Published online 24 July 2012 *Nucleic Acids Research, 2012, Vol. 40, No. 18* ***8979–8992***
doi:10.1093/nar/gks694

Regulatory consequences of gene translocation in bacteria

Dena H. S. Block, Razika Hussein, Lusha W. Liang and Han N. Lim*

Department of Integrative Biology, 1005 Valley Life Sciences Building MC 3140, University of California, Berkeley, CA 94720-3140, USA

Received June 1, 2012; Revised June 22, 2012; Accepted June 25, 2012

ABSTRACT

Gene translocations play an important role in the plasticity and evolution of bacterial genomes. In this study, we investigated the impact on gene regulation of three genome organizational features that can be altered by translocations: (i) chromosome position; (ii) gene orientation; and (iii) the distance between a target gene and its transcription factor gene ('target-TF distance'). Specifically, we quantified the effect of these features on constitutive expression, transcription factor binding and/or gene expression noise using a synthetic network in *Escherichia coli* composed of a transcription factor (LacI repressor) and its target gene (*yfp*). Here we show that gene regulation is generally robust to changes in chromosome position, gene orientation and target-TF distance. The only demonstrable effect was that chromosome position alters constitutive expression, due to changes in gene copy number and local sequence effects, and that this determines maximum and minimum expression levels. The results were incorporated into a mathematical model which was used to quantitatively predict the responses of a simple gene network to gene translocations; the predictions were confirmed experimentally. In summary, gene translocation can modulate constitutive gene expression levels due to changes in chromosome position but it has minimal impact on other facets of gene regulation.

INTRODUCTION

Gene translocations can occur by several common mechanisms including intra-chromosomal recombination, transposition and inversions (1,2) resulting in the rapid generation of novel phenotypes without changing the composition of the genome. The effects of translocations on gene regulation are not well understood in bacteria and it must be stressed that findings in eukaryotes may not be applicable given the fundamental differences in their genome organization and their mechanisms of transcription and translation. This study focuses on three key features of genome organization that may be altered by gene translocation: (i) chromosome position; (ii) gene orientation; and (iii) the distance between a target gene and its transcription factor gene ('target-TF distance').

In bacteria, chromosome position is thought to alter gene expression in two ways. Firstly, gene expression decreases with the distance of a gene from the origin of replication (*oriC*) (3–5). The relationship occurs because bacterial chromosome replication is initiated at *oriC* and proceeds bidirectionally until reaching the terminus region. As a consequence, genes located near *oriC* have more copies, particularly at high rates of growth when multiple rounds of chromosome replication are initiated (6). Secondly, gene expression appears to increase and decrease periodically along the chromosome due to DNA compaction and supercoiling (7–10). Eukaryotic studies suggest that chromosome position may possibly alter other aspects of gene regulation including silencing (11,12), timing of expression (13,14), tissue distribution of expression (14,15), and the burst size of transcription events and stochastic fluctuations in gene expression ('gene expression noise') (16–18). Whether these effects also occur in bacteria is unclear given their very different chromosome structures (19), cell division rates and the absence of histones in bacteria, which are largely responsible for chromosome position effects in eukaryotes (11,16).

Gene orientation may also affect bacterial expression levels (20) and/or gene expression noise (19). This has been proposed based on the greater number of highly expressed, essential genes on the leading strand (19) and the presence of specific mechanisms in the cell that terminate the transcription of genes on the lagging strand to prevent collisions between RNA polymerases and replisomes (21). The chromosomal distance between two interacting genes

*To whom correspondence should be addressed. Tel: +1 510 643 5915; Fax: +1 510 643 6264; Email: hanlim@berkeley.edu

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

(e.g. a transcription factor and its target) is also thought to be functionally important because it determines the distance the transcription factor needs to diffuse and thus its efficacy (note: chromosomal distance does not necessarily correspond to the physical distance due to DNA folding); it has been postulated that this may explain why many transcription factor genes and genes in the same pathway are located in close proximity to their target genes (22–25).

To assess the effect on gene regulation of chromosome position, gene orientation and the target-TF distance we created a synthetic network consisting of a transcription factor (LacI) gene and its target gene (*yfp* regulated by PLlacO-1). This system allowed the measurement of gene expression in single cells and the tuning of transcription factor activity by adding varying concentrations of an inducer molecule (isopropyl-β-D-thiogalactopyranoside, IPTG) to the media. Therefore, we were able to quantify multiple features of gene regulation including constitutive expression, transcription factor binding and cooperativity, and gene expression noise. The synthetic gene network enabled the effects of each genome organizational feature to be evaluated independently and decoupled from physiological control mechanisms (26–29). In contrast, random translocation events that occur naturally or experimentally typically alter several genome organizational features simultaneously and cause a multitude of other changes that include the disruption and formation of operons (30) and the alteration of cis-regulatory sequences; this complexity would have made it very difficult to characterize the specific effects of each genome organizational feature.

The study had four main parts. The first part characterized the effect of chromosome position and gene orientation on constitutive expression and transcription factor binding. We found that chromosome position only alters constitutive expression due to differences in gene copy number and that gene orientation had no effect. We also demonstrated that in the absence of flanking terminators, the maximum expression level of a gene is more likely to be altered by local sequences at different chromosome positions. The second part created a model and this was used to predict how gene regulation is modulated by translocation of the transcription factor gene; these predictions were confirmed experimentally. This demonstrated that gene translocations can be used to rationally reprogram the output of simple gene networks. The third part examined whether target-TF distance alters transcription factor activity, and we found that it does not. The fourth part investigated whether chromosome position affects stochastic fluctuations in gene expression ('gene expression noise') and we demonstrated that it does not.

MATERIALS AND METHODS

Plasmids and strains

Details of the strains and plasmids, and the oligonucleotides used to construct them, are in Supplementary Tables S1–S3. Briefly, plasmids were constructed containing monomeric *cfp* or *yfp* (provided by R. Tsien, UC San Diego, CA, USA) under the control of the PLlacO-1 promoter which is regulated by LacI (31). Translation of *cfp* and *yfp* was controlled by the T7 10 5′ UTR sequence ('T7 10 RBS') obtained from the pET-11a plasmid (Stratagene). These plasmids were used as template for PCR amplification and the PCR products were inserted into the genome using the lambda Red system (32).

Five sets of strains were used in the study. The first set maintains *lacI* at the native position and varies the chromosome position of the target gene (PLlacO-1::T7 10 RBS::*yfp*) on the leading strand (Figure 1B): *yjbI′* (HL5135), *galK* (HL5058), *jayE′* (HL5133), *arpB′* (HL5033), *intS* (HL5043), *yfjV′* (HL5141), *yhdW′* (HL5036) and *glvC′* (HL5137). The second set also maintains *lacI* at the native position and varies the chromosome position of the target gene but the target gene is on the lagging strand (Figure 1C): *ilvG′* (HL5116), *yjbI′* (HL5042), *yjiP′* (HL5037), *ykfC′* (HL5038), *galK* (HL5131), *jayE′* (HL5059), *arpB′* (HL5136), *intS* (HL5132), *yfjV′* (HL5035), *yhdW′* (HL5134) and *glvC′* (HL5034). The third set maintains *lacI* at the native position and varies the chromosome position of the target gene (on the leading and lagging strands) but there are no terminators upstream or downstream of the target gene (Figure 3). In the fourth set we varied the chromosome position of *lacI* (with its native cis-regulatory sequences) and kept constant the position of PLlacO-1::T7 10 RBS::*cfp* at *intS* and PLlacO-1::T7 10 RBS::*yfp* at *galK* (Figure 4). *lacI* was also inserted into the chromosome using lambda Red recombinase (32). In the fifth set, we inserted *lacI* directly upstream of PLlacO-1::T7 10 RBS::*cfp* at *intS* or PLlacO-1::T7 10 RBS::*yfp* at *galK* (Figure 5).

Measurements of gene expression, mRNA concentrations and gene copy number

Measurements of gene expression were performed by fluorescence microscopy as recently described (33) except that all cells were included in the analysis. Cells were grown in LB media unless otherwise stated. mRNA concentrations were measured in duplicate independent cultures on independent blots using previously described protocols and probes (33). Gene copy number was measured by quantitative PCR using oligonucleotides to amplify the 5′ end of the *yfp* and *rrsB* (control) genes and the relative amounts of DNA were calculated as previously described (33). PCR efficiencies were determined by serial dilution of template DNA. The standard error of the mean and error propagation were calculated by standard methods (34).

