Genetic context effects can override canonical *cis* regulatory elements in *Escherichia coli*

Scott A. Scholz[1,*], Chase D. Lindeboom[1] and Peter L. Freddolino [1,2,*]

[1]Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, MI, USA and [2]Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA

ABSTRACT

Recent experiments have shown that in addition to control by *cis* regulatory elements, the local chromosomal context of a gene also has a profound impact on its transcription. Although this chromosome-position dependent expression variation has been empirically mapped at high-resolution, the underlying causes of the variation have not been elucidated. Here, we demonstrate that 1 kb of flanking, non-coding synthetic sequences with a low frequency of guanosine and cytosine (GC) can dramatically reduce reporter expression compared to neutral and high GC-content flanks in *Escherichia coli*. Natural and artificial genetic context can have a similarly strong effect on reporter expression, regardless of cell growth phase or medium. Despite the strong reduction in the maximal expression level from the fully-induced reporter, low GC synthetic flanks do not affect the time required to reach the maximal expression level after induction. Overall, we demonstrate key determinants of transcriptional propensity that appear to act as tunable modulators of transcription, independent of regulatory sequences such as the promoter. These findings provide insight into the regulation of naturally occurring genes and an independent control for optimizing expression of synthetic biology constructs.

GRAPHICAL ABSTRACT
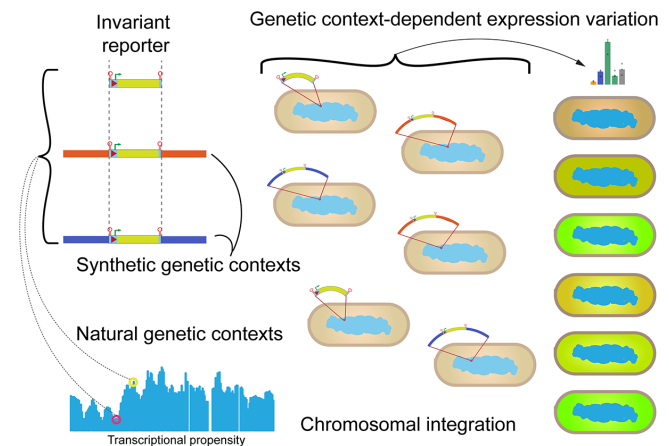
INTRODUCTION

Unicellular microbes, particularly fast-growing bacteria, typically have dense genomes with little intergenic DNA. Whereas variation in the transcribability of chromosomal regions in many eukaryotes is correlated with gene density, transcription from the *Escherichia coli* chromosome can vary many fold across regions of similarly high gene density (1). Several studies have reported variation in expression from a standardized reporter integrated into different locations in the *E. coli* genome (2–6). Factors that affect the bacterial chromosome, some of which may also contribute to position-dependent expression variation, have been recently reviewed (7,8). More recent studies have also examined the effects of metabolic burden and nutrient source on position-dependent expression variation (9). The variation in gene expression at different genomic loci in bacteria has been used to great effect in the field of synthetic biology to generate bacterial strains with improved production and circuit performance (10–15).

We recently generated an empirical, high-resolution map of transcription from a standardized reporter integrated at >144,000 sites in the *E. coli* genome. We termed the

*To whom correspondence should be addressed. Tel: +1 734 647 5839; Email: petefred@umich.edu
Correspondence may also be addressed to Scott A. Scholz. Tel: +49 6421178433; Email: scott.scholz@mpi-marburg.mpg.de
Present address: Scott A. Scholz, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany.

genomic position-dependent effects on reporter transcription 'transcriptional propensity' ([1]). The chromosomal features that were observed to most strongly correlate with transcriptional propensity were local GC content, proximity to ribosomal RNA operons, and binding of the nucleoid associated proteins (NAPs) H-NS and Fis. For example, H-NS binding sites are apparent as troughs on the transcriptional propensity map (i.e. low transcriptional propensity). Subsequent high-throughput assessment of fluorescent protein expression from transposon integrations largely confirmed these results, and also linked H-NS binding sites to low reporter expression ([16]). H-NS binds to and reduces expression of genes with AT rich sequences ([17–20]). H-NS alone and in combination with other factors has been shown to function to drive RNAP pausing after transcription initiation ([21]). In our previous work, we observed that reporters that integrated into H-NS binding regions also became silenced, despite the fact that the GC-neutral reporter sequence was unaltered.

At the other extreme, the most readily apparent features of transcriptional propensity topology are broad peaks of high transcriptional propensity that are centered on ribosomal RNA operons (*rrn*; see Supplementary Figure S1A). *rrn* are the most heavily transcribed genes in the *E. coli* genome ([22,23]), and additional highly-transcribed genes (including elements of the amino acid biosynthetic pathways) are enriched in regions of high transcriptional propensity around *rrn* ([1]). A model in which *rrn* colocalize into nucleolus-like domains ([24,25]) and thereby increase RNAP access to the surrounding genes is consistent with this data and may explain the transcriptional propensity peaks at *rrn*, but requires further experimental testing.

Context-dependent expression variation is also caused by interference between transcription of neighboring genes, as has been demonstrated through rational reporter integration and construct design ([3,26–29]). In general, terminator-flanked genes downstream of highly expressed promoters are depressed in their expression, consistent with a build-up of positive supercoiling. Supercoiling interference effects are also apparent for genes encoded in the divergent orientation ([26,30]).

In this study, we investigated the expression consequences of chromosomal genetic context in *E. coli* that is further from the transcriptional start site (TSS) than is typically considered for gene regulation and synthetic biological designs. With an invariant core reporter sequence, including standardized sequences 162 bp upstream and 852 bp downstream of the TSS, and strong flanking transcriptional terminators, the reporter could nonetheless range from high expression to undetectable levels depending upon its location of insertion, and the flanking genomic context. We found that a reporter is silenced if it is flanked by 1 kb of non-coding low GC content DNA, while high GC flanks can cause de-repression in a context that would otherwise be silencing. Natural and synthetic genetic contexts can have strong effects on reporter expression independent of growth phase and under several different growth conditions. Despite the substantial differences in reporter expression depending on the chromosomal integration site and the GC content of flanking DNA, the time by which the reporter reaches maximal fluorescence level after induction is similar. We also demonstrate that although *rrn* are at the center of large transcriptional propensity peaks, removal of the *rrn* itself does not diminish the expression of a reporter that has integrated in close proximity to it. Finally, the transcriptionally insulated reporter used in this study was affected by interference from the highly expressed *rrnE* operon, depending on relative gene orientation, despite an absence of direct transcriptional readthrough. This observation is a confirmation of previously reported supercoiling competition effects. Genetic context effects, reported throughout this article at the population and single-cell level, appear to cause tunable and uniform changes in expression across a clonal population. In other words, differences in the average fluorescence between strains are revealed via flow-cytometry as uniform shifts of the entire population distribution around different medians, and not due to the appearance of different sub-populations shifting the population average fluorescence.

MATERIALS AND METHODS

Reagents

Enzymes/kits. DNAse digestion for RNA purification was performed using the Qiagen RNase Free DNase kit (product # 79254, Qiagen, Germantown, MD). Reverse transcription for RT-qPCR was performed using NEB Protoscript II Reverse Transcriptase (product # M0368, New England Biolabs, Ipswich, MA). qPCR then used the Bio-Rad iTaq Universal SYBR Green Supermix (product # 1725120, BioRad, Hercules, CA, USA).

Equipment. All microplate-based growth and fluorescence experiments were performed using a BioTek Synergy H1 multi-mode plate reader (BioTek, Winooski, VT); growth-only experiments during pregrowth were instead performed on a BioTek Epoch plate reader. Flow cytometry was performed using a MACSQuant VYB instrument (Miltenyi Biotech, Auburn, CA). Quantitative PCR experiments were performed using a BioRad CFX96 instrument (BioRad, Hercules, CA).

Biological resources

All strains were constructed from ecSAS17, a derivative of *E. coli* K12 MG1655 with chromosomal integrations of constitutively expressed tet repressor and mCherry ([1]). The mNeonGreen reporter construct was chromosomally integrated into different genomic locations with or without synthetic genetic contexts, but otherwise unaltered from our previous study. For strain construction, ecSAS17 harboring the lambda Red plasmid pSIM5 (gift from Prof. Don Court) was transformed with linear PCR fragments of the reporter or reporter flanked by synthetic GC content (sequences of the plasmids used for strain construction are given in Supplementary Data S1, and a full listing of strains used in the present study is given in Supplementary Data S2). The reporter kanamycin resistance marker was then removed from

each strain with the helper plasmid pCP20. Targeted integration of the reporter construct, with or without flanking synthetic genetic contexts as indicated, was confirmed by PCR genotyping across integration junctions and sanger sequencing of junction PCR products (a complete listing of primers is given in Supplementary Data S3). For all strain construction, LB Lennox broth (10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl) with the appropriate antibiotics added was used for propagation and growth. Confirmed strains were cryopreserved indefinitely at $-80°C$ in 15% glycerol.

