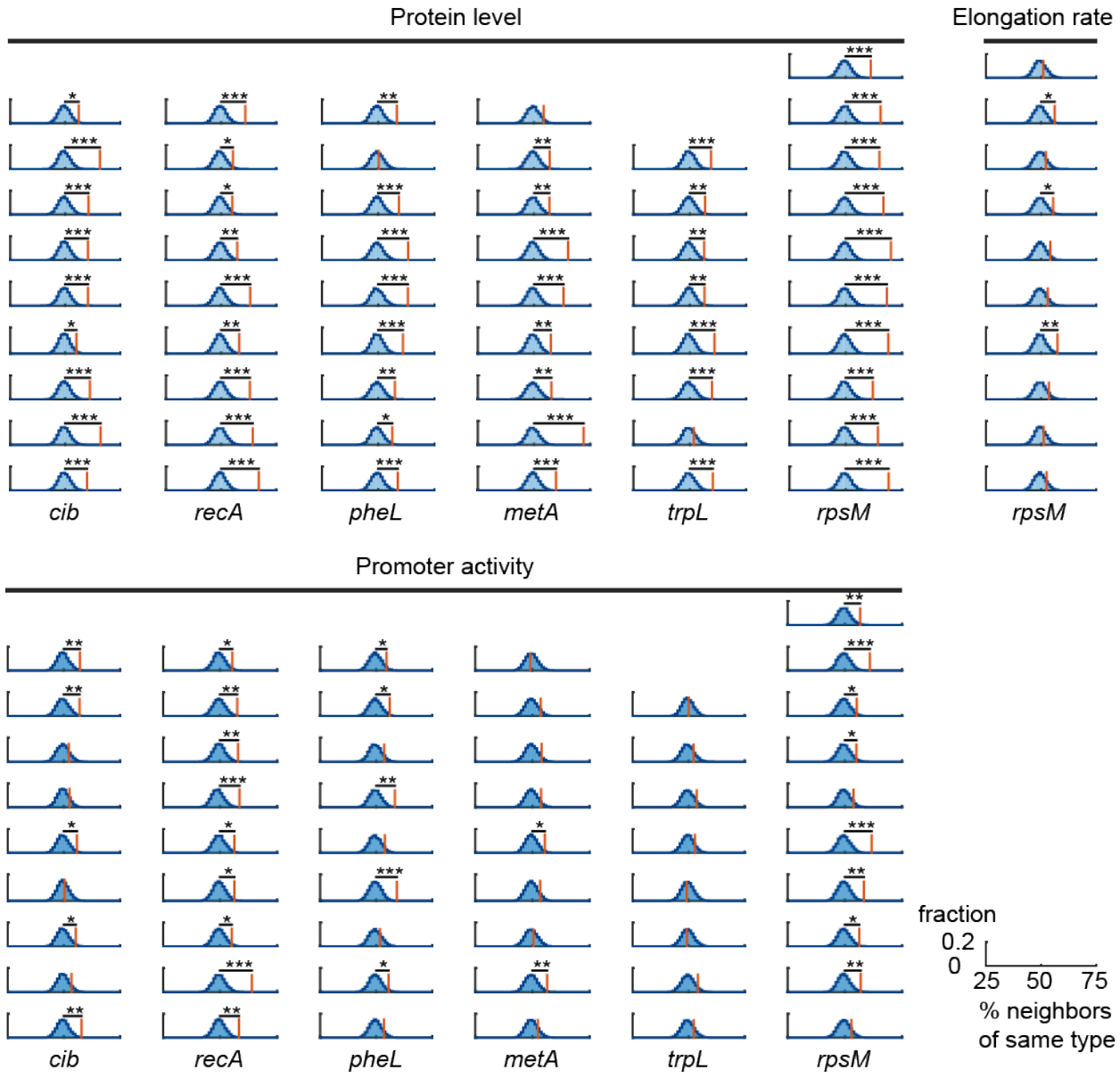

Figure S2. Related to Figure 1. Neighboring cells have similar expression levels. For all measured phenotypes (except *trpL* promoter activity) we observed that neighbors are significantly more similar than can be expected by chance ($p<10^{-6}$, except for: elongation rate ($p=3\cdot10^{-5}$), *metA* promoter activity ($p=3\cdot10^{-4}$), and *trpL* promoter activity ($p=0.05$), Chi-squared test on combined p-Values using Fischer's method). Spatial correlations in phenotypes were quantified using a randomization test: cells were grouped into two clusters (intensity above or below median intensity). Each cell was compared to all its neighbors and the percentage of neighbors of the same cell type was calculated (red line). This procedure was repeated 10^4 times after randomly permuting measured intensities among all cells (blue distribution). Each column shows 8 to 10 replicate microcolonies of 117-138 (mean=128) cells for each reporter. Data is shown for three phenotypes: protein level (upper left), elongation rate (upper right), and promoter activity (bottom). *p<0.05, **p<0.01, ***p<0.001, randomization test.