Fits and parameter estimation

For the model that predicts the effect of *lacI* translocations (Equations (3)–(5)) we calculated φ to be 7.44 ± 0.68 by subtracting one from the average dynamic range (see Figure 3D). We attempted to obtain *h* by fitting Equation (2) to the data shown in Figure 1B but the error in the parameters was unacceptably high (Supplementary Table S4). Therefore we used the value for $n = 1.51 \pm 0.03$,

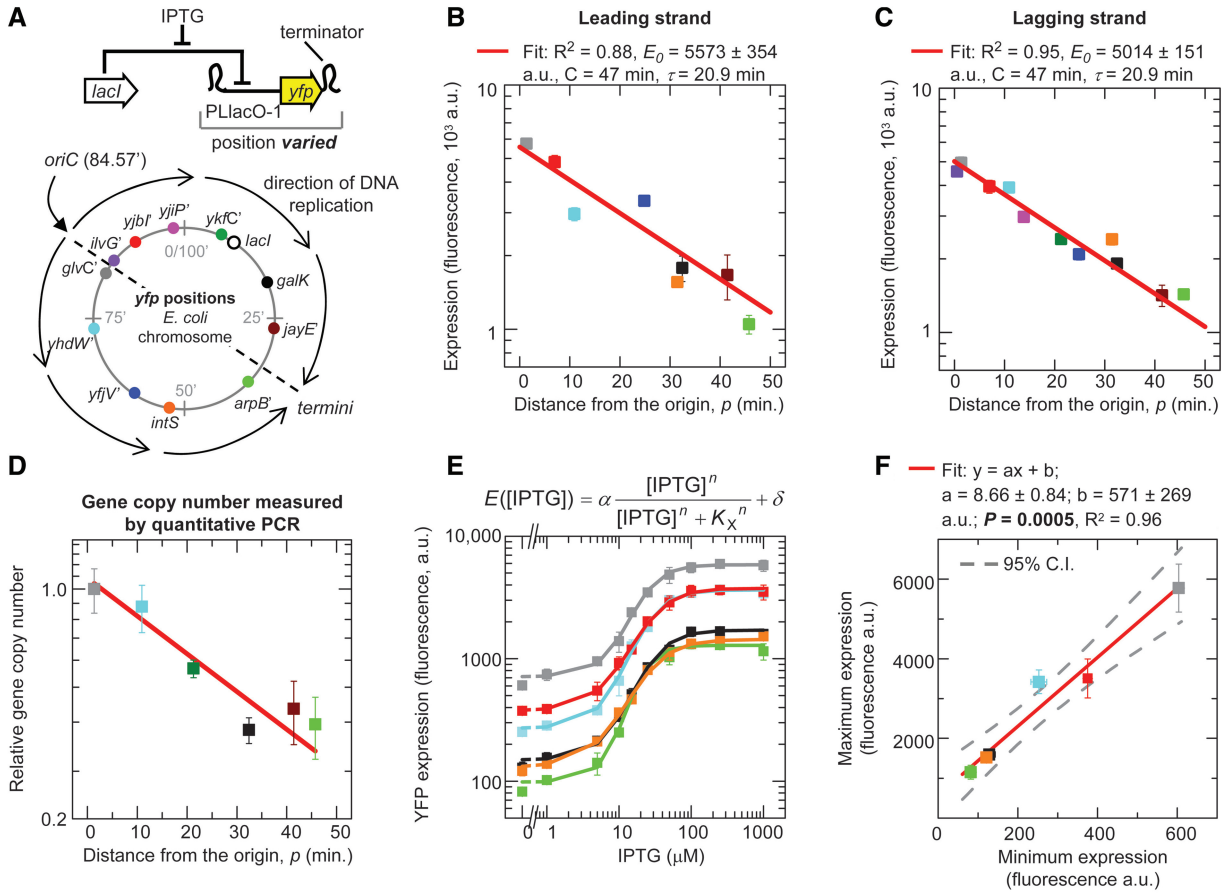

Figure 1. Gene expression at different chromosome positions and gene orientations. Error bars indicate s.e.m. of duplicate measurements. Color of the data symbols indicate positions in panel A. Strain numbers in the Materials and Methods section. (**A**) Diagram showing the gene circuit used in these experiments and different chromosome positions of *yfp*, origin and termination of replication, and *lacI*. (**B**, **C**) YFP expression on the *leading* and *lagging* strands as a function of the shortest distance of *yfp* from the origin (*oriC*). Data fitted to Equation (1). (**D**) Gene copy number as a function of chromosome position. (**E**) Induction curves for *yfp* at six positions with four on the leading strand (*galK*, *arpB'*, *intS* and *yhdW'*) and two on the lagging strand (*yjbI'* and *glvC'*). Data fitted to equation shown where α is the maximum induced amount of expression, δ is the minimum expression, n is the Hill coefficient and K_X is the IPTG concentration at half-maximal induction. The observed maximum expression = $\alpha+\delta$ = constitutive expression. (**F**) Maximum YFP expression as a function of minimum YFP expression. Data from panel E. Maximum and minimum expression were obtained experimentally (i.e. not from fits) at 1 and 0 mM IPTG.

which was the weighted value obtained from our measurements (Supplementary Table S5), as an estimate for h. This value of h was close to the Hill number for the repressor–operator interaction at the *lacZYA* promoter without DNA looping ($n = 1.45$) (35). Since the *yfp* reporter is at the *galK* position, β was set to the value for the maximum expression at this location (936 ± 121 a.u.; Supplementary Table S5). β_{lacI} which is the maximum level of LacI expression at each position was estimated by the amount of YFP expression at the same location. The *yfjV'* induction curves were similar to *lacI* at the native position indicating that LacI production at this location is likely to be similar and therefore it was used as an estimate for β_0 (225 ± 28 a.u.; Supplementary Table S5). Fits were performed using the non-linear least squares algorithm of Levenberg–Marquardt in Origin (version 7.5, OriginLab).

Stochastic simulations

Stochastic simulations were generated using Gillespie's algorithm programmed in Matlab (R2008a, MathWorks). The initial state (repressed or unrepressed) was randomly selected and the simulations ran until a steady state protein concentration was achieved. Switching between the repressed and unrepressed states and transcription events in the unrepressed state were stochastic. We also included the stochastic generation of proteins from the mRNA and stochastic degradation events for the proteins and mRNAs. The rate constants for translation (k_P) and mRNA (k_{-M}) and protein (k_D) degradation had the same values for all simulations ($k_P = 1$ protein (mRNA·time)$^{-1}$, $k_{-M} = 0.05$ time^{-1}, $k_D = 0.02$ time^{-1}). Other parameter values are in the figure legends.

RESULTS

Gene expression at different chromosome positions and gene orientations

In the first set of experiments, an identical gene (PLlacO-1::T7 10 RBS::*yfp*) was placed at eight chromosome positions: *intS* (53.1′), *galK* (17.0′), *jayE′* (26.0′), *arpB′* (38.8′), *yfjV′* (59.7′), *yhdW′* (73.7′), *glvC′* (83.2′)

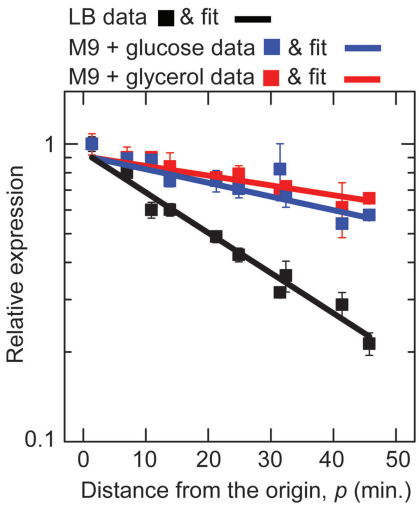

Figure 2. Effect of chromosome position on gene expression depends on cell growth rates. Error bars indicate s.e.m. of duplicate measurements. Normalized gene expression as a function of the shortest distance from the target gene to the origin (*oriC*) in three different media. Chromosome positions are on the leading (*galK*, *arpB′*, *intS* and *yhdW′*) and lagging (*yjbI′*, *yjiP′*, *ykfC′*, *jayE′*, *yfjV′* and *glvC′*) strands (strain numbers in Materials and Methods section). Gene expression was normalized to the level measured at the position closest to the origin (*glvC′*). The predicted function is determined by Equation (1). The function was completely determined by the values for C and the doubling times (τ) which were obtained independently of this plot (Supplementary Figure S1). Intercept is 1.0 because the data were normalized.

and *yjbI′* (91.6′) (Figure 1A). We chose the first two positions because they are well characterized reference sites (36) and the latter six positions because they harbor pseudogenes and therefore their replacement with *yfp* should have minimal physiological impact. *yfp* was inserted with flanking upstream and downstream terminators on the leading strand and its transcription was controlled by the PLlacO-1 promoter which is repressed by LacI encoded by *lacI* at the native position (7.9′). In addition, 78 nucleotides of the T7 bacteriophage 10 gene were fused to the 5′ end of *yfp* to enhance translational efficiency. *yfp* transcription was varied by adding different concentrations of inducer (isopropyl-β-D-thiogalactopyranoside, IPTG) to the media and its expression was quantified by measuring YFP fluorescence.

We measured the maximum YFP expression (1 mM IPTG) at the different positions on the leading strand and plotted it as a function of p, the shortest distance from its position to the origin of replication (*oriC* at 84.6′) (Figure 1B). We found that gene expression decreased with p. This finding is consistent with Cooper and Helmstetter′s model (4,6,37,38) which predicts that gene copy number decreases with increasing p due to chromosome replication. Assuming that gene expression (E) is proportional to the gene copy number then

$$E(p) = E_0 \cdot 2^{-p\frac{C}{\tau}} \quad (1)$$

where E_0 is the steady state expression of a gene at the origin (i.e. $p = 0$) in units of fluorescence, C is the average time to replicate the chromosome and τ is the average cell doubling time. In non-synchronized, exponentially growing cells at steady state $C = 47$ min (39) and the average doubling time in our experiments was 20.9 min (Supplementary Figure S1). Equation (1) was fitted to the data with only one free parameter (E_0, which has no effect on the slope) to generate a predicted relationship between the expression level and the distance of a gene from *oriC* and it was found to be in good agreement with the data (Figure 1B).

yfp was next placed on the lagging strand at eight sites plus three other sites (*ykfC′* at 5.9′, *yjiP′* at 98.4′ and *ilvG′* at 85.1′). Again, we found that expression decreased with the distance from *oriC* and the data agreed with the relationship predicted by Equation (1) (Figure 1C). The difference in the predicted intercept for the leading and lagging strands (5573 ± 354 a.u. and 5014 a.u. ± 151 a.u., respectively) was thought to be due to measurement error between the different experiments, and this was confirmed by measuring YFP expression in both orientations at four positions in the same experiment and showing there was no consistent difference in expression (Supplementary Figure S2).

We directly tested the explanation that gene expression decreases with distance from *oriC* due to decreasing gene copy number and consequently decreasing mRNA concentrations. Gene copy number was measured by quantitative PCR (Materials and Methods section) and it was found to decrease with the distance from *oriC* as predicted (Figure 1D).