Statistical analyses

Population-level fluorescence per OD values are presented in several bar-plots in this article. Individual replicate values represent the fluorescence divided by OD, averaged over all time points in which OD450 was between 0.5 and 1 (after normalization to a 10 mm path length), which is within the log growth phase for all strains reported here. The included bar is a summary statistic of the replicate data; either geometric mean or median, as indicated. Specifically, values collected from microplate reader experiments were processed and reported herein as blank (media-only) subtracted fluorescence and OD 450 nm. Fluorescence and OD values were multiplied by 3.33 simply to match the path length for a 10 mm standard cuvette. Fluorescence per OD is reported in bar charts throughout this work (except Figure 3) as the average Fluor/OD for the multiple time points that were collected between OD 450 nm 0.5 and 1. Cells grown in M9RDM are in the exponential growth phase in this OD range and also have a similar Fluor/OD at each time point across this OD range. Finally, the Fluor/OD value for each strain within the individual replicates was divided by the total Fluor/OD average of all strains within that replicate in order to normalize for small replicate-specific technical differences in medium and microplate reader assessment. For growth rate analysis, blank subtracted OD values were $\log_2$ transformed. Then, over the range of the $\log_2$ OD values that were linear with respect to time, we performed a linear regression to determine the slope (aka Doubling Time, as reported in Supplementary Figure S7). Somewhat different ranges were appropriate for growth rate and Fluor/OD calculations for cells grown in different media. For Figure 3 and Supplementary Table S1, data from the following specific OD ranges were plotted and some time restrictions were also implemented in order to avoid obscured OD signals from bubbles at early time points, as are apparent in the complete data over time (Supplementary Data S4): M9RDM: 0.06–1 OD after 80 minutes, M9 Casamino Acids: 0.06–0.5 OD after 330 min, M9 Glucose: 0.06–0.25 OD after 80 min, M9 Succinate 0.06–0.125 after 80 min, M9 Acetate 0.06–0.125 OD after 600 min.

The autofluorescence per OD of the reporter-free base strain ecSAS17 calculated over the same OD range (grown in parallel within the same microplate) was subtracted from the fluorescence per OD value from each strain to obtain the plotted data. Analysis of other data types (flow cytometry, RT-qPCR, etc.) are described in the relevant method-specific sections below.

Strain growth for microplate reader and flow cytometry strain assessment

Each strain was independently streaked onto an LB Lennox agar plate from a cryopreserved stock and incubated overnight at 37°C. Plates were stored for up to 12 days at 4°C. A single colony from the LB agar plate was inoculated into 3 ml of LB Lennox and grown at 37°C overnight with shaking. 10 μl of the overnight culture was inoculated into 1 ml M9RDM (Glucose 4 g/l, NH$_4$Cl 1 g/l, KH$_2$PO$_4$ 3 g/l, NaCl 0.5 g/l, Na$_2$HPO$_4$ 6 g/l, MgSO$_4$ 240.7 mg/l, ferric citrate 2.45 mg/l, CaCl$_2$ 0.111 mg/l, 200 ml/l 5× Supplement EZ, 100 ml/l 10× ACGU (Teknova cat # M2103), 1 ml/l MOPS micronutrient solution (31)) + a final concentration of 100 μg/l anhydrotetracycline (aTc) and incubated at 37°C for 2 h with shaking. 1 μl of the log-phase cells was inoculated into 150 μl of M9RDM + aTc in a microplate (96-well flat clear bottom, black sides or flat bottom all clear for microplate reader and flow cytometry, respectively) and topped with 100 μl mineral oil. For microplate reader assessment, mNeonGreen fluorescence (Ex. 506 nm, Em. 550 nm), mCherry fluorescence (Ex. 580 nm, Em. 610 nm) and optical density (OD) at 450 nm were measured every 10 minutes. For growth before flow cytometry assessment OD 450 was measured every 10 minutes. We note here that OD at 450 nm maximizes the sensitivity of our OD measurements at low cell density due to stronger scattering of lower wavelength light. All strains were otherwise grown at 37°C with orbital shaking in a Synergy H1 or Epoch BioTek plate reader. For flow cytometry assessment, the microplate was rapidly moved from the microplate reader into an ice water bath after 3 h of growth and immediately transferred for auto-sampled flow cytometry.

Unless otherwise noted, plate reader experiments were performed entirely using the M9RDM recipe given above. For experiments using other nutrient sources, we instead used as a base the following M9 minimal media base: NH$_4$Cl 1 g/l, KH$_2$PO$_4$ 3 g/l, NaCl 0.5 g/l, Na$_2$HPO$_4$ 6 g/l, MgSO$_4$ 240.7 mg/l, CaCl$_2$ 11.1 mg/l, ferric citrate 2.45 mg/l, plus 1× MOPS micronutrients. Carbon sources were used as indicated in the text, at the following concentrations: glucose 0.2% (w/v), casamino acids 0.4% (w/v), sodium succinate 0.2% (w/v), sodium acetate 0.2% (w/v).

Flow cytometry and single-cell analysis

Rapidly cooled cells grown to log phase were sampled from an ice-cold 96-well plate on a Miltenyi Biotec MACSQuant YVB flow cytometer cooling auto-sampler. After daily calibration with MACSQuant Calibration Beads, forward scatter (FSC), side scatter (SSC), mCherry (Ex. 561 nm, Em. band pass 615/20 nm) and mNeonGreen (Ex. 488 nm, Em. band pass 525/50) signals were collected with an automated sampling and washing procedure. We calculated the correlation between the mNG signal and the signals for FSC, SSC and mCherry, as well as the slope of the linear regression of log-transformed values with respect to log-transformed mNG within each strain. Since the correlation between mNG and SSC was the strongest, we corrected log mNG for each strain by subtracting the product of the average slope of log mNG with respect to log SSC for each cell in order to calculate mNG concentration. This correction tightens the

distribution of mNG signal within each strain by normalizing the natural variation attributable to cell morphology.

Induction timing of strains passaged in a microplate reader

Since the autofluorescence per OD from reporter-free cells growing in M9RDM medium changes at different ODs, we performed a kernel regression of the autofluorescence signal of ecSAS17 cells growing in M9RDM with respect to the optical density at 450 nm (OD 450) in order to calculate the fluorescence signal attributable to the growth medium and cells at every OD. As the reporter-free cells grow in M9RDM, the fluorescence signal initially dips slightly but substantially (presumably due to consumption of a medium component contributing to autofluorescence signal, as this decrease was not observed in cell-free medium) before increasing as a result of autofluorescence of the cells. Therefore, we subtracted the expected autofluorescence signal from cells growing in M9RDM derived from the kernel regression from the fluorescence signal from each strain at every measured OD value and finally divided by OD in order to report only the fluorescence per OD resulting from mNG reporter expression itself in Figure 4B. The fold-difference between fluorescence per OD and maximal induction level was calculated by averaging the fluorescence per OD between 410 and 530 min and reporting the $\log_2$ fold-change of the fluorescence per OD at each timepoint from that maximal induction value of each strain.

qPCR based reporter measurements

The calibration of reporter expression by RT-qPCR shown in Supplementary Figure S2 was performed as described in (1). In brief, cells were grown to OD 450 0.2 in M9RDM, at which point 650 μl of the culture was mixed with 1.3 ml RNAProtect Bacterial Reagent (Qiagen) and frozen according to the manufacturer's instructions. RNA was then purified using the RNeasy mini kit (Qiagen). An additional DNaseI treatment (Qiagen RNase-Free DNase Set) was performed on the purified sample in order to remove residual gDNA and re-purified using the RNeasy kit. cDNA was finally synthesized using the NEB Protoscript II First Strand cDNA synthesis kit (New England Biolabs) following the manufacturer's instructions in parallel to control samples without the Reverse Transcriptase enzyme (-RT).