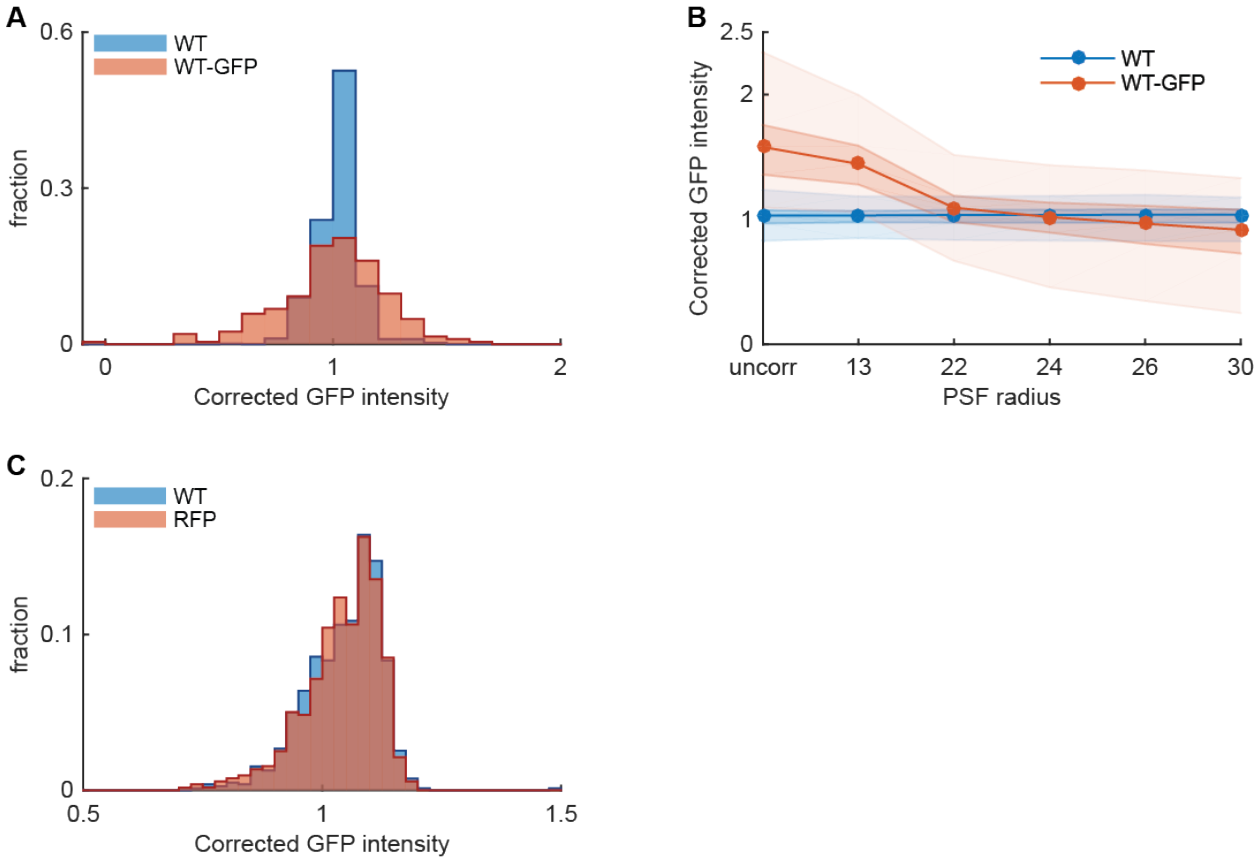

Figure S3. Related to Figure 1. Correcting for fluorescence halos and bleed-through. A) Image correction fully compensates for fluorescence halos. Cells expressing GFP (MG1655+pGFP) were mixed with non-fluorescent wild-type cells (MG1655). There is no significant difference in GFP intensity between isolated wild type cells (WT, blue, median=1.04, interquartile range=(0.97,1.08), n=1118) and wild type cells neighboring a GFP labeled cell (WT-GFP, red, med=1.02, IQR=(0.89,1.13), n=206; p=0.47, Wilcoxon rank sum test). For this figure a point spread function of 24x24 pixels was used (see panel B). **B)** The radius of the point spread function (PSF) strongly affects the accuracy of the deconvolution routine. When the radius is too small (<24) the deconvolution is not strong enough to remove halos from bright cells, i.e. the corrected GFP intensity in WT cells next to GFP labeled cells (red curve) is higher than the intensity in isolated WT cells (blue curve). When the radius of the PSF is too large (>24) artifacts are introduced causing dark halos to appear around bright cells, thus leading to an over correction: i.e. the corrected GFP intensity in WT cells next to GFP labeled cell (red curve) is lower than the intensity in isolated WT cells (blue curve). When a radius of 24 pixels is used, a close to perfect correction is obtained (see panel A). All images are corrected using a PSF of this size. Uncorr: gfp intensities for images that are only background corrected; no shading correction or deconvolution has been performed. Points: median values, dark shaded region: 25-75 percentile range, light shaded region: 2.5-97.5% range. **C)** There is no detectable fluorescence bleed-through from the RFP channel into the GFP channel. Cells expressing RFP (MG1655+pRFP) were mixed with non-fluorescent wild-type cells (MG1655). There is no significant difference in corrected GFP intensities for wild type cells (WT, blue, med=1.06, IQR=(1.00,1.10), n=781) and for RFP labeled cells (RFP, red, med=1.05, IQR=(1.00,1.10), n=517; p=0.51, Wilcoxon rank sum test).