To identify the specific aspects of gene regulation that depend on chromosome position we measured the induction of gene expression at six chromosome positions by varying the IPTG concentration (Figure 1E and Supplementary Table S6). The induction curve was fitted to a standard Hill-type function (40,41) to determine the Hill coefficient (n), which measures the steepness of the curve and provides a measure of LacI cooperativity, and to determine the IPTG concentration at half-maximal induction (K_X), which is a measure of LacI affinity for the DNA. We found the Hill coefficient and the IPTG concentration at half-maximal induction did not alter with chromosome position (Supplementary Figure S3). In contrast, maximum and minimum expression levels altered with the position but their ratio ('dynamic range') was relatively constant, with maximum expression being 8.66-fold (± 0.84) the minimum expression (linear regression, $P = 0.0005$, $R^2 = 0.96$) (Figure 1F and Supplementary Figure S3). The ratio is constant because the level of LacI activity is constant (as measured by LacI cooperativity and the IPTG concentration required for half-maximal induction) and does not depend on the position of the target gene therefore it is simply proportional to the amount of 'leaky' constitutive expression.

In summary, our results show that constitutive expression but not transcription factor activity depends on the chromosome position of a target gene. As a consequence, both maximum and minimum expressions alter with the chromosome position of a target gene but their ratio is constant. We additionally demonstrate that gene regulation does not depend on gene orientation.

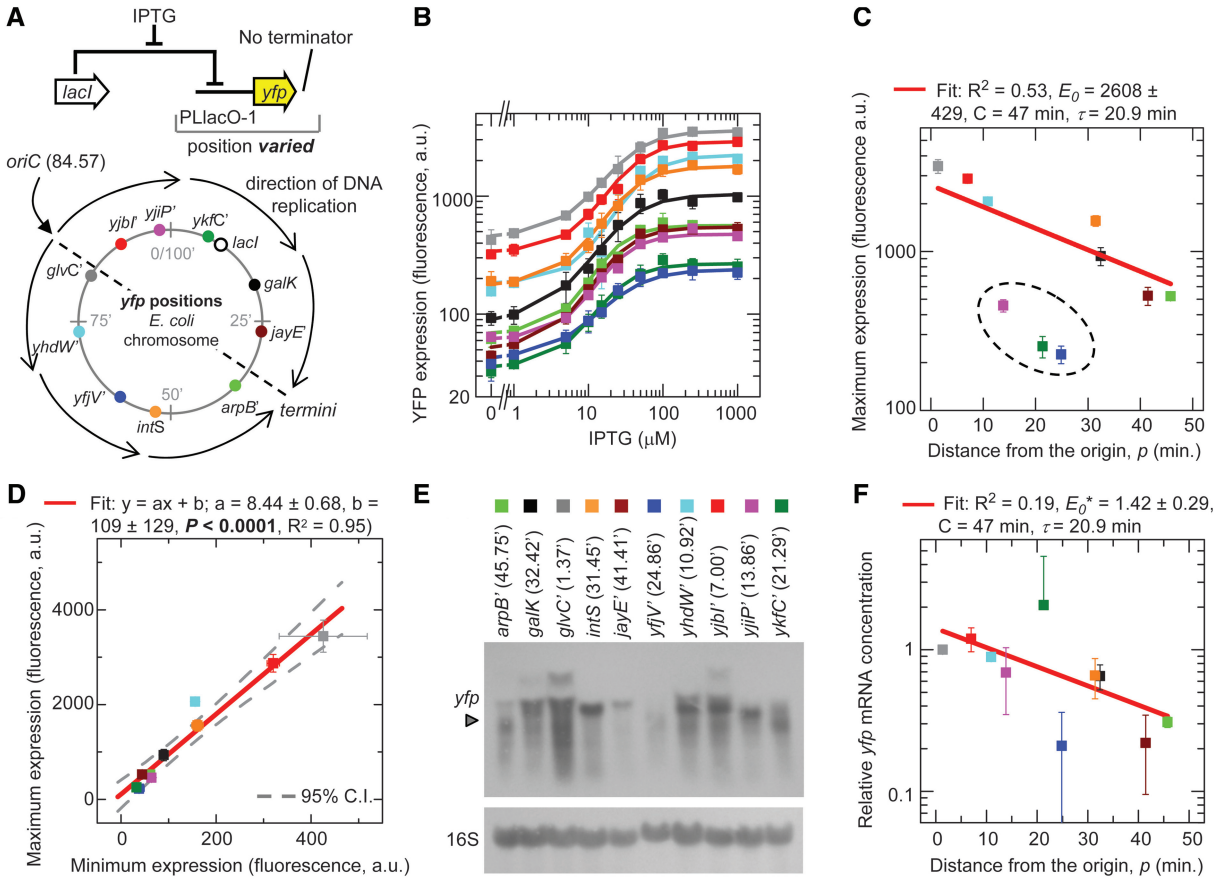

Figure 3. The expression of the target gene without flanking terminators at different chromosome positions. Error bars indicate s.e.m. Color of the data symbols throughout the figure indicates positions in panel A. (**A**) Gene circuit used in these experiments and chromosome positions of *yfp*, origin and termination of replication, and *lacI* are reshown. Strains used are: *yjbI′* lagging (HL3779), *yjiP′* lagging (HL4268), *ykfC′* lagging (HL4237), *galK* leading (HL1951), *jayE′* lagging (HL4238), *arpB′* leading (HL3776), *intS* leading (HL2821), *yfjV′* lagging (HL4236), *yhdW′* leading (HL4235) and *glvC′* lagging (HL3778). (**B**) Induction curves for *yfp* measured in triplicate. Lines indicate fits using the equation at the top of Figure 1E. (**C**) Maximum YFP expression as a function of the shortest distance from the origin. (**D**) Maximum YFP expression as a function of minimum YFP expression for data shown in panel B. (**E**) Representative northern blot of *yfp* mRNA and 16S loading control at different chromosome positions for the strains in panel B. The shortest distance from the origin (min.) is in parentheses. Contrast and brightness were adjusted solely to enhance visualization of the printed figure; no bands were obscured or selectively enhanced. (**F**) The relative *yfp* mRNA concentration obtained from independent duplicate samples in separate northern blots as a function of chromosome position. Fit is the same as panel C except the *y*-intercept value (E_0*) is the extrapolated normalized mRNA concentration at the origin.

Effect of chromosome position on gene expression depends on cell growth rates

Equation (1) shows that gene expression at different chromosome positions depends on the doubling time (τ). This prediction was tested by measuring maximum gene expression (at 1 mM IPTG) at 10 positions in 3 types of media. The media were LB, M9 + 0.4% glucose and M9 + 0.4% glycerol which resulted in doubling times (τ) of 20.9 ± 0.4, 61.9 ± 1.1 and 86.6 ± 3.2 min, respectively (Supplementary Figure S1). Expression at the different positions was normalized by the level obtained for the gene closest to the origin (*glvC′*) for each growth condition (Figure 2). These measurements showed the relative difference in expression with distance from the origin (i.e. the slope of the function) decreased as the doubling time increased. The doubling times were substituted into Equation (1) (C = 47 min and E_0 = 1) and this theoretical function, which had no free parameters, agreed with the data. These results are in agreement with an earlier study (3) and show that relative differences in gene expression due to chromosome position can be predictably tuned by altering the cell growth rate. In addition, the results provide further support for the differences in constitutive expression being due to differences in gene copy number.

Effect of local sequences on gene regulation

The above experiments examined the effect of chromosome position on a gene that is isolated from neighboring sequences by an upstream and two downstream terminator sequences. Therefore we now examine the expression of the target gene without flanking terminator sequences at 10 chromosome positions. At each position the gene was inserted in the reverse direction to the native orientation of the gene it replaced to minimize the effect of native regulatory mechanisms and to prevent the reporter gene being transcribed within existing operons (Figure 3A).