qPCR reactions on the cDNA and -RT control samples described above were performed with primers 63–64 for *mNG* (length = 100 bp) and *mdoG* primers 65–66 (length = 65 bp) (see Supplementary Data S3) using Bio-Rad iTaq Universal SYBR Green Supermix, on a BioRad CFX96 instrument. 2 μl of 1:5 cDNA reactions diluted in water were then directly added to 20 μl final-volume qPCR reactions. Thermal cycling parameters were as follows:

1. 95°C, 2 min (initial denaturation)
2. 40 cycles of:
   a. 95°C, 10 s (denaturation)
   b. 55°C, 15 s (annealing/extension)
   c. 68°C, 15 s (additional extension)
3. Melt curve (65°C to 95°C, incremented by 0.5°C every 5 s)

Plate reads were taken at the end of step 'b' during the amplification cycle (step 2 above), and at every temperature increment during the melt curve (step 3 above). Except for the WT no-reporter control (ecSAS17 strain, 30 Ct, thus representing background non-specific amplification under our conditions), samples containing cDNA had 13–18 Ct compared to 30–34 Ct for no-template and -RT controls for both primer sets. The melting temperature for the mNG-specific products (80.5°C) and mdoG-specific products (76.5°C) was virtually identical for standard curve DNA samples and experimental samples, except for no-template water controls and the ecSAS17 WT control DNA with mNG primers; equivalent results were observed for each of three independent replicates conducted on separate days. To generate standard curves, gDNA extracted from ecSAS20 with a Qiagen Blood and Tissue kit was serially diluted from 40 ng/μl to 1 ng/μl. The $\log_2$ of known DNA concentrations compared to Ct slope ranged from 1 to 1.1 (R^2 ranging from 0.98–0.99), indicating a qPCR efficiency close to 100% in all cases. The relative target RNA level was quantified by calculating the cDNA concentration from the standard curve generated for each primer set, which accounts for small variations in qPCR efficiency. Finally, the mNG RNA signal was divided by the *mdoG* RNA signal for each sample as a sample loading normalization and used directly for Supplementary Figure S2.

RESULTS

Genome context has similar effects on reporters with different promoter strengths

To study the impact of genetic context on gene expression, we measured expression from a fluorescent reporter construct integrated at several genomic loci in the *E. coli* genome with manipulated local genetic contexts. Our invariant reporter is schematized in Figure 1A, and includes an inducible promoter, mNeonGreen coding sequence, and highly efficient flanking bidirectional terminators to provide transcriptional insulation (1,32). As detailed in Supplementary Text S1, we found that translated protein levels in our reporter serve as a faithful readout of transcriptional propensity across a wide range of expression levels, both in bulk (plate reader) and at a single cell level (flow cytometry) (Supplementary Figure S1). Thus, for the sake of convenience, we made use of fluorescent readouts of protein production as a proxy for the context-dependent transcriptional changes in reporter expression in the work described here. An important corollary of the data shown in Supplementary Text S1 is that we see no evidence for any position-dependent 'translational propensity' in *E. coli* (at least for the limited set of sites considered here). If the translation of an mRNA depends upon its location on the chromosome, we would expect to see position-dependent differences in the protein:RNA ratio; instead, fluorescent protein levels are well correlated both with transcriptional propensity and RNA levels (Supplementary Figure S2). Unless otherwise noted, all experiments were conducted in a rich defined (M9RDM) medium, which is also the growth medium used to generate the transcriptional propensity map, referenced throughout this work (1).
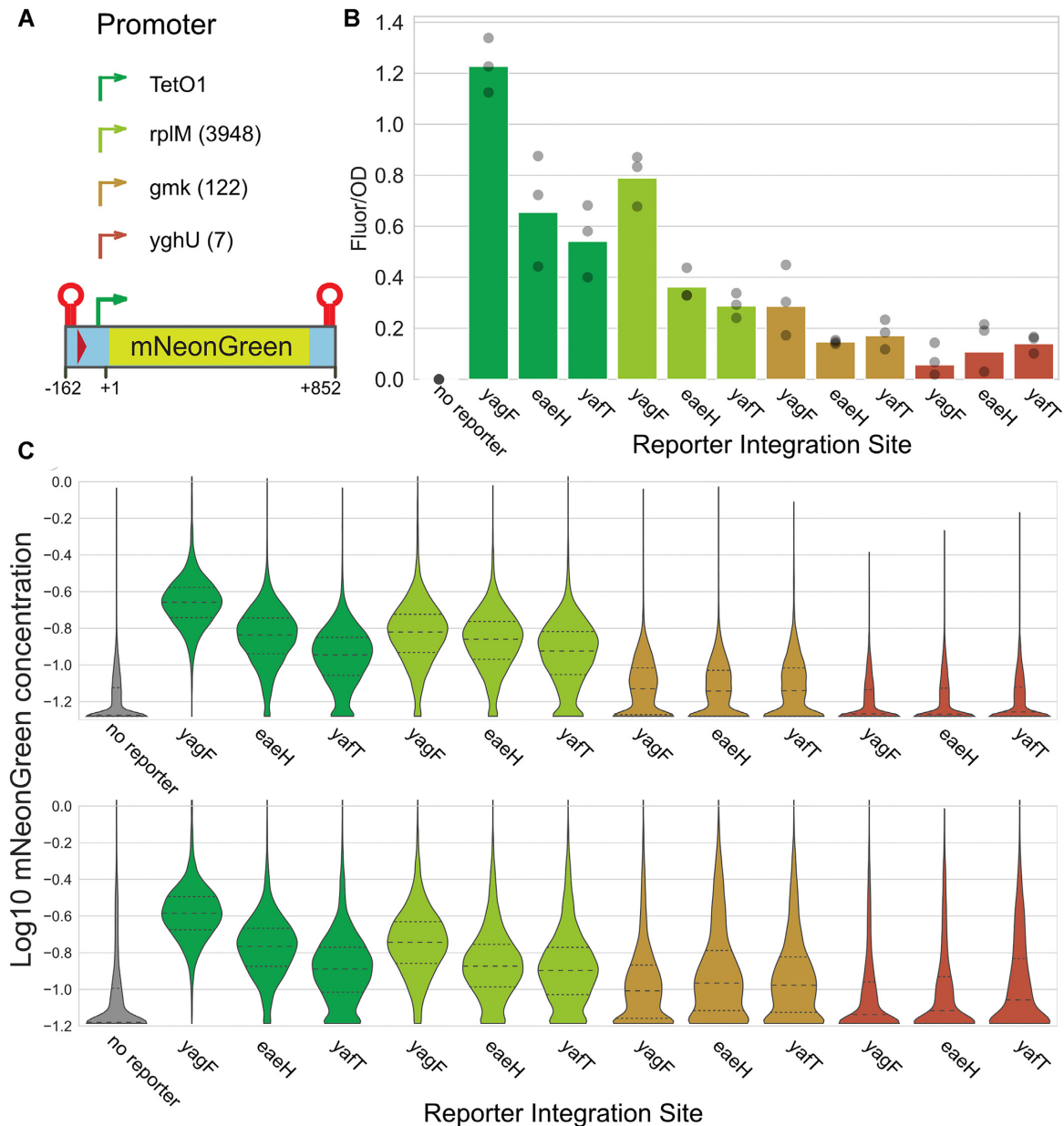
Figure 1. Genetic context affects gene expression from strong and weak promoters. (**A**) Diagram of the reporter after removal of an FRT-flanked kanamycin resistance cassette (The red triangle represents the remaining FRT scar). The mNeonGreen CDS is downstream of a strong RBS and different strength promoters flanked by very strong transcriptional terminators (red hairpins). The promoter strength is indicated by the transcripts per million normalized RNA count from the native gene in parentheses (64). Coordinates for the reporter construct diagram are relative to the transcription start site of the TetO1 promoter. (**B**) Microplate reader normalized Fluorescence/OD for mNeonGreen reporters with four different promoters integrated into three adjacent genomic loci (see Supplementary Figure S1A). Bars show the geometric mean signal across three biological replicates, with individual replicates shown as gray circles. (**C**) Two replicate flow cytometry readings of the same strains measured on two different days show that the distribution of mNeonGreen concentration per cell is consistent with microplate data. All cells with fluorescence below the median for the 'no reporter' cells are clamped at a lower limit equal to the 'no reporter' median, representing the limit of detection in the experiment; the apparent accumulation of cells at the bottom of the axis this does not represent a secondary definable population, but rather, simply the presence of a portion of cells that do not have detectable fluorescence. The dashed line shows the median, and dotted lines 25th/75th percentile.