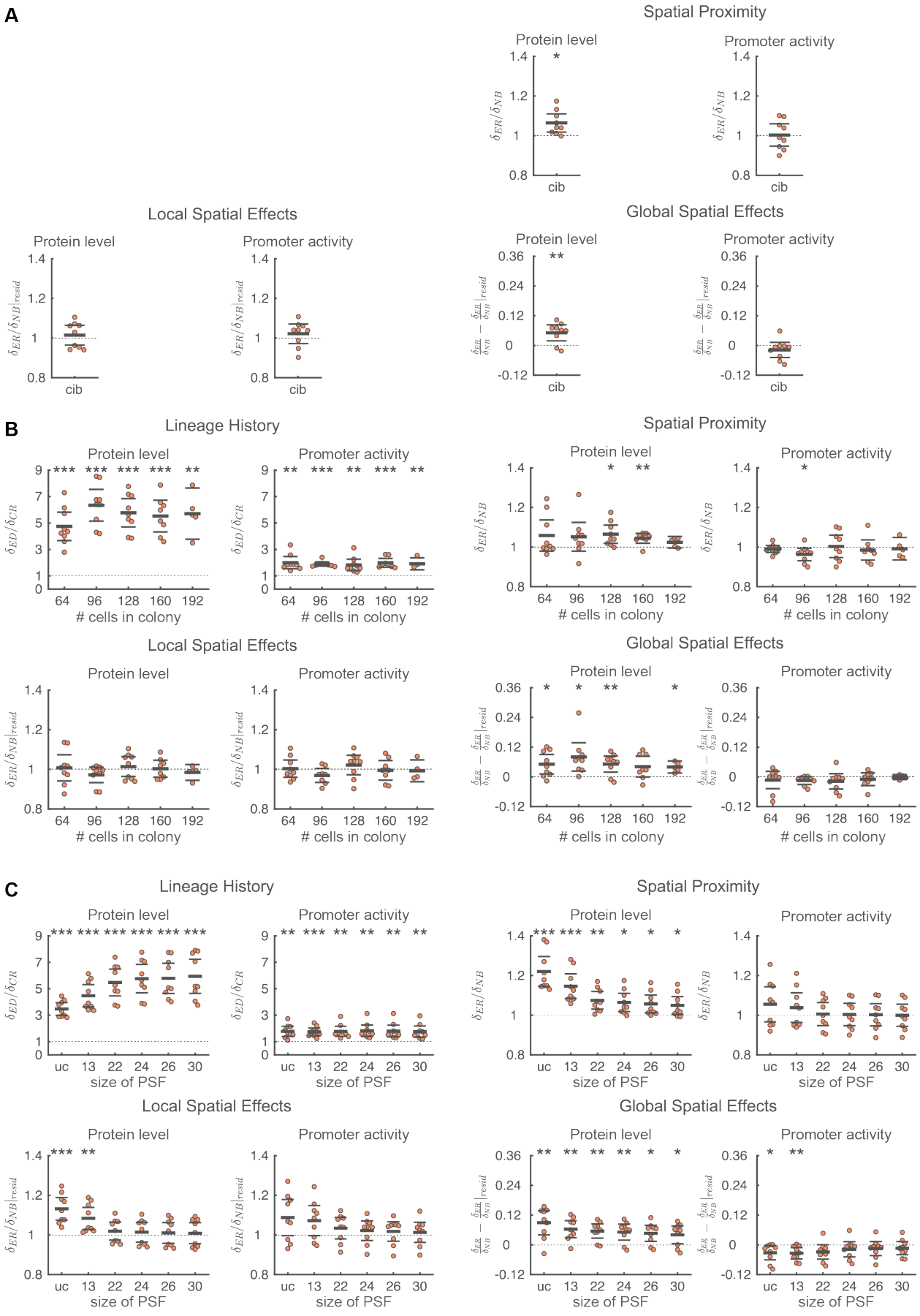

Figure S4. Related to Figure 3. Testing robustness of statistic. A) Robustness to choice of equally-related cells. Same as Figure 3B-D, except that all equally related cells are considered instead of only the most distant one. Specifically, the median phenotypic distance to all equally related cells is used (see *Methods*). **B).**

Robustness of statistic to time of analysis. Same as Figure 3A-D, except that the analysis is repeated for 5 different colony sizes. **C). Robustness of statistic to variation in deconvolution procedure.** Same as Figure 3A-D, except that the statistics were calculated for uncorrected intensities (uc) and for deconvoluted data where point spread functions (PSF) of various sizes were used (see also Figure S3B). PSF radii of 13 (strong under correction), 22 (mild under correction), 24 (optimal correction), 26 (mild over correction) and 30 pixels (strong over correction) were used. The effect of spatial proximity is systematically over estimated for uncorrected and strongly under corrected data (PSF radius=13), but is robust to mild under and over correction (PSF radius between 24 and 30). Uncorrected intensities were calculated from images that are background corrected, but where no shading correction or deconvolution has been applied. **A-C)** Each point corresponds to a microcolony with 117-138 (mean=128) cells, points are horizontally offset. Thick horizontal lines indicate mean, thin lines 95% confidence intervals. Dashed lines indicate the expected value under the null hypothesis. Null hypothesis rejected with: *p<0.05, **p<0.01, ***p<0.001, t-test, n=9.