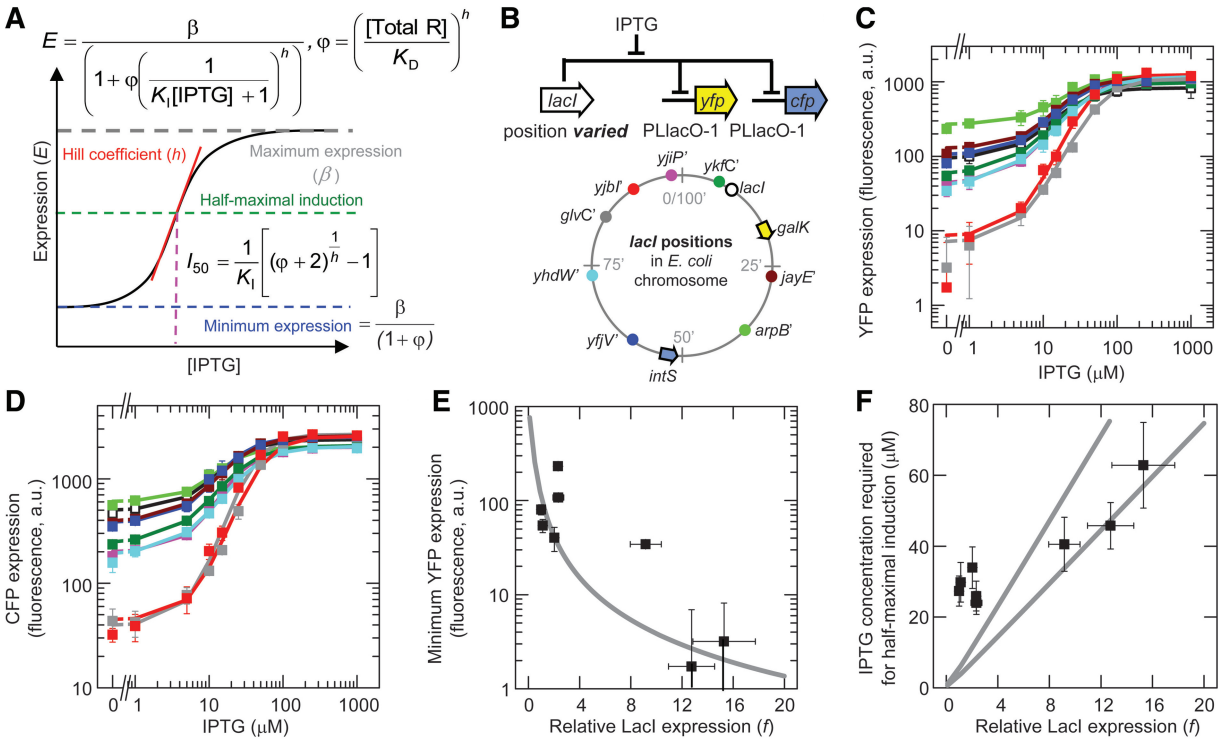

Figure 4. Using chromosome position to modulate gene regulation. Error bars indicate s.e.m. (**A**) Incorporating chromosome position effects into a model of gene regulation. β is the expression efficiency which is equal to the maximum expression, [IPTG] is the IPTG concentration, K_I is the equilibrium association constant for IPTG binding to LacI, [Total R] is the total LacI concentration, K_D is the repressor concentration required for half-maximal induction in the absence of IPTG, and h is the Hill coefficient for LacI binding to operator sites at PLlacO-1. (**B**) Diagram showing the gene circuit used in these experiments and the positions of *lacI* and the target genes (*cfp* and *yfp*). The positions of *lacI* are: *yjbI′* lagging (HL2329), *yjiP′* lagging (HL4269), *ykfC′* lagging (HL4349), *jayE′* lagging (HL4271), *arpB′* leading (HL2328), *yfjV′* lagging (HL4270), *yhdW′* leading (HL4272), *glvC′* lagging (HL4273) and at the native position (7.88′) on the lagging strand (HL1852). *cfp* and *yfp* are at *intS* and *galK* and both are on the leading strand. (**C, D**) YFP and CFP expression as a function of IPTG with *lacI* at different positions. (**E**) Minimum YFP expression as a function of the relative LacI expression. Black symbols indicate data and gray line is the relationship predicted by Equation (3). (**F**) IPTG concentration required for half-maximal induction (I_{50}) as a function of the relative LacI expression. Black symbols indicate data and gray lines are the relationships predicted by Equation (4) with $K_I = 6.3 \times 10^5$ and 1.0×10^6 M^{-1} (upper and lower lines, respectively).

Without the flanking terminators, transcription can occur into and out of the gene causing RNA polymerases to collide and prematurely terminate transcription (42), and it can result in the inclusion of additional sequences into the transcript that alter its degradation and folding. These potential effects on gene expression which are not specifically due to chromosome position were grouped together and termed 'local sequence effects'.

We measured YFP expression at different IPTG concentrations and these induction curves were fitted to a standard Hill-type function (40,41) (Figure 3B, Supplementary Table S5 and Supplementary Figure S4). We found that the maximum and minimum expression levels were the parameters that varied most among the positions, as was observed above when terminators were present. Maximum expression generally decreased with distance from the origin but there were clearly positions (circled in Figure 3C) where expression was lower than predicted by Equation (1); this indicates that local sequence effects at some positions have a strong effect on the constitutive expression from the translocated gene. The ratio of maximum and minimum expression across all the positions was relatively constant at 8.44 ± 0.68 (linear regression: $P < 0.0001$ and $R^2 = 0.95$), which was a similar value to that obtained with terminators present (compare Figure 3D and Figure 1F). Therefore we show again that transcription factor activity does not depend on the site of the target gene. In support of this the Hill coefficient and the IPTG concentration required for half-maximal induction were relatively constant at the different target gene positions.

To provide a quantitative assessment of the impact of local sequences on constitutive expression we calculated the 'relative displacement' which is the relative distance of the observed value for the maximum expression from the predicted value at each gene position (red line in Figures 1B, C and 3C). That is, the relative displacement = (observed maximum expression − predicted maximum expression)/predicted maximum expression. For genes with terminators, there was little relative displacement with almost all observed values differing by <25% from the predicted maximum expression (Supplementary Figure S5). In contrast, genes without terminators showed much greater relative displacement with most chromosome positions having observed maximum expression levels that differed by more than 25% of the

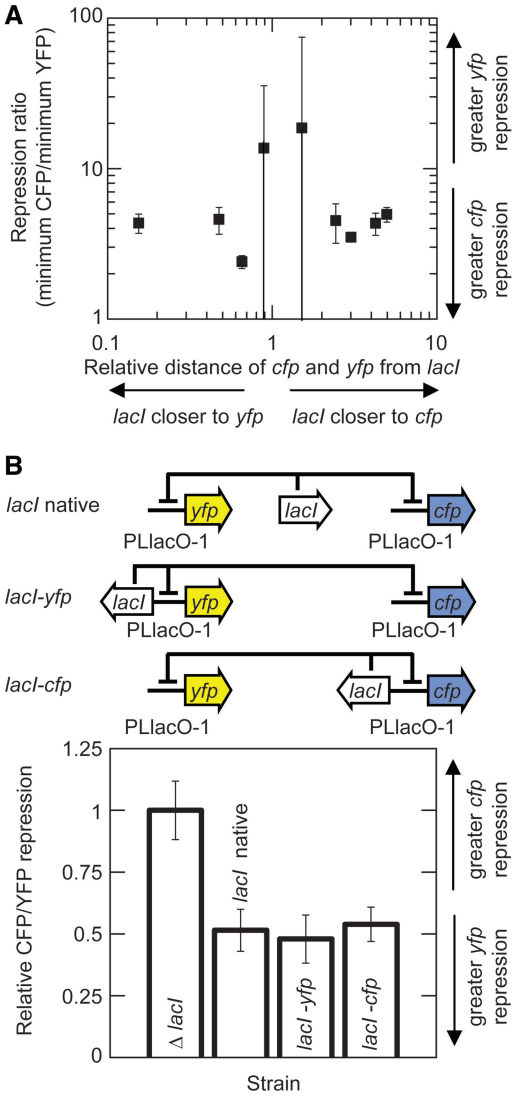

Figure 5. Target-TF distance does not affect LacI repression. Error bars indicate s.e.m. of duplicate measurements. (**A**) Repression ratio (i.e. ratio of minimum expression) of two target genes *cfp* and *yfp* under PLlacO-1 as a function of their relative distances from *lacI*. *lacI* chromosome position is varied and target gene positions are constant. CFP and YFP expression were measured at 0 mM IPTG. (**B**) Relative CFP/YFP repression (defined in main text) at very short distances with *lacI* immediately upstream of *yfp* or *cfp* (HL2620 and HL2664, respectively), at the native position (HL1852) or deleted (Δ *lacI*, HL2028).

predicted value (Supplementary Figure S5). This greater displacement indicates the larger impact of local sequences at different positions.

Northern blots were performed and quantified as previously described (33) with a probe located at the 5′ end of *yfp* to determine the concentration and length of the mRNAs in the absence of the terminators (Figure 3E and Supplementary Figure S6). We found a range of different length mRNAs within and between each chromosome position. Some mRNAs were shorter than the expected length of the T7 10::*yfp* fusion (804 nucleotides) indicating possible premature termination or partial degradation and many mRNAs were longer indicating additional upstream or downstream sequences adjoined to the coding sequence. Since there were a range of mRNA lengths we measured the total concentration of T7 10::*yfp* fusion mRNA that is at least the minimum length for the coding sequence and found that in general it decreased with distance from the origin as expected (Figure 3F). Some of the positions that had lower than expected YFP expression had mRNA concentrations that were higher than predicted (e.g. *ykfC′*, green symbol in Figure 3C and F), indicating decreased translation (possibly due to altered mRNA folding). Other positions had reduced mRNA concentrations (e.g. *yfjV′*, blue symbol in Figure 3C and F) indicating decreased transcription or increased mRNA degradation. These differences suggest that altered mRNA lengths and sequences are altering mRNA lifetimes and translational efficiency via a variety of mechanisms.

In summary, local sequences can alter gene expression by various mechanisms but their effect is limited to when terminators are absent and they only alter constitutive expression. Local sequences did not substantially influence transcription factor binding (i.e. they did not alter cooperativity or the IPTG concentration required for half-maximal induction).

Using chromosome position to modulate gene regulation

We sought to build a model of our repressor-target gene circuit that incorporates chromosome position effects and then use the model to predict how translocation of the transcription factor gene would alter the output of our circuit. The Hill-type function used above regards maximum and minimum expression as independent variables, which was useful for obtaining empirical parameters from the induction curves; but we have now shown that this is not the case. Therefore an alternate form of the Hill function was used (Figure 4A and Supplementary Methods), which is similar to that reported by others (43,44), to determine the amount of expression E for a given IPTG concentration ([IPTG], units: M):

$$\mathrm{E}([\mathrm{IPTG}]) = \frac{\beta}{\left(1+\varphi\left(\frac{1}{K_\mathrm{I}[\mathrm{IPTG}]+1}\right)^h\right)}, \text{ where } \varphi = \left(\frac{[\mathrm{TotalR}]}{K_\mathrm{D}}\right)^h \quad (2)$$

β is the expression efficiency (a.u.), K_I is the equilibrium association constant (M^{-1}) for IPTG binding to LacI, h is the Hill coefficient (unitless), [Total R] is the total repressor concentration (M) and K_D is the repressor concentration required for half-maximal induction in the absence of IPTG (M). β depends on the mRNA concentration, mRNA translation and fluorescence yield per protein (Supplementary Methods). We showed earlier that φ is independent of *yfp* position which indicates that h is also independent of *yfp* position. K_I, which defines the affinity of IPTG for LacI, is also independent of *yfp* position. It should also be noted that while Equation (2) is able to predict expression with known parameters it is not useful for obtaining parameter values due to large fitting errors (Supplementary Table S4).