As the targeted reporter integrations using the TetO1 promoter (33) from our prior work are generally expressed as expected based on the transcriptional propensity map (1) (Supplementary Figure S1B, C; see also Supplementary Text S1), we sought to understand whether different promoters driving reporter expression would be similarly affected by genetic context during log-phase growth in rich medium (~32 min doubling time; Supplementary Figure S7). The mNeonGreen reporter (mNG), and variants thereof, meet criteria chosen to faithfully read out the effect of genetic context without drastically perturbing the chromosome (unless by design), and to be unaffected by transcriptional read-through from strong chromosomal promoters. To that end, the reporter is relatively small (1014 bp), flanked by strong transcriptional terminators with undetectable transcriptional read-through, and has neutral intrinsic GC content (54% GC) (Figure 1A). To isolate the effect of genetic context on a gene with different promoter strengths, we integrated reporters with four promoter variants into three genomic loci that were selected because they are proximal to each other (maximum distance is 76 kb or <2% of total genome length apart, to minimize dosage effects), but have very different transcriptional propensity (*yagF*, *eaeH* and *yafT*; see Supplementary Figure S1). Throughout this work, the name assigned to an insertion location refers to the closest gene with a coordinate number preceding reporter integrations. We observed via bulk fluorescence experiments that mNG expression driven by the same TetO1 promoter used to generate the transcriptional propensity map ranked as expected. mNG expression driven by different promoters was affected by genetic context in the same rank order (*yagF* > *eaeH* > *yafT*), unless very weak promoters are used, in which case the distributions become essentially indistinguishable (Figure 1A-B). The strains with the weakest promoter that we tested still produced a fluorescent signal above the reporter-free base strain autofluorescence level, but did not change based on genetic context, indicating there may be a baseline level of transcriptional activity that cannot be further silenced by local genetic context. We observed a similar rank ordering in flow cytometry experiments for the TetO1 and *rplM* promoters, although in this case the distinctions between locations for the weaker *gmk* and *yghU* promoters were not apparent (Figure 1C). Flow cytometry may lack the sensitivity to reliably detect such differences at the extreme low end of expression level, especially given that a large fraction of the cells with the *gmk* and *yghU* promoters had fluorescence levels below the average for reporter-free cells (Figure 1C). Importantly, the flow cytometry results demonstrate that the contexts we have tested here only shift the fluorescence of cells as a unimodal population, with no instances of the appearance of a secondary population; thus, the effects of chromosomal context represent tunable, rather than all-or-nothing, changes in expression.

Synthetic genetic context can strongly influence reporter expression

Transcriptional propensity is strongly correlated with genomic features such as local GC content, binding of the nucleoid associated protein H-NS, and proximity to ribo-

somal RNA operons (1). We set out to directly test whether systemic alteration of the most informative of these genetic features could affect expression of an integrated reporter. In all cases, our standard reporter with a TetO1 promoter was used (and as shown in Figure 1A, the sequence context was fixed and constant across all integrations for 162 bp upstream of the transcription start site). An identical reporter construct was integrated into two sites, one with naturally high (*nfi*) and one with naturally low (*wbbH*) transcriptional propensity. In order to isolate the effect of flanking GC content on reporter expression, we then integrated the same reporter into both sites and included 1 kb of flanking synthetic context on each side with either 35% average GC content or 65% average GC content. These GC contents represent extremes compared to the average *E. coli* GC content of 50.4% (Figure 2A). We amplified two independent sequences for both our 35% and 65% GC synthetic flanks from eukaryotic yeast or human DNA ('35% GC' A/B and '65% GC' A/B respectively), which are not expected to bear specific regulatory elements that are functional in *E. coli*. All sequences are given in Supplementary Data S1.

Fluorescence from the reporter expressed from within low GC content synthetic contexts was a fraction of the reporter with no synthetic flanks, regardless of the intrinsic transcriptional propensity of the integration site itself, as shown in Figure 2B, C. The addition of low-GC flanking regions substantially decreased expression of the reporter at both the *nfi* and *wbbH* loci, with the effects more dramatically evident at *nfi* due to the initially high expression there (Figure 2B, C). Compared to the flank-free reporter, high GC content synthetic flanks did not have a strong impact on reporter expression when the reporter construct was integrated at the high transcriptional propensity site (Figure 2B). However, they substantially increased fluorescence at the low transcriptional propensity site (Figure 2C). We observed consistent results via flow cytometry, and again note tunable changes in expression in response to the changing GC content of the flanking regions (Supplementary Figures S3 and S4; note that mCherry is always at the same site as part of the base strain and is integrated near *yihG*). Likewise, we also observed that synthetic flanks have strong effects on mNG expression from the same reporter encoded on a plasmid, with 65% GC flanks substantially increasing fluorescent reporter expression compared with 35% GC flanks (see Supplementary Figure S5). Together, our findings suggest that the addition of ~1kb flanks with extreme GC content can substantially alter the transcriptional propensity of a site even when the promoter region itself is held fixed. However, GC content cannot be the sole factor leading to position-dependent transcriptional propensity given that altered flanking sequence composition cannot reliably convert the propensity at *wbbH* to that at *nfi* (at the very least, dosage effects (34,35) almost undoubtedly make orthogonal contributions, and other factors may as well).

Genetic context-dependent expression variation is robust to growth condition

Transcriptional propensity and the genetic context effects demonstrated in Figures 1 and 2 were measured in rich
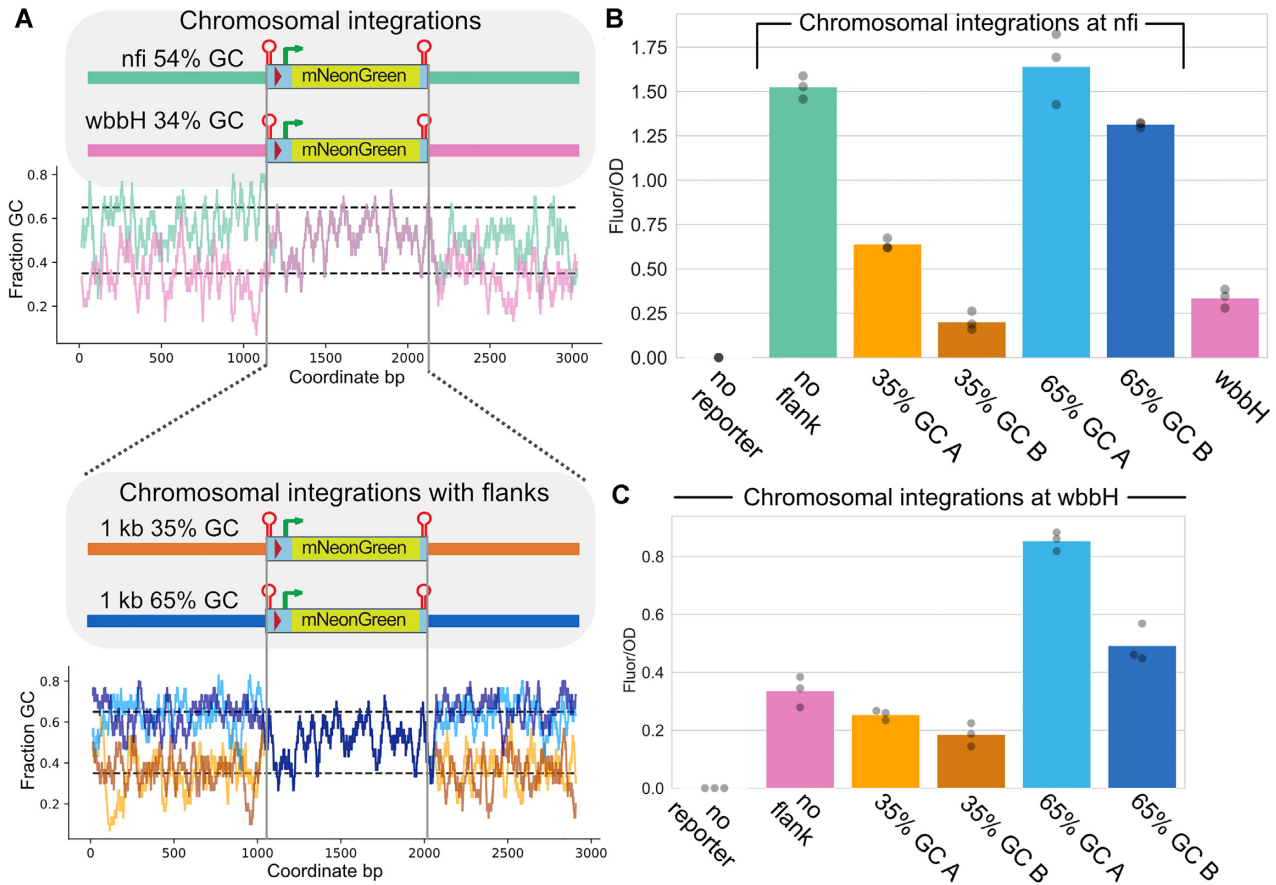
Figure 2. Local low GC content silences reporter gene expression at two sites. (**A**) Diagram of the mNG reporter integrated into the *nfi* (teal) or *wbbH* (pink) locus or the construct flanked by synthetic genetic contexts of 35% (orange) or 65% (blue) GC content and integrated into either the *nfi* or *wbbH* genomic locus. For the synthetic flanks, the light and dark colors represent the two different variants that were used (and correspond to colors in the following panels). The accompanying plot shows GC fraction over a 30 bp sliding window of the mNG construct integrated into the genome (top) or into the genome with synthetic flanks (bottom). 35% GC content is the 1.67 percentile and 65% GC content is the 99.88 percentile of all 500 bp sliding windows of GC content. (**B**) Reporter fluorescence per OD 450 measured in three independent microplate reader experiments for reporters integrated at the *nfi* locus. A chromosomal integration without synthetic flanks at *nfi* and *wbbH* is included for reference. Bars show the geometric means of three biological replicates (which are themselves shown as gray dots). (**C**) Reporter fluorescence per OD 450 measured in three independent microplate reader experiments for reporters integrated at the *wbbH* locus. Bars show the geometric means of three biological replicates (which are themselves shown as gray dots).