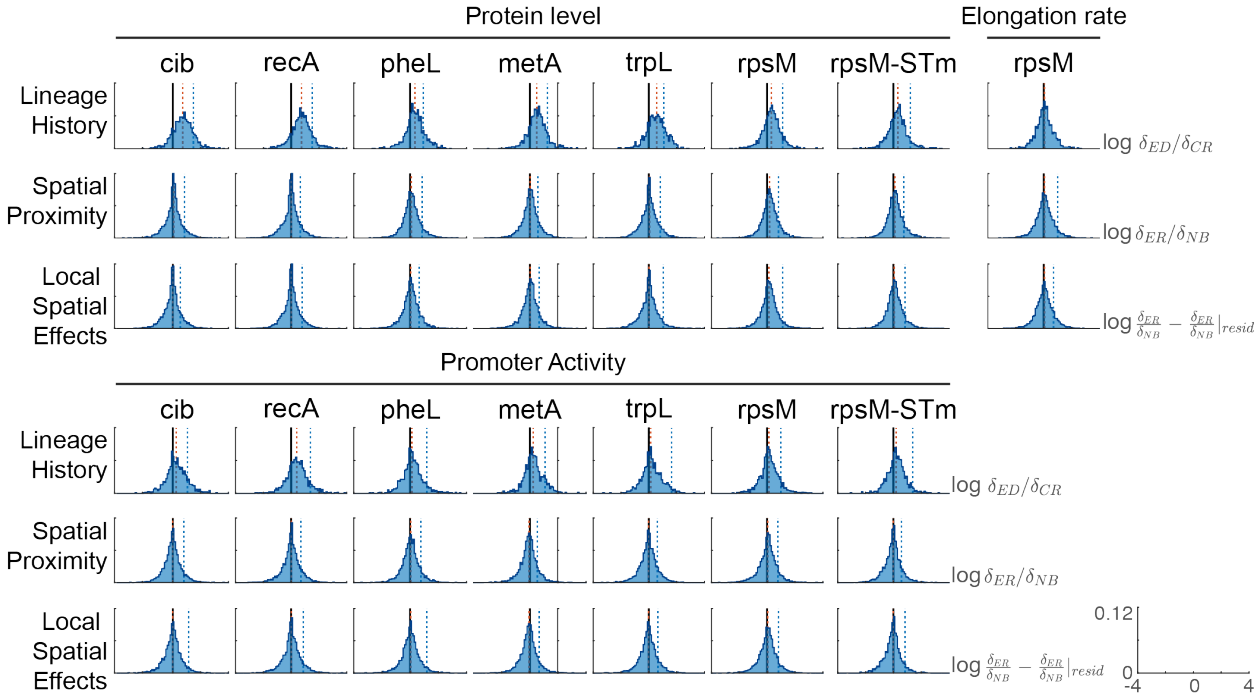

Figure S5. Related to Figure 3,5, & 6. Full distribution of statistics. The full distribution of the effects of lineage history ($\langle\delta_{ED}/\delta_{CR}\rangle$), spatial proximity ($\langle\delta_{ER}/\delta_{NB}\rangle$), and local spatial effects ($\langle\delta_{ER}/\delta_{NB}|_{resid}\rangle$) is shown for all measured reporters. Data for all replicate microcolonies was pooled together. The statistics are shown for both protein levels (top) and promoter activities (bottom), and for elongation rate for the *rpsM* reporter. The data is log transformed to visualize the long tails of the distributions. Red dashed vertical lines indicate median values, blue dashed lines mean values and black solid line indicates the expected value under the null hypothesis. Rpsm-STm: rpsM transcriptional reporter integrated in the chromosome of *Salmonella Typhimurium*.

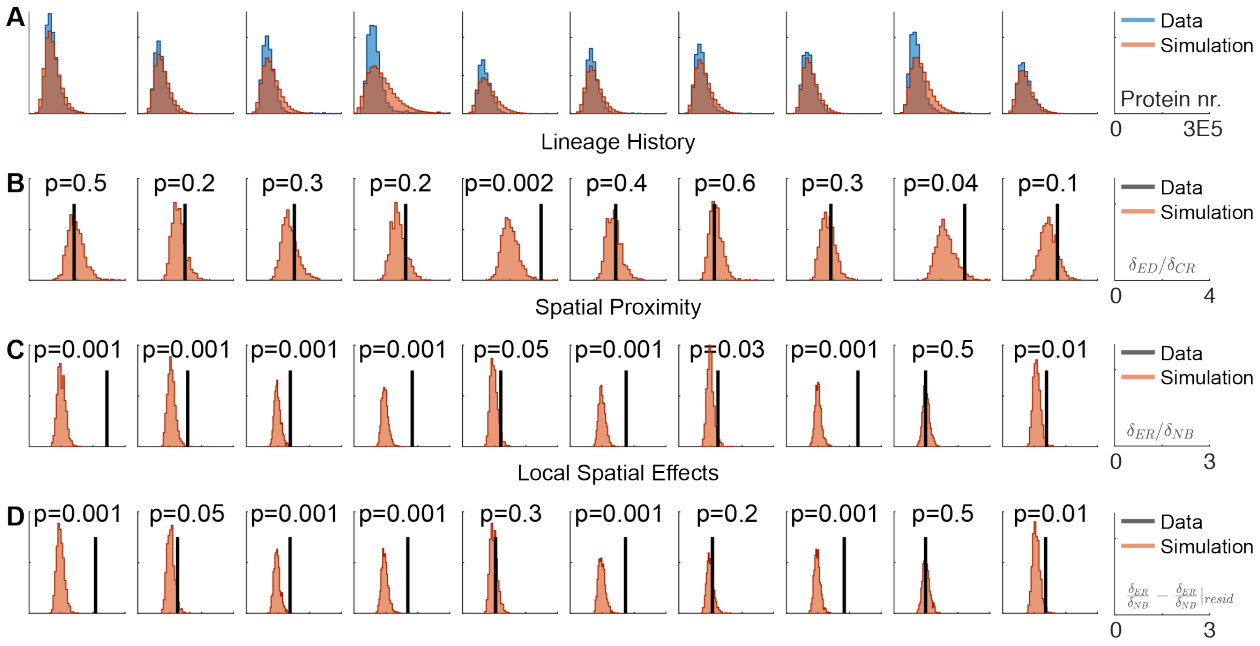

Figure S6. Related to Figure 3 and 6. Validation of statistical method using simulated data. A) Distribution of protein numbers as estimated from data (blue) and as calculated during 1000 runs of a stochastic model of unregulated gene expression (red). For each microcolony parameters were separately estimated. Each panel shows an independent microcolony. **B)** The effect of lineage history is shown for the real data (black line) and for 1000 runs of the stochastic model (red distribution). P-values show the percentage of simulated runs for which $\langle \delta_{ED}/\delta_{CR} \rangle$ of the simulated data exceeded that of the real data. **C)** The effect of spatial proximity is shown for the real data (black line) and for 1000 runs of the stochastic model (red distribution). P-values show the percentage of simulated runs for which $\langle \delta_{ER}/\delta_{NB} \rangle$ of the simulated data exceeded that of the real data. **D)** Same as C, but the statistics were calculated using the residuals of a linear regression of a cell's phenotype to the distance of a cell to the colony edge Note that in 8 out of 10 microcolonies the real statistic for linage history is not significantly different from that observed in the simulated data, showing that the stochastic model accurately incorporates lineage history effects as desired. Furthermore, in 8 out of 10 microcolonies the effect of spatial proximity is significantly stronger in the real data compared to the simulated data, showing that the measured effect of spatial proximity cannot be explained under the null-model.