Equation (2) shows that at saturating IPTG concentrations, gene expression is maximal and approximately equals the expression efficiency (β). The expression efficiency is determined by the chromosome position of the target gene (*yfp*). At [IPTG] = 0, gene expression is at its minimum and equals $\beta/(1+\varphi)$. The ratio of maximum to minimum expression (i.e. dynamic range) equals $1+\varphi$. That is, the dynamic range is independent of the expression efficiency and therefore independent of chromosome position as was observed experimentally (Figures 1F and 3D).

Equation (2) also predicts the effect of *lacI* position on target gene regulation. If altering the chromosome position of *lacI* changes the total repressor concentration by f [the ratio of LacI expression at the new (β_{lacI}) and original (β_0) positions], the new minimum YFP expression will be equal to

$$\frac{\beta}{(1+\varphi \cdot f^h)}, \text{ where } f=\beta_{\mathrm{lacI}}/\beta_0 \quad (3)$$

Because the position of *lacI* alters minimum expression but not maximum expression it will also change the dynamic range. Furthermore, the position of *lacI* will also affect [IPTG] needed for half-maximal induction ([I_{50}]) by (derivation in Supplementary Methods)

$$[I_{50}] = \frac{1}{K_I}\left[(\varphi \cdot f^h+2)^{\frac{1}{h}} - 1\right] \quad (4)$$

In summary, the model explains the observed effects of chromosome position on target gene (*yfp*) expression (i.e. altered maximum and minimum expression and a constant dynamic range) and it predicts the effect of *lacI* position on gene regulation (i.e. altered minimum expression, [I_{50}] and dynamic range).

We tested the model's predictions by placing *lacI* at eight chromosome positions without terminator sequences and measuring its effect on two target genes (*yfp* at *galK* and *cfp* at *intS*) (Figure 4B). If the model is predictive for genes without terminator sequences then the model is robust. We found that as predicted, minimum expression changed with the position of *lacI* and the maximum expression did not (Figure 4C and D). We then examined whether minimum expression decreased as a function of the relative LacI expression according to the relationship specified by Equation (3). LacI expression at each position (β_{lacI}), which was assumed to be proportional to maximum YFP expression at the same location, was divided by expression at the native position (β_0) to yield the relative LacI expression. β_0 was estimated to be equivalent to that at *yfjV′* ($\beta_0 = 225 \pm 28$ a.u.) since *lacI* at this position produced similar levels of repression (Figure 4C). Values for φ (7.44 ± 0.68), h (1.51 ± 0.03) and β (936 ± 121 a.u.) were obtained from the measurements for *yfp* without terminators (Materials and Methods section) and substituted into Equation (3). This yielded the predicted relationship between minimum YFP expression and relative LacI expression (with no free fitting parameters), which agreed with the observed values (Figure 4E).

We next examined whether the IPTG concentrations required for half-maximal induction matched the predictions of the model. Values for φ, h and a lower and upper value for the equilibrium association constant for IPTG binding to LacI [$K_I = 6.3 \times 10^5$ and 1.0×10^6 M^{-1}, respectively, which are calculated from the reciprocal of the equilibrium *dissociation* constant (45–47)] were substituted into Equation (4). The resulting functions were consistent with the data at high levels of LacI (Figure 4F). At low LacI concentrations, the [IPTG] required for half-maximal induction was higher than expected. A probable explanation is that the ratio of intracellular to extracellular IPTG is not constant but varies with the external [IPTG] due to active transport by membrane pumps (48,49) and positive feedback regulation (50).

The target-TF distance does not affect LacI repression

It is increasingly recognized that the intracellular environment is heterogeneous and the sites of gene transcription and translation are important (51,52). If the diffusion distance between the repressor gene (which is a site of repressor production due to the coupling of transcription and translation) and its target gene is important, we should expect the relative ratio of repression at two target genes to vary depending on their relative proximity to *lacI*. We calculated the ratio of maximum *repression* at two target genes by taking the ratio of their minimum expression (termed the 'repression ratio') with *lacI* at different chromosome positions (Figure 5A). The chromosome distance does not necessarily represent the physical distance between *lacI* and each target gene but if diffusion is important then the repression ratio should vary substantially with *lacI* position. However, we found the repression ratio was relatively constant for all *lacI* positions, indicating that diffusion is not a major contributor to gene regulation, at least over relatively large distances (note: the repression ratio $\neq 1$ because *cfp* and *yfp* do not have identical transcription, translation efficiencies, mRNA and protein degradation rates and/or quantum yields per protein).

We then examined whether there was increased effectiveness of LacI repression over very short distances as has been proposed (23) by placing *lacI* immediately upstream of *cfp* or *yfp* (Figure 5B). To measure the amount of repression we used the 'relative CFP/YFP repression' rather than the repression ratio to correct for any local effects of the *lacI* gene on the promoter of the adjacent *cfp* or *yfp*. The relative CFP/YFP repression was calculated from the ratio of maximum to minimum CFP expression (at 1 and 0 mM IPTG) divided by the ratio of maximum to minimum YFP expression and the resulting value was normalized to the CFP/YFP ratio in the absence of LacI to compensate for a small effect that IPTG has on CFP and YFP expression independently of LacI. In essence, we are taking the ratio of the dynamic ranges for CFP and YFP expression and in the presence of *lacI* there is a greater dynamic range for YFP than CFP, which causes the relative CFP/YFP repression to be <1. If a short diffusion distance is important then the relative CFP/YFP repression should be high when *lacI* is close to *cfp* and low

when *lacI* is close to *yfp*. However, we found no difference in the relative CFP/YFP repression with *lacI* at different positions (Figure 5B). Therefore, the diffusion distance between a transcription factor gene and its target gene was not important even at short distances.

Effect of chromosome position on gene expression noise

We investigated the effect of chromosome position on gene expression noise in our simple circuit. Gene expression noise was quantified by the coefficient of variance (C.V.), which is simply the standard deviation (S.D.) of the expression divided by the mean with background autofluorescence subtracted (53). We first analyzed the data from the induction curves with *yfp* at different chromosome positions where mean expression varies with IPTG concentration and chromosome position. We calculated the C.V. for all the cells in the sample and for only those cells within 2 S.D.s from the mean; the latter was to demonstrate the values are not unduly determined by a small number of outliers. For each chromosome position, increasing the mean expression by increasing the IPTG concentration resulted in a decreasing 'C.V. trend'. Across chromosome positions, this decreasing trend shifted to the left or right but not up or down (Figure 6A). That is, the chromosome position of a target gene alters mean expression with little effect on gene expression noise.

When *lacI* was placed at different chromosome positions we found the C.V. values for a given mean expression level were the same (Figure 6B). That is, translocation of *lacI* does not result in a shift up or down for the C.V. trend therefore it does not increase or decrease gene expression noise. The *lacI* position does affect the total LacI concentration and therefore minimum expression level; this establishes the lower bound of the downward 'trend' for each position and thus the maximum C.V. value attainable in the absence of IPTG (Figure 6B and Supplementary Figure S7).

A simple stochastic model of gene expression was created to examine the noise sources. In the model, switching occurs between a 'repressed' state with LacI bound and an 'unrepressed' state with LacI unbound at rates determined by k_{-1} and k_1 (note: k_{-1} is dependent on the LacI concentration) (Figure 7A). Initially, we assumed the gene copy number (which depends on chromosome position) simply alters the number of mRNAs produced per unit time and thus the magnitude of the rate constant for transcription in the unrepressed state (k_M). The model also includes mRNA translation and degradation of mRNAs and proteins which do not vary. It has been reported that when transcription factor binding and unbinding occur at rates that are low compared with the transcription rate, 'bursts' of transcription occur and this becomes the dominant source of noise (16,17,54,55). Under these conditions (i.e. $k_1 = k_{-1} = 0.01k_M$; note: all have units of min^{-1} because the LacI concentration is included within k_{-1}), altering mean expression by varying LacI binding or LacI concentration has a large effect on the C.V. compared with varying mean expression by varying the transcription rate k_M (Figure 7B).

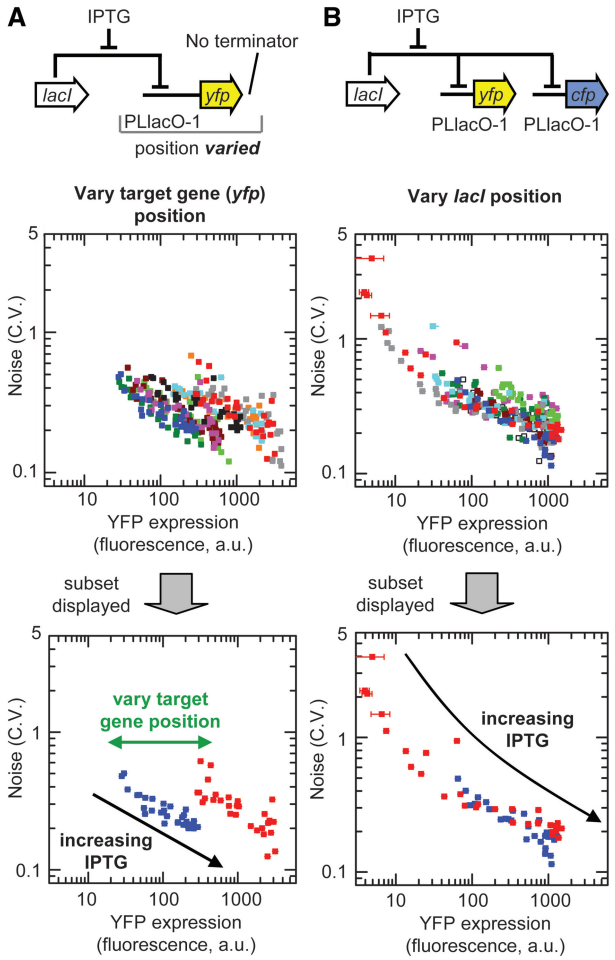

Figure 6. Effect of chromosome position on gene expression noise. Error bars indicate s.e.m. of each sample. Color of the data symbols indicates the chromosome positions in Figure 1A. Gene expression noise as a function of mean YFP expression at different target gene (*yfp*) positions (**A**) different *lacI* positions (**B**). At top we show the gene circuit for each experiment. Upper plots show all positions and lower plots display only two positions. Gene expression noise for *cfp* is displayed in Supplementary Figure S7.