medium. We therefore sought to test genetic context effects from a representative set of reporter strains grown in different conditions supporting a wide range of growth rates; the selected growth conditions are summarized in Supplementary Table S1. The included strains sample strong natural genetic context effects at *yagF* and *yafT* with the mNG reporter under the standard TetO1 promoter and the *rplM* promoter. We also examined the effect of synthetic 1 kb flanks composed of 35% or 65% GC content at the *nfi* and *wbbH* sites. Although the different media led to large differences in growth rate (Supplementary Table S1); data for every 10 minutes over 48 hours is visualized in heatmaps in Supplementary Data S4), the genetic context effects tested were generally robust across the conditions tested during logarithmic growth (Figure 3A). There were minor changes to the rank ordering of mNeonGreen expression (assessed by the fluorescence intensity/OD ratio) at lower growth rates, particularly in strains with the mNG reporter integrated at *wbbH*, near the chromosomal terminus, which is consistent with reduced chromosomal dosage effects at

slow growth rates (Figure 3B). Nonetheless, the silencing effects of 35% GC 1 kb flanks remained apparent and cannot arise due to dosage effects; likewise, the *yagF*/*yafT* distinction remained under all growth conditions, and these loci are separated by less than 50 kb. We also collected fluorescence measurements from reporter expression during the stationary phase for several of the faster growth conditions. Although average fluorescence intensity per OD varied strongly depending on the growth medium in the stationary phase, the same effects in terms of both location (*yagF* vs *yafT*, *nfi* versus *wbbH*) and flanking GC content (at the *nfi* and *wbbH* loci) hold under all conditions, although the exact magnitudes of the effects vary (Figure 3C). We also note that cells in stationary phase after growth in rich media (RDM) showed substantially lower fluorescence, regardless of reporter location, potentially due to differences in the regulation of the reporter or other aspects of metabolism during stationary phase entry. The observed differences in media-dependent stationary phase behavior are not dependent upon the exact time in station-
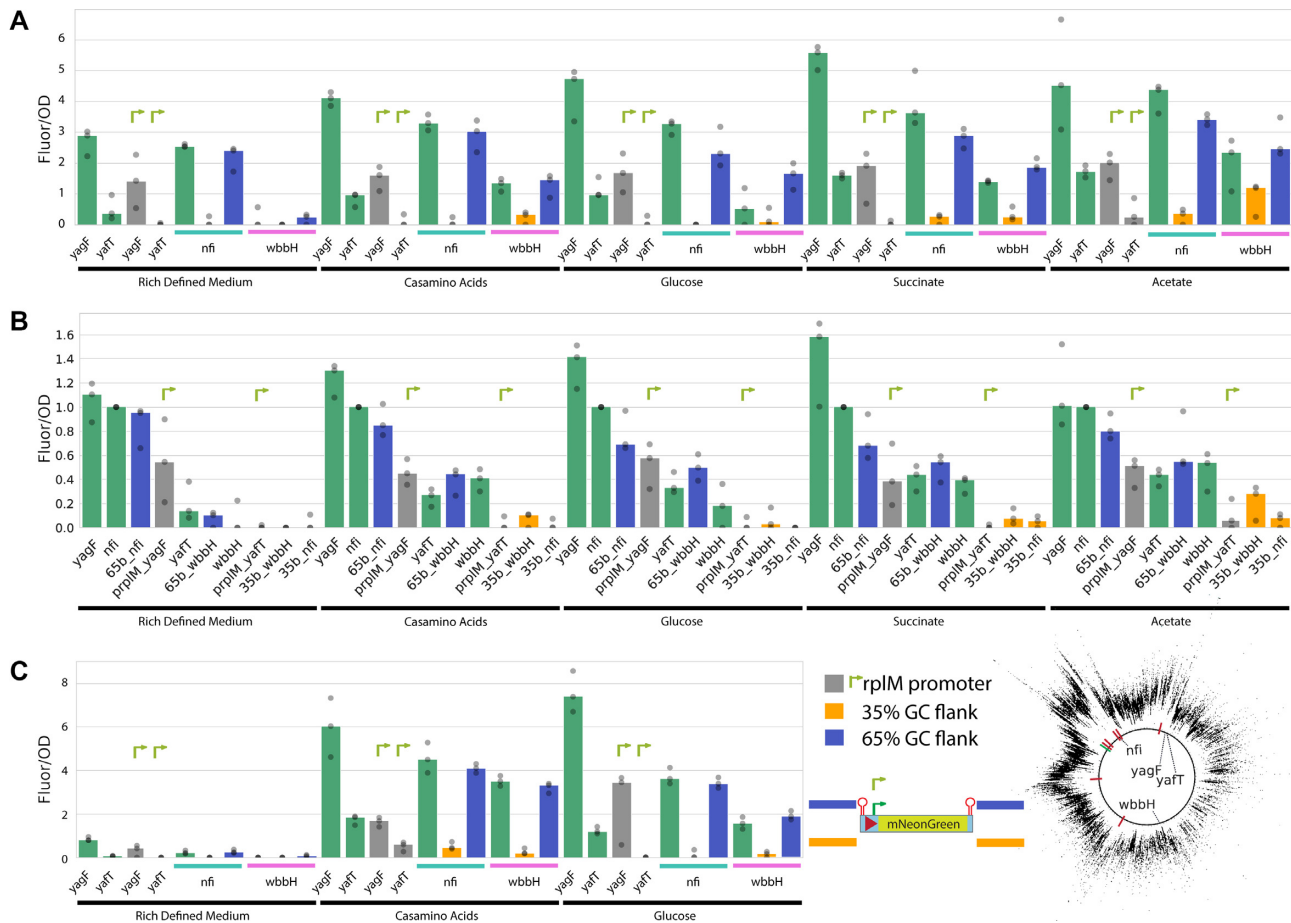
Figure 3. Genetic context affects reporter gene expression similarly under different growth conditions. (**A**) OD-normalized fluorescence during logarithmic growth of strains in different media (see Materials and Methods for details) with the mNG reporter in various representative genetic contexts. The bar represents the median value; each of three replicates is plotted independently (dots). The *yagF* and *yafT* site integrations have mNG under TetO1 or the *rplM* promoter as indicated. mNG integrated at the *nfi* and *wbbH* sites include either 35%, 65% GC 1 kb flanks or no flank, all under the TetO1 promoter. (**B**) The data from (**A**) normalized to *nfi* reporter site strain's fluorescence per OD, within each replicate, and reordered by the intensity of strains grown in RDM shows a similar trend across different media conditions. (**C**) Stationary phase fluorescence per OD was plotted as the average fluorescence per OD between 33 and 42 hours of incubation. The cartoon at the bottom right indicates the locations of the reporter variants shown here (no flank, 35% or 65% GC content flanks, TetO1 or *rplM* promoter). The chromosomal integration locations are indicated on a diagram of the *E. coli* genome with the Origin of replication (green line), ribosomal RNA operons (red line), and transcriptional propensity level (black dots, data from (1)) shown for reference.

ary phase; similar results would be observed anywhere in a wide window of times after stationary phase entry (Supplementary Data S4). Even in the RDM stationary phase samples, though, the *yagF* and *nfi* insertion locations continue to show notably higher fluorescence than other locations (Figure 3C).