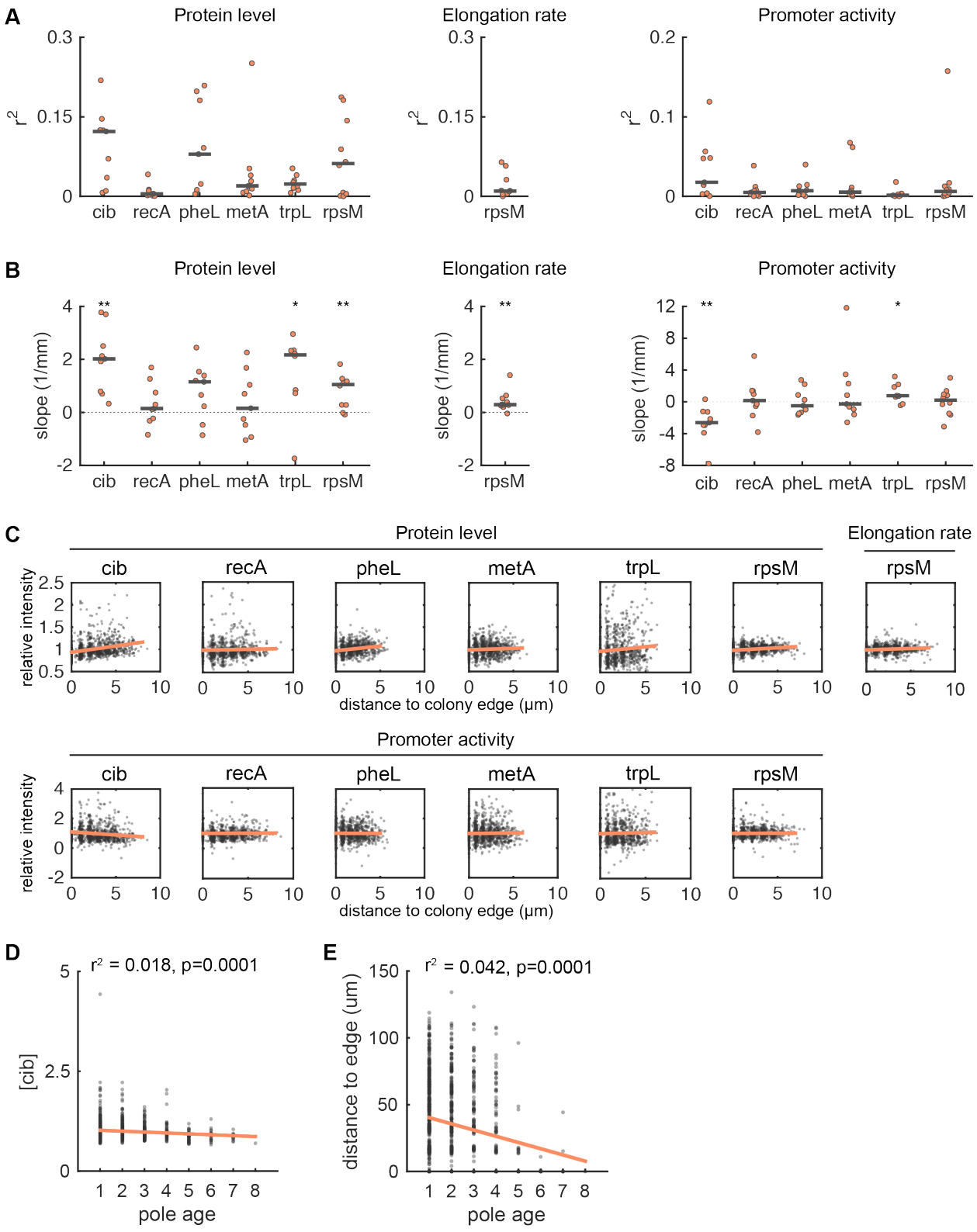

Figure S7. Related to Figure 3,5 & 6. Dependence of reporter expression dynamics on position in colony. The dependence of protein level, elongation rate, and promotor activity on the position of a cell in the colony. **A)** R-square value of a linear regression of normalized phenotype (i.e. protein level, elongation rate, or promoter activity) with the distance of a cell to the colony edge. Each point corresponds to a microcolony with 117-138 (mean=128) cells, points are horizontally offset. Thick horizontal lines indicate median value. **B)** Slope (1/mm) of a linear regression of normalized phenotype with the distance of a cell to the colony edge. Each point corresponds to a microcolony with 117-138 (mean=128) cells, points are horizontally offset. Thick horizontal lines indicate median value. Dashed lines indicate a slope of 0. Median slope significantly different from zero with: *p<0.05, **p<0.01, Wilcoxon rank sum test, *n*=9, except for *trpL* (*n*=8) and *rpsM* (*n*=10).**C)** Normalized phenotype as function

of the distance of a cell to the colony edge. Each point corresponds to a single cell. Data for 8 to 10 microcolonies are pooled together. The red line shows a linear fit to the data. In all cases, normalization is done by dividing the phenotype of a cell by the mean value in the colony. **D)** Dependence of *cib* expression level on cell pole age and **E)** Dependence of the distance of a cell to the colony edge on cell pole age. **D,E)** Each point corresponds to a single cell. Data for 9 microcolonies are pooled together. The red line shows a linear fit to the data. Person R-square and p-value (determined by randomization test) are indicated.