Experimentally this was observed as an increase in maximal C.V. that accompanies lower minimum expression levels due to higher LacI concentrations. These simulations show that altering mean expression by changing the chromosome position and consequently the mRNA concentration has minimal effect on gene expression noise (compare Figure 7C and bottom panel Figure 6A) which is in agreement with our observations.

We now compare two scenarios for transcription among multiple copies of the target gene (Scenarios 1 and 2, Figure 7D). To be clear, we refer to copies of the same gene at the 'same' position which arise during DNA replication. In Scenario 1, switching between the repressed and unrepressed state, and transcription events are highly correlated because fluctuations in the LacI concentration are the dominant source of noise. This results in all the gene copies behaving the same which is essentially the scenario in the above simulation where we assumed there was one copy of the gene and varied k_M. In Scenario 2,

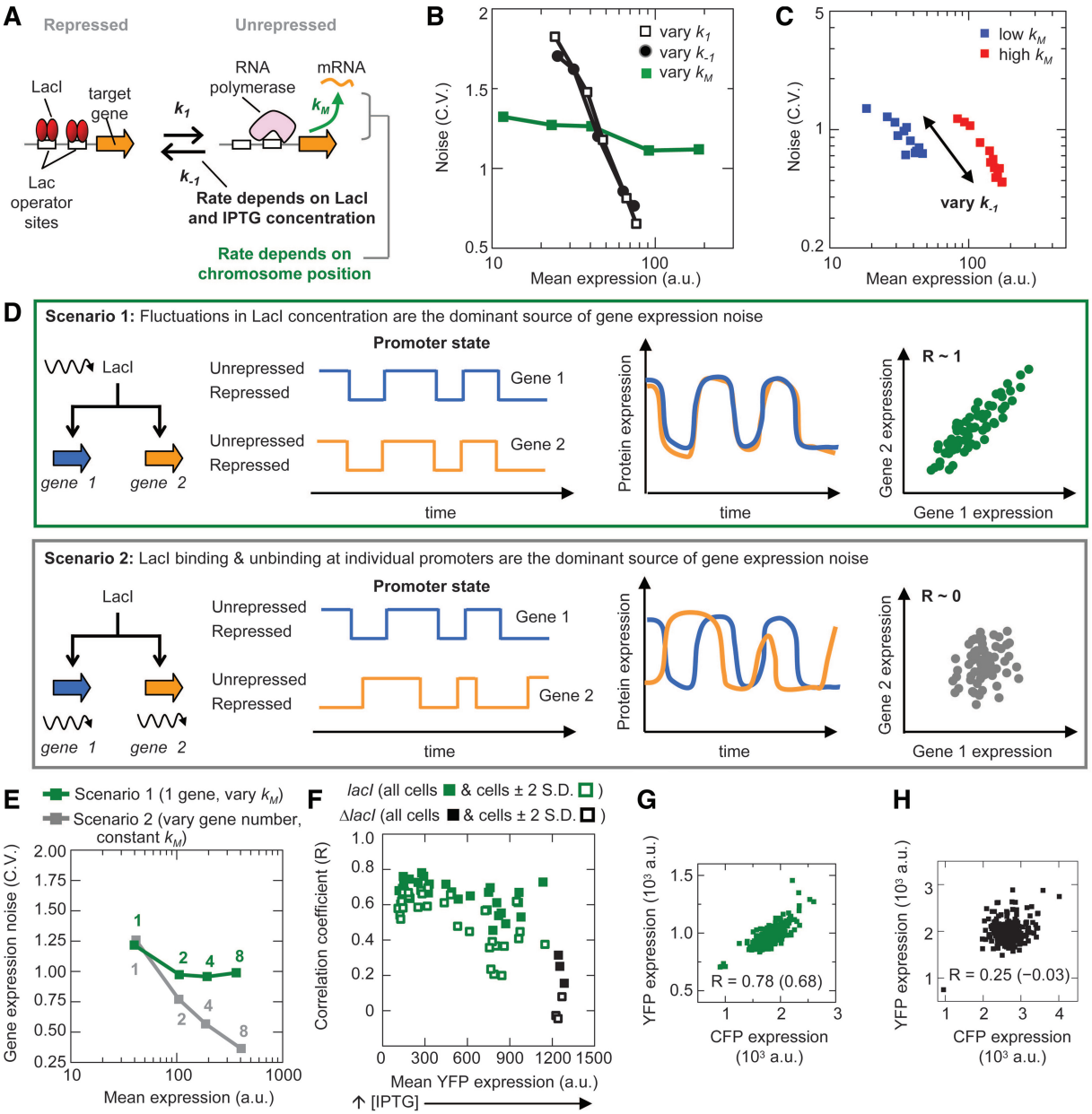

Figure 7. Chromosome position has minimal impact on gene expression noise. (**A**) Model of gene expression as described in main text. (**B**) Simulated gene expression noise as a function of mean expression at varying values for k_1, k_{-1} and k_M. Initially $k_1 = k_{-1} = 0.001$ concentration·time^{-1} and $k_M = 0.1$ concentration·time^{-1}. Each value was independently varied by 0.25-, 0.5-, 2- and 4-fold. (**C**) Varying the rate of LacI binding at two different rates of transcription. Simulation was performed as above with k_{-1} = 0.1, 0.129, 0.167, 0.215, 0.278, 0.359, 0.464, 0.600, 0.774 and 1.000 x 10^{-3} concentration·time^{-1} at two values for k_M (0.05 and 2.00 concentration·time^{-1}, blue and red symbols, respectively). (**D**) Predicted effect on the promoter state and gene expression where fluctuations in LacI concentrations (Scenario 1) or LacI binding and unbinding at individual promoters (Scenario 2) are the dominant sources of expression noise. (**E**) Stochastic simulation where mRNA concentration was varied by altering gene copy number or the value of k_M (gray and green numbers indicate fold change in gene copies or k_M, respectively). (**F**) Correlation coefficient (R) as a function of mean YFP expression. Data derived from measurements of HL1852 at different [IPTG] in Figure 4C. Solid symbols indicate R-values calculated from all cells and unshaded symbols indicate R-values calculated from only cells within 2 S.D.s of the mean. Green symbols indicate strain HL1852 and black symbols indicate strain HL2028, which is HL1852 with *lacI* deleted. (**G, H**) Representative scatter plots of CFP and YFP expression for HL1852 and HL2028. R-values are calculated from all cells or for only cells within 2 S.D. of the mean (values in parentheses).

switching and transcription events at each copy of the gene are independent and most of the noise occurs due to random switching events at individual promoters. Under this scenario, increasing the copy number of a gene by moving it closer to the origin leads to a more stable average for the occupancy of the repressed and unrepressed states. That is, increasing mRNA production by increasing the number of independent copies of a gene (thereby indirectly increasing k_M) results in a lower C.V. than simply increasing the k_M for a single copy gene

(Figure 7E). However, our experimental data found no evidence to support Scenario 2 (Figure 6A). That is, genes that are closer to the origin did not have a lower C.V. for a given mean indicating that transcription from each gene copy is highly correlated; we now show that this is most likely due to stochastic fluctuations in the LacI concentration.

If fluctuations in LacI are the dominant source of noise as the above suggests, then two genes under LacI control should increase and decrease together resulting in a positive correlation coefficient ($R >> 0$). It is not possible to measure the expression noise of different copies of the same gene at the same position. However, the above also applies to genes under LacI control at different positions. Therefore, we measured the correlation coefficient for *cfp* and *yfp* under the control of LacI at two positions that are approximately equidistant from *oriC* (*intS* and *galK*). We found that at most levels of expression R was >0.5, which is consistent with a source of noise causing CFP and YFP expression to increase and decrease together (i.e. highly correlated expression) (green symbols, Figure 7F and G). To demonstrate that LacI is responsible for *cfp* and *yfp* expression being correlated we deleted *lacI* and showed that this causes the correlated expression to decrease from 0.78 to 0.25 (black symbols, Figure 7F and H). That is, transcription becomes independent and less correlated without LacI.

Together our experimental findings and stochastic simulations show that chromosome position, which alters mean expression via its effect on gene copy number and thus the mRNA concentration, has minimal impact on expression noise compared with fluctuations in the LacI concentration. That is, changing gene position, and therefore the gene copy number and local sequence effects, has minimal impact on gene expression noise. Increasing the mRNA concentrations by increasing promoter strength without altering transcription factor binding would be expected to have a similar effect to increasing gene copy number and this has been shown to be the case in yeast (56).

DISCUSSION

This study examined the effect on gene regulation of three features of genome organization that may alter with gene translocation (chromosome position, gene orientation and the relative distance between interacting genes). Remarkably, only chromosome position had a demonstrable effect and it primarily determined the maximum and minimum expression levels of the target gene. More specifically, maximum and minimum expression levels decreased with the distance of the target gene from the origin and we showed that this was associated with a decrease in gene copy number. These findings are consistent with Cooper and Helmstetter's model (6). Surprisingly, transcription factor activity (as measured by the amount of repression, the Hill coefficient and the IPTG concentration required for half-maximal induction) was constant; which contrasts with results reported in eukaryotes. This likely reflects the prominent role that long distance regulatory mechanisms such as enhancers and silencers, and chromatin modification have in the eukaryotes [reviewed in (57)] and which are absent in *E. coli*.