Context-driven variations in reporter expression affect equilibrium expression levels but not induction kinetics

Due to the experimental setup of our high-throughput transcriptional propensity profiling, which involved induction for a fixed period of time prior to measurement of transcript levels, low measured transcriptional propensity could arise due to either slow induction of transcription, or low equilibrium levels of transcription at those sites. In principle, It is possible that measured low propensity sites may take longer to reach maximal expression level and might eventually match the fluorescence level of cells with reporters at

high-propensity sites after additional cell doublings in inducing conditions.

To assess the induction rates and maximal expression during induction in the context of our engineered high- and low-propensity reporters (the high- and low-GC flanked reporters discussed in the context of Figure 2, plus a medium propensity unflanked reporter at *eaeH*, in all cases under control of the TetO1 promoter), we passaged growing cells from M9RDM into M9RDM + anhydrotetracycline (aTc) with additional subsequent passages into fresh induction medium in a microplate to maintain the cells in exponential growth. Fluorescence and optical density at 450 nm were monitored every 10 min (Figure 4A; induction begins at 210 min in the timeline on that figure). The maximal fluorescence reached by cells with the reporter integrated at three different sites with different synthetic flanks varied greatly even after multiple passages in the induction medium. Fluorescence levels were consistent with that observed in microplate reader and flow cytometry measure-
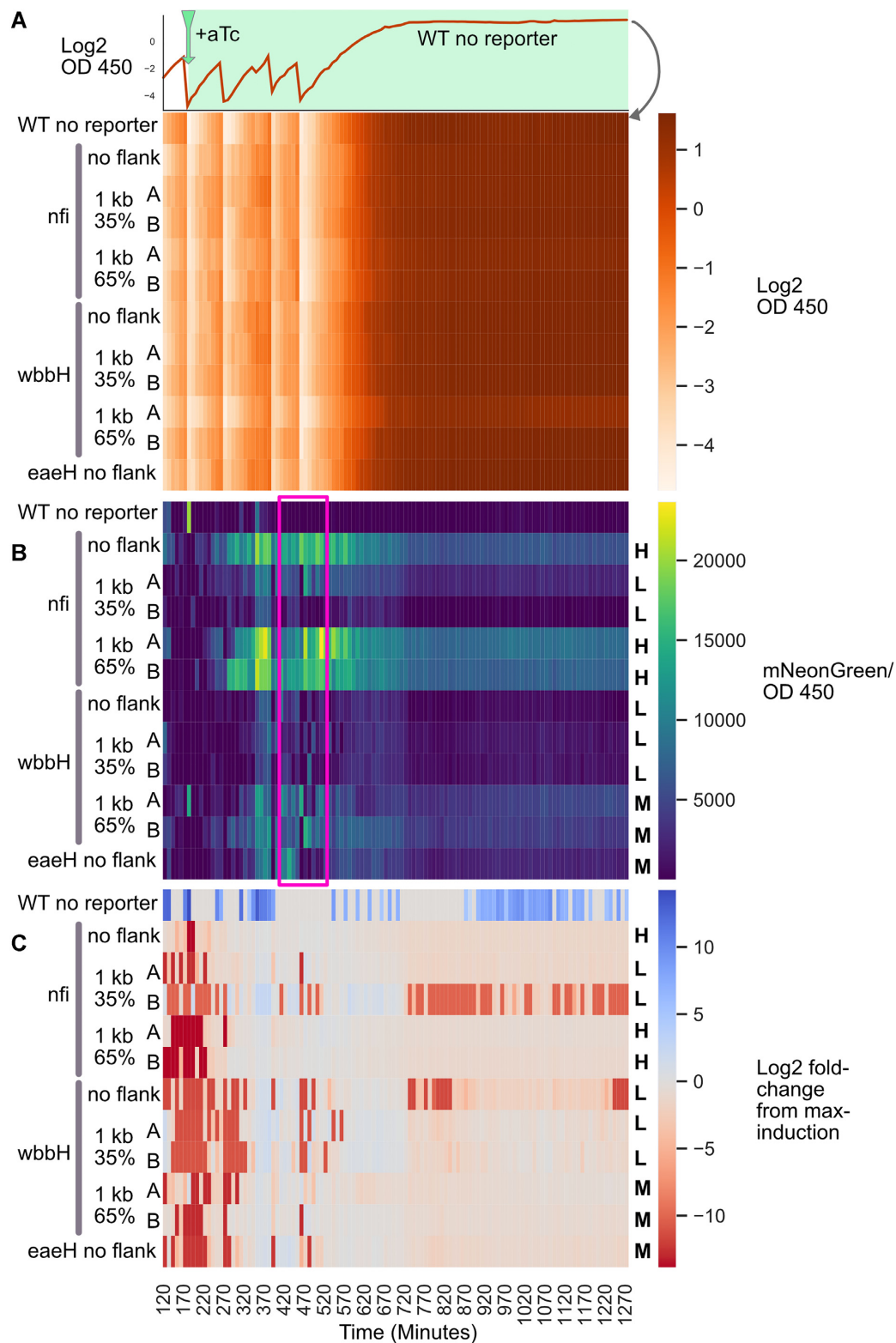
Figure 4. Genetic context effects are robust across timepoints. 12 strains, with and without synthetic flanks, were grown and passaged into fresh medium during growth in a microplate reader. The inducer aTc was added at the first passage time (Green arrow, 210 min) and included in every subsequent passage. A reporter integration at the *eaeH* site was included as a 'medium' expression site for comparison (1 kb of flanking sequence has a GC content of 37%). (**A**) The heat map coloration indicates the $\log_2$ of OD measured at 450 nm. The data for the $\log_2$ OD 450 for only the WT strain without a reporter is shown in a line plot directly above the heat map as a clarifying example representation of the heat map data. (**B**) Fluorescence per OD 450 for all 12 strains. The H, M and L letters stand for high, medium and low expression, respectively, expectations as measured from traditional microplate reader experiments (without passaging). (**C**) $\log_2$ fold-change from maximal induction (defined by the average Fluor/OD included in the purple box in panel B) shows that induction timing from reporter strains with different synthetic genetic contexts is similar.

ments from previous experiments (e.g. Figure 2), and most notably, the equilibrium fluorescence level achieved after several passages – and retained after stationary phase entry – was strongly determined by the transcriptional propensity of the reporter site (Figure 4B). The time required to reach maximal fluorescence was very similar for all strains (Figure 4C), typically stabilizing within 120 min of induction (noting the lag before the observed differences between the observed and maximum values stabilized near 0). Tracking fluorescence at the lower end of cell density after culture passaging introduced enough technical noise to make precise calculations of time required to reach maximal induction unreliable, as transient spikes/dips in the OD-normalized mNeonGreen signal are observed at low OD after dilution. Nevertheless, we can assert that variations in induction timing, if any, are on the order of tens of minutes (Figure 4C). For comparison, the mixed reporter library strain that was used to generate our previous high-resolution transcriptional propensity map was induced for 210 min prior to harvest (1), equivalent to the 420th minute in the induction experiment, which corresponds well to the equilibrium expression level. Thus, we find that the expression levels reported throughout this work and in (1) are comparably in the log-phase of growth and correspond to fully induced time-points (Figure 4).