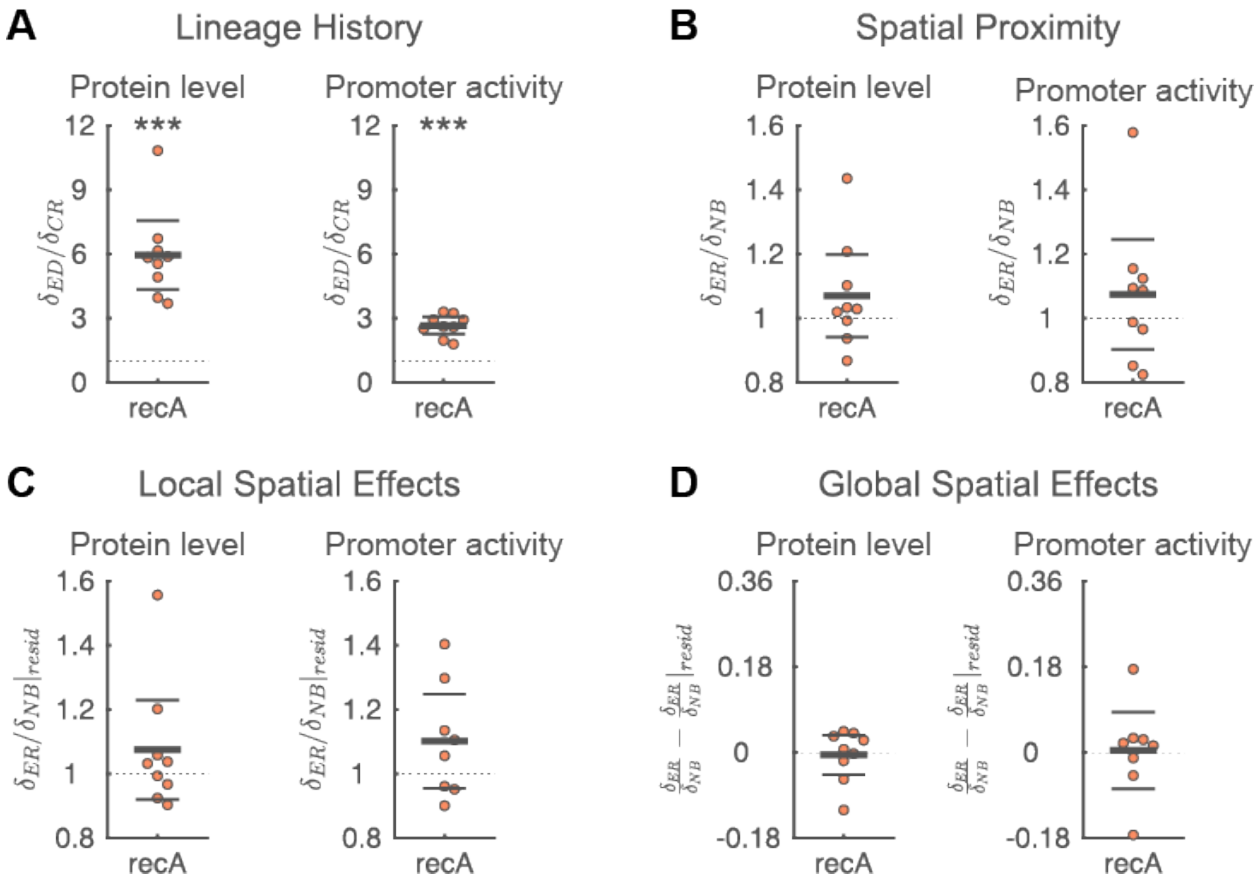

Figure S8. Related to Figure 3. Expression dynamics of *recA*. Same as Figure 3, but for the expression dynamics of the SOS response gene *recA*. Spatial correlations in the protein level and promoter activity of *recA* are solely due to shared lineage history (panel A), spatial proximity has no significant effect (panel B). Each point corresponds to a microcolony with 117-138 (mean=128) cells, points are horizontally offset. Thick horizontal lines indicate mean; thin lines 95% confidence intervals; dashed lines indicate the expected value under the null hypothesis (1 for panel A-C, 0 for panel D). Null hypothesis rejected with: *p<0.05, **p<0.01, ***p<0.001, t-test, *n*=9.

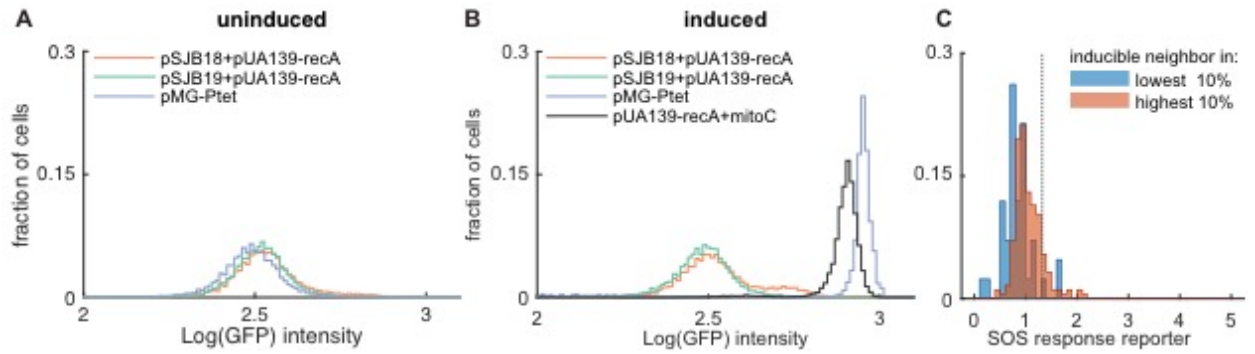

Figure S9. Related to Figure 4. Inducible SOS response. Histogram of log transformed GFP intensities measured with flow cytometry in uninduced conditions (left) and induced condition (100ng/ml AHT or 0.5µg/ml mitomycin C, right). The expression profiles are shown for 4 strains: **pSJB18+pUA139-recA** (red): containing an inducible nuclease and a GFP transcriptional reporter for *recA*; note that *recA* expression levels increase for a small fraction of cells in the presence of the AHT inducer. **pSJB19+pUA139-recA** (green): same as pSJB18, except that the plasmid also contains the Colicin E2 immunity protein that prevents nuclease activity inside the cell; note that *recA* expression levels do not increase in the presence of AHT. **pMG-pTet** (purple): cell with GFP under control of the P_{tet} promoter; note that all cells have increased GFP levels in the presence of AHT. **pUA139-recA** (grey): cell with GFP transcriptional reporter for *recA* where SOS response was induced by adding 0.5µg/ml mitomycin C (mitoC) to the culture medium; note that all cells have high levels of *recA* expression. **C)**. Same as Figure 4C, but considering only direct neighbors. The distribution of SOS response levels is shown for reporter cells (pUA66-recA) that are in direct contact with inducible cells (pSJB18 + pSV66-recA-rpsM) with low levels (dimmest 10% of inducible cells, n=59 cells) of SOS response (blue) and for reporter cells that are in direct contact with inducible cells with high levels (brightest 10% of inducible cell, n=503 cells) of SOS response (red). The distributions were obtained by pooling the data of 4 biological replicates. The dashed vertical line indicates an SOS response level of 2 standard deviation above average (mean and standard deviation were determined over all 11718 cells, irrespective of their neighborhood properties).