Our study showed that local sequences at different chromosome positions can modulate gene regulation so that gene expression deviates from that predicted to occur solely due to changes in gene copy number. Local sequences can alter expression via alterations in mRNA length, concentration and translation. However, local sequence effects appear to be limited to modulating constitutive gene expression levels and are relatively small unless the translocated gene lacks flanking terminators. In studies that have translocated genes for LacZ (5) and histidinol dehydrogenase (3), the difference between the observed values and the values predicted by the Copper–Helmstetter model also appears to be relatively small (i.e. a small relative displacement), indicating that local sequences also have limited effect on the constitutive expression of non-fluorescent genes.

We did not specifically assess if there were any periodic changes in expression with distance from the origin that would be compatible with regular DNA compaction and coiling along the chromosome. However, the expression data clearly indicate that any periodicity that exists must have a relatively small effect on gene expression because the pattern observed deviated very little from that predicted solely by the gene copy number. That is, if DNA looping and coiling had a large effect on expression then the measured values would be expected to display large deviations from the predicted fits depending on whether the gene was situated in a position that is affected or not; and this was not observed in our study (Figure 1B and C) or in a previous study that placed LacZ at different chromosome positions (5).

Gene orientation had no demonstrable effect on gene expression despite our measurements being performed at the fastest growth rates in LB media when collisions between RNA polymerases and the replisomes should be most common. This indicates that collisions do not affect genes on the leading and lagging strands differently, in agreement with comparative genome analyses that suggest that gene orientation is associated with the functional class of a gene rather than its expression level (19).

The distance between the target gene and the transcription factor gene did not affect the amount of transcription factor (LacI) activity at the target gene. Although LacI is a common paradigm for studying diffusion of transcription factors, it's relatively high association rate constant compared with many other proteins, may mean that its facilitated diffusion rate and the time taken to bind its targets may not be representative (58). Furthermore we could not examine whether the target-TF distance provides an advantage in the dynamics when a transcription factor gene is first turned on or when the transcription factor's half-life is very short. With these limitations, the possibility that the target-TF distance may be functionally important for some proteins and under some conditions cannot be excluded. We stress that while we find no evidence that the distance between a target gene and its transcription factor gene affects transcription factor activity, we are not suggesting that the spatial

organization within the cell is not important for gene regulation; in fact, there is clear evidence that it is (59).

We found that most of the target gene's expression noise arises from fluctuations in the transcription factor (LacI) concentration. However, LacI fluctuations did not scale with the number of LacI molecules produced (i.e. the relationship between gene expression noise and mean expression did not alter with *lacI*'s position). Consequently the primary source of target gene expression noise is not the intrinsic fluctuations in LacI production but other extrinsic factors such as the transcription factors that regulate *lacI*. Therefore there appears to be a consistent pattern with the dominant source of expression noise for both *lacI* and *yfp* being due to extrinsic factors that regulate their expression. In comparison, chromosome position has minimal impact on gene expression noise. This finding differs from reports in yeast and mammalian cells and is most likely due to bacteria lacking enhancers, heterochromatin and other factors (16,60,61) that are associated with infrequent, stochastic switching between the active and silent expression states (16,17).

This study demonstrates that gene translocation is a potential mechanism for reprogramming the output of synthetic gene networks. Moreover the Cooper–Helmstetter and/or Hill-type functions can be used to reprogram gene regulation and networks in a predictable manner. Alternative methods for controlling expression levels through transcription factor regulation and the introduction of mutations at the promoter (62) and RBS (63) can produce a greater fold change in expression than was observed but they also have disadvantages. Tuning gene expression by altering transcription factor binding can alter gene expression noise as we have shown. Introducing mutations to vary the strength of the promoter or RBS may disrupt sites needed for transcription factor and sRNA regulation. These problems are potentially avoided by selecting different chromosome positions to vary expression because the integrity of the gene is maintained. Our observation that LacI expression can be empirically predicted from YFP expression at the same position (despite different promoters, ribosome binding sequences and coding sequences, and the absence of flanking terminators) indicates that chromosome position effects are largely sequence independent, and therefore broadly applicable. This is supported by earlier studies that have shown chromosome position effects with LacZ (5) and histidinol dehydrogenase (3) in *E. coli* and *Salmonella typhimurium*.

Determining the contribution of spatial information to signaling in gene networks is challenging but it is essential to understanding the evolution of genome organization and for choosing the optimal position of genes in engineered genomes and circuits. Comparative genome sequence analyses have identified multiple features of genome organization that are common and conserved. Several functional and non-functional (e.g. genetic linkage) hypotheses have been proposed for why these organizational features arise. Synthetic gene circuits are ideal for directly testing these hypotheses of function and causality. Furthermore synthetic gene circuits allow specific features of genome organization to be isolated and systematically manipulated thereby providing detailed information about the reaction steps and dose-response relationships that are necessary for mathematical modeling and predictive analyses. This level of detail and direct testing of function cannot generally be readily extracted from the analyses of microarrays and genome sequences. These advantages make synthetic genetic circuits a powerful adjunct to high-throughput and bioinformatics studies as we have shown here. Other common features of genome organization (e.g. gene colocalization) and gene arrangements (e.g. operons and overlapping genes) that are believed to be functionally important should also be characterized using synthetic gene circuits.

In conclusion, our study assessed how common features of genome organization may alter the regulation of translocated genes and consequently the output of gene networks. We identified features of genome organization that are likely to have an impact on gene regulation (i.e. chromosome position) and those that are generally not likely to have an impact (i.e. gene orientation and target-TF distance). Our data not only provide an understanding of the potential regulatory consequences of gene translocation but also yield insight into fundamental aspects of gene transcription, translation and intracellular signaling. The insights include: (i) the impact of gene copy number on mean expression as well as gene expression noise; (ii) the mechanisms by which local sequences alter gene expression; (iii) the minimal effect of diffusion distances on transcription factor activity; and (iv) the demonstration that collisions between the transcription machinery and replisomes do not modulate gene expression. These findings provide important constraints and bounds on the relative rates and magnitudes of different processes in gene regulation, which may be incorporated into general models of gene regulation. These results in *E. coli* are likely to be applicable to other closely related bacteria such as Salmonella (and perhaps more broadly) due to similarities in their genome organization and mechanisms of DNA replication and gene expression.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–6, Supplementary Figures 1–7, Supplementary Methods and Supplementary References [64–69].

ACKNOWLEDGEMENTS

We thank David Adamson for helpful discussions and comments on the manuscript. D.H.S.B., R.H. and L.W.L. made the strains and performed the measurements. H.N.L. designed the project, analyzed the data and wrote the paper.

FUNDING

The start-up funds and a COR Faculty Research Grant (FRG) from UC, Berkeley. Funding for open access charge: Berkeley Research Impact Initiative (BRII).

Conflict of interest statement. None declared.