Ribosomal RNA operons themselves are not necessary for ribosomal RNA operon-proximal transcriptional propensity peaks

Because several transcriptional propensity peaks are centered on ribosomal RNA operons (*rrn*; see Supplementary Figure S1A), we sought to determine whether the *rrn* operons cause the peak formation. To that end, we either integrated our standard reporter construct upstream of *rrn* promoters and their associated Fis binding sites (*rrn* int, in which the Fis sites are unperturbed), or replaced the entire operon and its promoter region with the reporter (*rrn* KO; see Figure 5). Despite the striking centrality of *rrn* in transcriptional propensity peaks, replacing an *rrn* with the reporter (*rrn* KO) did not cause any decrease in reporter expression compared to the upstream integration. Additionally, we did not detect any changes in growth rate (measured via plate reader growth curves) or overall cell protein expression (using our internal mCherry control as measured by flow cytometry), either of which could otherwise confound interpretation of fluorescence data if they changed (Supplementary Figures S6 and S7). Thus, we find that the effects of proximity to the *rrn* are completely independent of the presence of *rrn* transcription or even the *rrn* operon itself, and must instead reflect some more general characteristic of the chromosomal regions in which *rrn* operons are found. To further investigate the roles played by different portions of the *rrn*-proximal regions in setting the transcriptional propensity near them, we also tested a third *rrnE* variant in which the ribosomal operon and its annotated promoters were removed, but the upstream Fis binding sites (known to be important in regulating transcription) were retained – we refer to this construct as 'KO+'. The 'KO+' variant showed a modest increase in fluorescence even over the corresponding 'KO', suggesting that the Fis binding sites may

yet further increase RNA polymerase recruitment to this region even in the absence of the *rrn* operon itself (Figure 5).

Although the *rrn* KOs generally had little to no effect on reporter expression, we noticed that the *rrnD* KO in particular had higher reporter expression compared to the *rrnD* int (*rrnD* upstream integration). To test whether the *rrnD* KO caused a local or general increase in mNG expression, we replaced *rrnD* with a kanamycin resistance marker in strains with mNG integrated into different locations. The *rrnD* KO caused similar increases in reporter fluorescence in all of the sites we tested, whereas two other *rrn* deletions (*rrnA* and *rrnC*) did not (Supplementary Figure S8). However, the *rrnD* KO did not result in an increase of mCherry fluorescence (Supplementary Figure S9). It is possible that removal of *rrnD* has a particularly strong effect on other high transcriptional propensity sites (all three sites considered in Supplementary Figure S8 are high propensity and fairly close to other *rrn* operons), although the reason for the outsized effect of *rrnD* deletion relative to deletions of other *rrn* operons remains unclear.

Transcriptionally insulated reporters are affected by strongly expressed genes in a orientation specific manner

Consistent with previous studies, we observed that expression from the mNG reporter integrated directly adjacent to the highly-expressed *rrnE* operon varied depending on the relative gene orientation (3,26). These effects are apparent in a re-analysis of our original transcriptional propensity data from (1), and demonstrate that on average, integrations upstream of the ribosomal RNA show higher expression than those downstream of a ribosomal RNA, and that reporters integrated downstream of the *rrn* and head-on with its transcriptional direction are particularly low in expression relative to other *rrn*-proximal locations (albeit still higher than the genome-wide average; see Figure 6A). Because orientation-specific effects are strong on a short distance scale (26) but our transcriptional propensity data must be averaged over a window to ensure robust results, we performed targeted reporter integrations near the *rrnE* operon with consistent spacing in order to directly measure the effects of relative location and orientation of our reporter. By testing all possible orientations of the well-insulated mNG reporter integrated at *rrnE*-adjacent sites, we confirmed that sites downstream of this highly-expressed operon show lower transcription compared to upstream integrations (Figure 6). Reporter integrations transcribed in the same direction (tandem orientation) with respect to the *rrnE* operon, whether upstream or downstream, were expressed more highly than the corresponding reporter integrated into the same site, but in the divergent or convergent (head-on) orientation. Our findings are generally consistent with a previous study demonstrating high transcriptional interference between convergently oriented fluorescent protein genes encoded on a plasmid and assessed in a microplate reader (26), and are fully consistent with expectations based on a buildup of negative supercoiling upstream of a highly expressed operon, and positive supercoiling downstream of it. Unlike virtually all other cases considered through our work, in the case of the directional integrations shown in Figure 6, we observed
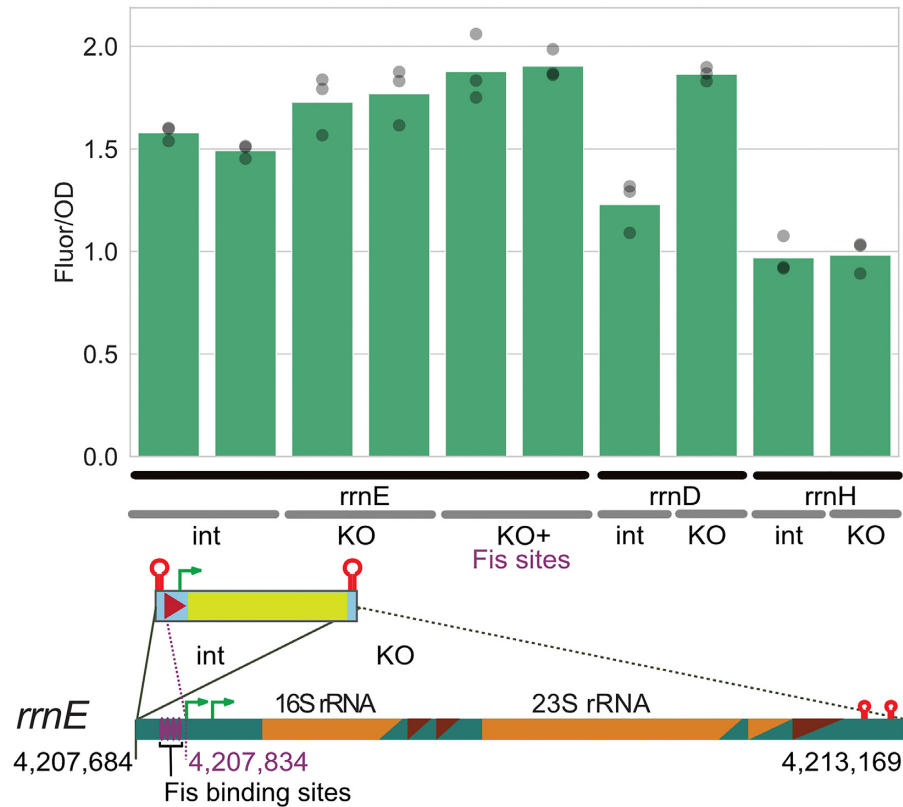
Figure 5. Ribosomal RNA operons are dispensable for regional high transcriptional propensity. The reporter was either integrated upstream (int) of the promoters of ribosomal RNA operons (*rrnE, rrnD, rrnH*) or replacing the entire operon and regulatory elements from the same site upstream of the promoter until after the ribosomal operon terminators (KO), as indicated in the diagram. A KO at *rrnE* that retains the Fis binding sites upstream of the reporter was also tested (KO+; dashed purple line indicates the upstream integration site). Reporter fluorescence per OD 450 was measured in three independent microplate reader experiments. Integration coordinates are for the U00096.3 complete MG1655 genome, the current EcoCyc standard. Cases where two bars are shown for a given genotype indicate two independently constructed clonal lineages; for each bar, the geometric mean of three biological replicates is shown, with the individual replicates as gray points.

a discrepancy in rank ordering between the plate reader-based bulk fluorescence data and flow cytometry data (Supplementary Figure S10); in the latter case, the lowest expression was observed for the Downstream-Tandem rather than Downstream-Convergent orientation. While the vast majority of measurements were concordant between the methods, for an unidentified reason, there was a discrepancy between fluorescent measurements in the microplate reader and flow cytometer in the isolated case of strains with mNG integrated downstream of *rrnE* in the convergent orientation (highlighted in Supplementary Figure S11 and all-strain correlations in Supplementary Table S2). Therefore, we repeated microplate measurements for a subset of specific strains showing a discrepancy between the approaches. This repeated trial confirmed a strong reduction in fluorescence of the mNG reporter integrated downstream of *rrnE* in the convergent orientation when measured by microplate reader (Supplementary Figure S4), supporting the conclusions drawn from Figure 6. The remaining divergence between the measurement modalities is possibly due to some specific sensitivity of the *rrn* regions to the harvesting conditions required for flow cytometry measurements.