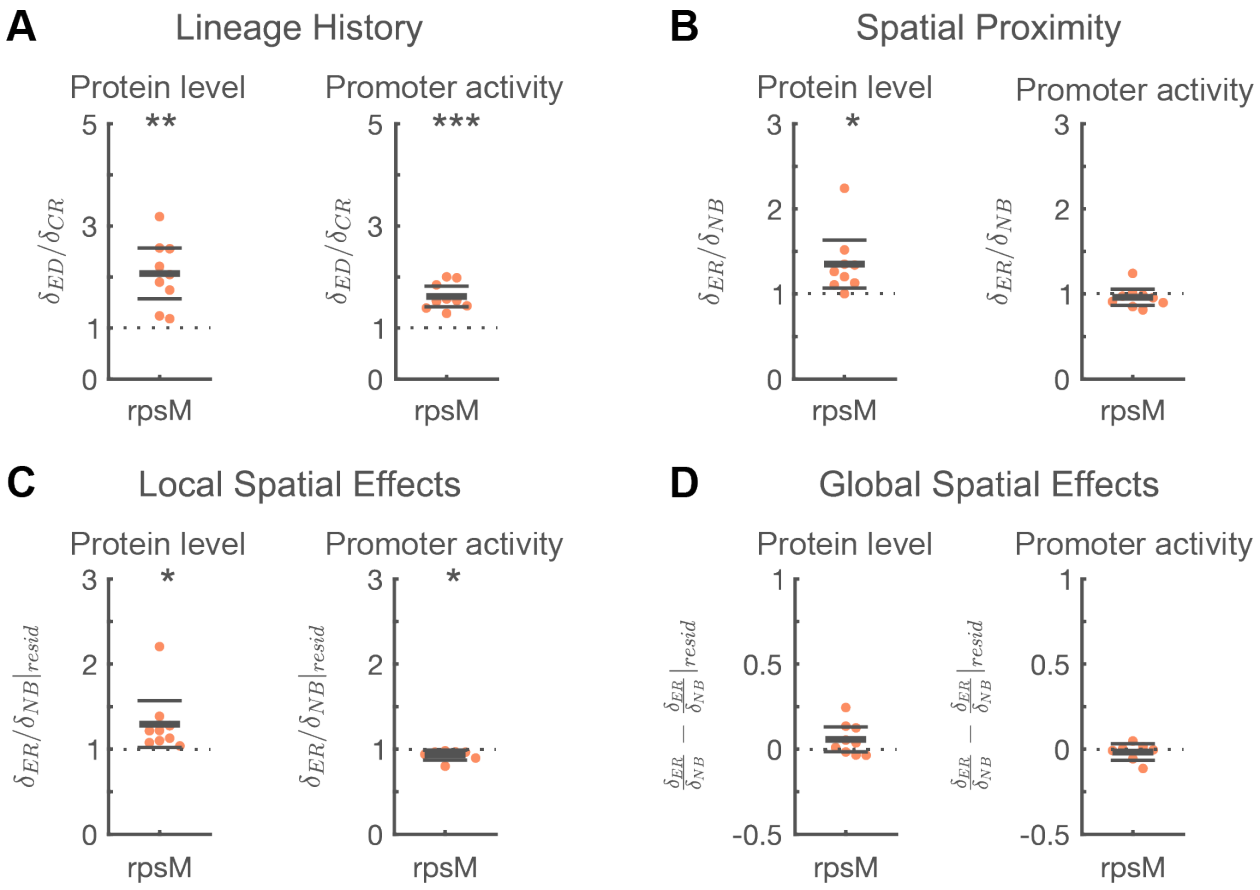

Figure S10. Related to Figure 6. Expression dynamics of *rpsM* integrated in the chromosome of *Salmonella Typhimurium*. Spatial correlations in the protein level and promoter activity of *rpsM Salmonella Typhimurium*. **A)** The effect of lineage history on *rpsM* protein levels is not significantly different between *E. coli* and *Salmonella Tm* (mean values 2.19 vs 2.07, p=0.7); the effect on promoter activity is also not significantly different between *E. coli* and *Salmonella Tm* (1.44 vs 1.63 p=0.16). **B)** The effect of spatial proximity on *rpsM* protein levels is not significantly different between *E. coli* and *Salmonella Tm* (mean values 1.60 vs 1.35, p=0.2), but the effect on promoter activity is significantly lower in *Salmonella Tm* (1.14 vs 0.96, p=0.02). **C)** The effect of local spatial effects on *rpsM* protein levels is not significantly different between *E. coli* and *Salmonella Tm* (mean values 1.44 vs 1.29, p=0.4), but the effect on promoter activity is significantly lower in *Salmonella Tm* (1.13 vs 0.93, p=0.02). **D)** The effect of global spatial effects on *rpsM* protein levels is not significantly different between *E. coli* and *Salmonella Tm* (mean values 0.16 vs 0.06, p=0.1), the effect on promoter activity is also not significantly different between *E. coli* and *Salmonella Tm* (0.01 vs -0.01 p=0.4). In all cases: (mean value pUA66-rpsM in *E. coli* vs mean value NF06 (*Salmonella Tm*), p-Value of t-test).

Table S1. Related to STAR Methods section "Strains and reporter plasmids". List of strains and plasmids used in this study. †:pM1437 and p2-camR were co-transformed into MG1655 kanR.