REFERENCES

1. Hughes,D. (2000) Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. *Genome Biol.*, **1**, REVIEWS0006.
2. Shapiro,J.A. (2009) Letting Escherichia coli teach me about genome engineering. *Genetics*, **183**, 1205–1214.
3. Schmid,M.B. and Roth,J.R. (1987) Gene location affects expression level in Salmonella typhimurium. *J. Bacteriol.*, **169**, 2872–2875.
4. Chandler,M.G. and Pritchard,R.H. (1975) The effect of gene concentration and relative gene dosage on gene output in Escherichia coli. *Mol. Gen. Genet.*, **138**, 127–141.
5. Sousa,C., de Lorenzo,V. and Cebolla,A. (1997) Modulation of gene expression through chromosomal positioning in Escherichia coli. *Microbiology*, **143(Pt 6)**, 2071–2078.
6. Cooper,S. and Helmstetter,C.E. (1968) Chromosome replication and the division cycle of Escherichia coli B/r. *J. Mol. Biol.*, **31**, 519–540.
7. Jeong,K.S., Ahn,J. and Khodursky,A.B. (2004) Spatial patterns of transcriptional activity in the chromosome of Escherichia coli. *Genome Biol.*, **5**, R86.
8. Kepes,F. (2004) Periodic transcriptional organization of the E.coli genome. *J. Mol. Biol.*, **340**, 957–964.
9. Peter,B.J., Arsuaga,J., Breier,A.M., Khodursky,A.B., Brown,P.O. and Cozzarelli,N.R. (2004) Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. *Genome Biol.*, **5**, R87.
10. Allen,T.E., Price,N.D., Joyce,A.R. and Palsson,B.O. (2006) Long-range periodic patterns in microbial genomes indicate significant multi-scale chromosomal organization. *PLoS Comput. Biol.*, **2**, e2.
11. Brand,A.H., Breeden,L., Abraham,J., Sternglanz,R. and Nasmyth,K. (1985) Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell*, **41**, 41–48.
12. Nasmyth,K.A., Tatchell,K., Hall,B.D., Astell,C. and Smith,M. (1981) A position effect in the control of transcription at yeast mating type loci. *Nature*, **289**, 244–250.
13. Ferguson,B.M. and Fangman,W.L. (1992) A position effect on the time of replication origin activation in yeast. *Cell*, **68**, 333–339.
14. Jaenisch,R., Jahner,D., Nobis,P., Simon,I., Lohler,J., Harbers,K. and Grotkopp,D. (1981) Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. *Cell*, **24**, 519–529.
15. al-Shawi,R., Kinnaird,J., Burke,J. and Bishop,J.O. (1990) Expression of a foreign gene in a line of transgenic mice is modulated by a chromosomal position effect. *Mol. Cell. Biol.*, **10**, 1192–1998.
16. Batenchuk,C., St-Pierre,S., Tepliakova,L., Adiga,S., Szuto,A., Kabbani,N., Bell,J.C., Baetz,K. and Kaern,M. (2011) Chromosomal position effects are linked to sir2-mediated variation in transcriptional burst size. *Biophys. J.*, **100**, L56–L58.
17. Skupsky,R., Burnett,J.C., Foley,J.E., Schaffer,D.V. and Arkin,A.P. (2010) HIV promoter integration site primarily modulates transcriptional burst size rather than frequency. *PLoS Comput. Biol.*, **6**, e1000952.
18. Becskei,A., Kaufmann,B.B. and van Oudenaarden,A. (2005) Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat. Genet.*, **37**, 937–944.
19. Rocha,E.P. (2008) The organization of the bacterial genome. *Annu. Rev. Genet.*, **42**, 211–233.
20. Price,M.N., Alm,E.J. and Arkin,A.P. (2005) Interruptions in gene expression drive highly expressed operons to the leading strand of DNA replication. *Nucleic Acids Res.*, **33**, 3224–3234.
21. Washburn,R.S. and Gottesman,M.E. (2011) Transcription termination maintains chromosome integrity. *Proc. Natl Acad. Sci. USA*, **108**, 792–797.
22. Janga,S.C., Salgado,H. and Martinez-Antonio,A. (2009) Transcriptional regulation shapes the organization of genes on bacterial chromosomes. *Nucleic Acids Res.*, **37**, 3680–3688.
23. Kolesov,G., Wunderlich,Z., Laikova,O.N., Gelfand,M.S. and Mirny,L.A. (2007) How gene order is influenced by the biophysics of transcription regulation. *Proc. Natl Acad. Sci. USA*, **104**, 13948–13953.
24. Fritsche,M., Li,S., Heermann,D.W. and Wiggins,P.A. (2012) A model for Escherichia coli chromosome packaging supports transcription factor-induced DNA domain formation. *Nucleic Acids Res.*, **40**, 972–980.
25. Svetic,R.E., MacCluer,C.R., Buckley,C.O., Smythe,K.L. and Jackson,J.H. (2004) A metabolic force for gene clustering. *Bull. Math. Biol.*, **66**, 559–581.
26. Guido,N.J., Wang,X., Adalsteinsson,D., McMillen,D., Hasty,J., Cantor,C.R., Elston,T.C. and Collins,J.J. (2006) A bottom-up approach to gene regulation. *Nature*, **439**, 856–860.
27. Koide,T., Pang,W.L. and Baliga,N.S. (2009) The role of predictive modelling in rationally re-engineering biological systems. *Nat. Rev. Microbiol.*, **7**, 297–305.
28. O'Shaughnessy,E.C., Palani,S., Collins,J.J. and Sarkar,C.A. (2011) Tunable signal processing in synthetic MAP kinase cascades. *Cell*, **144**, 119–131.
29. Gardner,T.S., Cantor,C.R. and Collins,J.J. (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature*, **403**, 339–342.
30. Price,M.N., Arkin,A.P. and Alm,E.J. (2006) The life-cycle of operons. *PLoS Genet.*, **2**, e96.
31. Lutz,R. and Bujard,H. (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.*, **25**, 1203–1210.
32. Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl Acad. Sci. USA*, **97**, 6640–6645.
33. Lim,H.N., Lee,Y. and Hussein,R. (2011) Fundamental relationship between operon organization and gene expression. *Proc. Natl Acad. Sci. USA*, **108**, 10626–10631.
34. Taylor,J.R. (1997). An introduction to error analysis: the study of uncertainties in physical measurements. 2nd ed. University Science Books, Sausalito, pp. 327.
35. O'Gorman,R.B., Rosenberg,J.M., Kallai,O.B., Dickerson,R.E., Itakura,K., Riggs,A.D. and Matthews,K.S. (1980) Equilibrium binding of inducer to lac repressor.operator DNA complex. *J. Biol. Chem.*, **255**, 10107–10114.
36. Elowitz,M.B., Levine,A.J., Siggia,E.D. and Swain,P.S. (2002) Stochastic gene expression in a single cell. *Science*, **297**, 1183–1186.
37. Bremer,H. and Churchward,G. (1977) An examination of the Cooper-Helmstetter theory of DNA replication in bacteria and its underlying assumptions. *J. Theor. Biol.*, **69**, 645–654.
38. Chandler,M., Bird,R.E. and Caro,L. (1975) The replication time of the Escherichia coli K12 chromosome as a function of cell doubling time. *J. Mol. Biol.*, **94**, 127–132.
39. Kubitschek,H.E. and Freedman,M.L. (1971) Chromosome replication and the division cycle of Escherichia coli B-r. *J. Bacteriol.*, **107**, 95–99.
40. Elowitz,M.B. and Leibler,S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature*, **403**, 335–338.
41. Pedraza,J.M. and van Oudenaarden,A. (2005) Noise propagation in gene networks. *Science*, **307**, 1965–1969.
42. Crampton,N., Bonass,W.A., Kirkham,J., Rivetti,C. and Thomson,N.H. (2006) Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Res.*, **34**, 5416–5425.
43. Murphy,K.F., Balazsi,G. and Collins,J.J. (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl Acad. Sci. USA*, **104**, 12726–12731.
44. Nevozhay,D., Adams,R.M., Murphy,K.F., Josic,K. and Balazsi,G. (2009) Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proc. Natl Acad. Sci. USA*, **106**, 5123–5128.
45. Laiken,S.L., Gross,C.A. and Von Hippel,P.H. (1972) Equilibrium and kinetic studies of Escherichia coli lac repressor-inducer interactions. *J. Mol. Biol.*, **66**, 143–155.

46. Ohshima,Y., Mizokoshi,T. and Horiuchi,T. (1974) Binding of an inducer to the lac repressor. *J. Mol. Biol.*, **89**, 127–136.
47. Oehler,S., Alberti,S. and Muller-Hill,B. (2006) Induction of the lac promoter in the absence of DNA loops and the stoichiometry of induction. *Nucleic Acids Res.*, **34**, 606–612.
48. Bohn,C. and Bouloc,P. (1998) The Escherichia coli cmlA gene encodes the multidrug efflux pump Cmr/MdfA and is responsible for isopropyl-beta-D-thiogalactopyranoside exclusion and spectinomycin sensitivity. *J. Bacteriol.*, **180**, 6072–6075.
49. Hansen,L.H., Knudsen,S. and Sorensen,S.J. (1998) The effect of the lacY gene on the induction of IPTG inducible promoters, studied in Escherichia coli and Pseudomonas fluorescens. *Curr. Microbiol.*, **36**, 341–347.
50. Smits,W.K., Kuipers,O.P. and Veening,J.W. (2006) Phenotypic variation in bacteria: the role of feedback regulation. *Nat. Rev. Microbiol.*, **4**, 259–271.
51. Montero Llopis,P., Jackson,A.F., Sliusarenko,O., Surovtsev,I., Heinritz,J., Emonet,T. and Jacobs-Wagner,C. (2010) Spatial organization of the flow of genetic information in bacteria. *Nature*, **466**, 77–81.
52. Nevo-Dinur,K., Nussbaum-Shochat,A., Ben-Yehuda,S. and Amster-Choder,O. Translation-independent localization of mRNA in E. coli. *Science*, **331**, 1081–1084.
53. Maheshri,N. and O'Shea,E.K. (2007) Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.*, **36**, 413–434.
54. Choi,P.J., Cai,L., Frieda,K. and Xie,X.S. (2008) A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, **322**, 442–446.
55. Paulsson,J. (2005) Models of stochastic gene expression. *Phys. Life Rev.*, **2**, 157–-175.
56. Murphy,K.F., Adams,R.M., Wang,X., Balazsi,G. and Collins,J.J. (2010) Tuning and controlling gene expression noise in synthetic gene networks. *Nucleic Acids Res.*, **38**, 2712–2726.
57. De,S. and Babu,M.M. (2010) Genomic neighbourhood and the regulation of gene expression. *Curr. Opin. Cell Biol.*, **22**, 326–333.
58. Halford,S.E. and Marko,J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, **32**, 3040–3052.
59. Nevo-Dinur,K., Nussbaum-Shochat,A., Ben-Yehuda,S. and Amster-Choder,O. (2011) Translation-independent localization of mRNA in E. coli. *Science*, **331**, 1081–1084.
60. Sen,R. and Grosschedl,R. (2010) Memories of lost enhancers. *Genes Dev.*, **24**, 973–979.
61. Perez-Martin,J., Uria,J.A. and Johnson,A.D. (1999) Phenotypic switching in Candida albicans is controlled by a SIR2 gene. *EMBO J.*, **18**, 2580–2592.
62. Alper,H., Fischer,C., Nevoigt,E. and Stephanopoulos,G. (2005) Tuning genetic control through promoter engineering. *Proc. Natl Acad. Sci. USA*, **102**, 12678–12683.
63. Salis,H.M., Mirsky,E.A. and Voigt,C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950.
64. Alon,U. (2007) *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman and Hall/CRC, Boca Raton, FL.
65. Bell,C.E. and Lewis,M. (2000) A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Biol.*, **7**, 209–214.
66. Butler,A.P., Revzin,A. and von Hippel,P.H. (1977) Molecular parameters characterizing the interaction of Escherichia coli lac repressor with non-operator DNA and inducer. *Biochemistry*, **16**, 4757–4768.
67. Barkley,M.D., Riggs,A.D., Jobe,A. and Burgeois,S. (1975) Interaction of effecting ligands with lac repressor and repressor-operator complex. *Biochemistry*, **14**, 1700–1712.
68. Hussein,R. and Lim,H.N. (2012) Direct comparison of small RNA and transcription factor signaling. *Nucleic Acids Res*, **40**, 7269–7279.
69. Hussein,R. and Lim,H.N. (2011) Disruption of small RNA signaling caused by competition for Hfq. *Proc. Natl Acad. Sci. USA*, **108**, 1110–1115.