Name	Carrier strain	Genotype	Description	Source
pM1437†	MG1655 kanR (TB60)	pBR322-derivative, P*cib-gfpmut2*, ampR	*colicin Ib* transcriptional reporter, used together with p2-camR	(Nedialkova *et al*, 2014)
p2-camR†	MG1655 KanR (TB60)	Salmonella enterica serovar Typhimurium plasmid pColB9, camR	Contains colicin Ib operon, used together with pM1437	(Stecher *et al*, 2012)
pUA139-recA	MG1655	pUA139 derivative, P*recA-gfpmut2*, kanR	*recA* transcriptional reporter	(Zaslaver *et al*, 2006)
pUA139-trpL	MG1655	pUA139 derivative, P*trpL-gfpmut2*, kanR	*trpL* transcriptional reporter	(Zaslaver *et al*, 2006)
pUA139-rpsM	MG1655	pUA139 derivative, P*rpsM-gfpmut2*, kanR	*rpsM* transcriptional reporter	(Zaslaver *et al*, 2006)
pUA66-metA	MG1655	pUA66 derivative, P*metA-gfpmut2*, kanR	*metA* transcriptional reporter	(Zaslaver *et al*, 2006)
pUA66-pheA	MG1655	pUA66 derivative, P*pheA-gfpmut2*, kanR	*pheA* transcriptional reporter	(Zaslaver *et al*, 2006)
pSV66-recA-rpsM	MG1655	pUA139 derivative, P*recA-gfpmut2*, P*rpsM-turboRFP,* kanR	*recA/rpsM* dual transcriptional reporter	This study
NF06	S. Typhimurium SL1344	ΔinvG, ssed::aphT, P*rpsM-gfpMut2-camR::putPA*	Chromosomally integrated *rpsM* transcriptional reporter	Lab strain-collection
pGFP	MG1655	pQE derivative, P*lac-gfpmut2*, ampR	IPTG inducible *gfp*	Lab strain-collection
pRFP	MG1655	pQE derivative, P*lac-mcherry*, ampR	IPTG inducible *rfp*	Lab strain-collection
pSJB18	MG1655 nalR smR	pMG-Ptet derivative, P*tet-cea*(nuclease domain), ampR	Tet inducible nuclease domain of colicin E2	This study
pSJB19	MG1655 nalR smR	pMG-Ptet derivative, P*tet-cea*(nuclease domain)-*cei*, ampR	Tet inducible nuclease domain of colicin E2 + colicin E2 immunity protein	This study
pMG-pTet	MG1655 nalR smR	P*tet-gfp*	Tet inducible GFP	(Neuenschwander *et al*, 2007)
pTurboRFP	MG1655	*turboRFP*	Source of turboRFP coding sequence	(Hol *et al*, 2014)
pZS2-123	MG1655	Triple reporter	Source of multiple terminator sequence	(Cox *et al*, 2010)
pColE2-p9	MG1655	Colicin E2 plasmid	Source of *cea* and *cei* coding sequence	(Hol *et al*, 2014)
NF06	*Salmonella Typhimurium* SL1344	putPA::P*rpsM-gfp-cam*†	*rpsM* transcriptional reporter, incorporated in chromosome	(Hautefort *et al*, 2003)

Table S2. Related to STAR Methods section "Plasmid construction". Primer sequences used to construct plasmids pSJB18, pSJB19, and pSV66. Underlined parts are tails added to primers to provide homology for Gibson assembly. Abbreviations: (gibson): primers are used to amplify end product of Gibson assembly; tRFP: turboRFP; ter: terminator

Name	Sequence	Target	Note
	Primers for pSJB18 and pSJB19		
ColE2_C-dom_tet_fwd	5' AAT TCT AGA *ATG* AAC AAT TTA ATC GAT TTG CCC	pColE2-p9	Contains ATG start codon and XbaI restriction site
ColE2_tet_rev	5' GAT CTC GAG TTA CTT ACC CCG ATG AAT AT	pColE2-p9	Contains XhoI restriction site
ImmE2_tet_rev	5' GAT CTC GAG TCA GCC CTG TTT AAA TCC	pColE2-p9	Contains XhoI restriction site
	Primers for pSV66		
Prpsm-fw	5' GTA TCA CGA GGC CCT TTC G	pUA139-rpsM / PrpsM-tRFP (gibson)	
Prpsm-rv	5' TGT TCT CCT TGA TCA GCT CGC TCA T ATG TAT ATC TCC TTC TTA AAT CTA GAC TCG AG	pUA139-rpsM	Homology to tRFP
rfp-fw	5' ATG AGC GAG CTG ATC AAG G	pTurboRFP	
rfp-rv	5' ATT TGA TGC CTG G TCA TCT GTG CCC CAG TTT G	pTurboRFP / PrpsM-tRFP (gibson)	Homology to ter
ter-rv	5' GAT CGA GAA GGA CAC GGT TAA TAC	pZS2-123 / PrpsM-tRFP-ter (gibson)	
ter-fw	5' GGG GCA CAG ATG A CCA GGC ATC AAA TAA AAC GAA AG	pZS2-123	Homology to tRFP
Prfp-fw	5' TAA CAA ACT AGC AAC ACC AGA ACA G GTA TCA CGA GGC CCT TTC G	PrpsM-tRFP-ter (gibson)	Homology to vector
vector-fw	5' CTG TTC TGG TGT TGC TAG TTT G	pUA139-rpsM	
vector-rv	5' AGT ATT AAC CGT GTC CTT CTC GAT C CCT GCA GGT CTG GAC ATT TAT TTG	pUA139-rpsM	Homology to ter
Pgfp-fw	5' CTG GCA ATT CCG ACG TCT AAG AAA C	pUA139recA	Homologous to gfpVec-rv
Pgfp-rv	5' CAA CAA GAA TTG GGA CAA CTC CAG TG	pUA139-recA	Homologous to gfpVec-fw
gfpVec-fw	5' CAC TGG AGT TGT CCC AAT TCT TGT TG	pSV66-rpsM-rpsM	
gfpVec-rv	5' GTT TCT TAG ACG TCG GAA TTG CCA G	pSV66-rpsM-rpsM	