DISCUSSION

We investigated the contribution of the chromosomal features with the strongest correlations to position-dependent transcriptional propensity. As local GC content, proximity to *rrn*, and binding of the NAPs H-NS and Fis co-vary with transcriptional propensity, they were our top candidates for chromosomal features that may be functionally responsible for position-dependent expression variation (1). In general, we designed a strategy to fix a reporter position and sequence while varying the genetic context around it in order to test the contribution of chromosomal features on gene expression. The transcriptional propensity from our previous study was broadly consistent with both the RNA level and fluorescent protein expression level measured from targeted integrations (Supplementary Figure S1-S2), which allowed us to report mNG fluorescence at the population and single-cell level to assess reporter expression due to position-dependent transcription effects. We note that in contrast to the majority of recent work on the effects of sequence contexts on transcription, which focus on the regulatory sequences immediately upstream of a gene (the promoter and UP-element region, 1–125 bp upstream of the transcription start site), here we have kept a fixed sequence
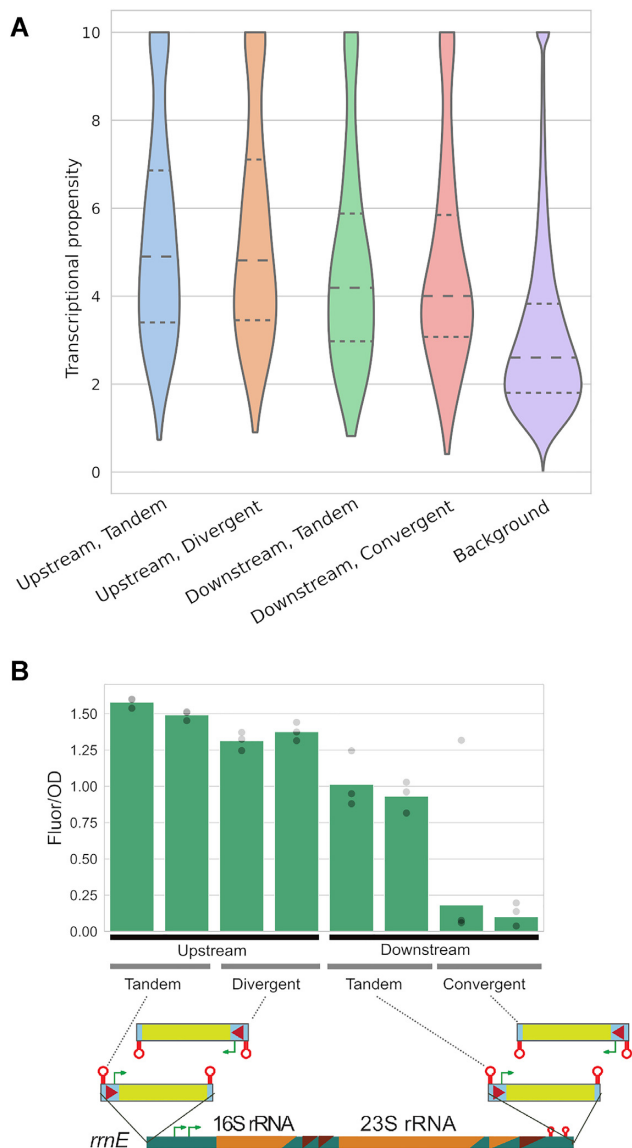
Figure 6. Ribosomal RNA operons affect local direction-dependent gene expression. (**A**) Re-analysis of the individual site-level transcriptional propensity values from (1), in which each site was categorized based on being either upstream or downstream of an *rrn* operon, and oriented either in the same direction (tandem) or opposite direction (divergent or convergent) as that operon. *rrn* operon boundaries were obtained from RegulonDB (65). For each category, the median and quartiles are shown as dashed lines; all transcriptional propensity values above 10.0 were clamped to 10.0 to provide simplified visualization. In each *rrn*-proximal case the distributions shown cover a 5 kb window adjacent to the operon in the specified direction. (**B**) The reporter construct was integrated into the genome either upstream of the *rrnE* promoters or downstream of the *rrnE* terminators in all possible orientations as indicated in the diagram. Reporter fluorescence per OD 450 nm measured in three independent microplate reader experiments. The green bar corresponds to the geometric mean of the three replicates (gray circles). Side-by-side independent lineages were derived from separate colonies in the initial reporter integration (shown as separate bars). We observed a single outlier in one of the technical replicates from a strain with the reporter integrated into the convergent (head-on) orientation with respect to *rrnE*, but it does not substantively affect our conclusions (Although this is the apparent outlier in microplate replicate 1, shown in Supplementary Figure S7, it is still included in the geometric averaging and data analysis shown; see also Supplementary Figure S4 for an independent repetition of *rrnE* downstream strains).

up to 162 bp upstream and 852 bp downstream of the transcription start site, and instead focused on the longer-range impacts of genetic context. Our data demonstrate that, in parallel to the extensively studied fundamental effects of regulatory sequences (6,36–40), genetic context further than 162 bp from the transcription start site has a profound effect on transcription level, and must be considered both in efforts to model transcription of natural genes, and in the design of synthetic biology constructs.

Transcriptional propensity affects natural promoters of varying strengths

In our previous work on empirical transcriptional propensity mapping, we used the TetO1 promoter to drive mNG expression and generate the map. As are many promoters used in synthetic biology, the TetO1 promoter is relatively strong, ranking in the 86th percentile of *E. coli* transcriptional output. We wondered whether weaker promoters would be similarly affected by genetic context. Competitive interplay between H-NS and RNAP may lead to greater position-dependent H-NS silencing effects on genes that are naturally more weakly expressed (41). On the other hand, context-dependent silencing of weak promoters may be less apparent, as silencing a weak promoter would result in a relatively small decrease in signal compared to silencing a strong promoter assuming some basal level of transcriptional activity (42). By swapping only the reporter promoter to one of three additional sequences with a range of strengths at three chromosomal sites (*eaeH*, *yagF*, and *yafT*) that are close to each other (and therefore have comparable DNA copy number) but have different transcriptional propensity, we demonstrate that natural promoters are subject to similar context-dependent expression variation (Figure 1). The same conclusion has also been drawn from many thousands of promoters tested at three different genomic loci (43). Only our weakest promoter was unaffected by the genomic context, which is consistent with a context-independent transcriptional activity baseline. Indeed, the detection of RNA that is antisense to genes or intergenic in high-coverage RNA-seq experiments, which may be deleterious to cells, is presumably a result of 'leaky' transcription in at least some cells (42,44,45) and may occur at some level independent of any transcriptional propensity effects or regional silencing.

Extreme GC content flanking regions overrides regional transcriptional propensity

H-NS binds to and silences sites of low GC content (17–20). In our high-resolution transcriptional mapping study, we found that the mNG reporter (54% GC content) became silenced when it integrated into H-NS bound regions at different sites across the genome (1). Here, we report that synthetic context flanks composed of 35% GC can strongly silence transcription even in an otherwise high transcriptional propensity site, and that 65% GC flanks can partially relieve the repression of a low transcriptional propensity site (Figure 2). How can flanking DNA have such a dramatic effect on gene expression? H-NS may be forming silencing filaments from high-affinity low GC sites flanking the reporter

(46) to invade and silence the reporter itself. Previous work has also shown that H-NS and associated proteins StpA and Hha can form bridge filaments leading to early termination of RNAP elongation (21,47). These mechanisms may explain the reduced transcriptional propensity for an otherwise highly-expressed and GC neutral (54% GC) reporter when integrated into low GC content genomic regions bound by H-NS, such as near *wbbH*.

We note that in our sampling of different promoters and placement contexts, the mNG concentration in individual cells within a population always shows a unimodal shift in cell-level expression (Figure 1 and Supplementary Figure S3). This indicates that the drivers of transcriptional propensity that are being modulated here (possibly silencing by H-NS) are not stochastic events that are biased towards a higher frequency due to context, but rather affect all cells in a tunable manner. Furthermore, we never observed the appearance of a secondary population of cells with a distinct fluorescence level in any of the strains tested in this study. Whether the reporter fluorescence was decreased by low flanking GC content or by transcriptional interference from *rrnE*, the cells fluorescence level shifted as a unimodal population. We emphasize that the accumulation of cells with low mNG concentration represented in the flow cytometry violin plots in this article is simply the result of assigning cells without detectable fluorescence above the median autofluoresce level (from reporter-free cells) to a minimum fluorescence level across experiments. This assignment clearly indicates the point at which cells do not have detectable fluorescence, while maintaining the median and quantile values of the population level fluorescence.

A previous study has shown that altering the GC content within a gene in a synthetic operon can affect the expression of an adjacent gene in the same operon (48), and attributed these to changes in RNA folding energy. However, given that our experiments show expression effects in the same direction as was observed in that study, simple genetic context effects may be responsible for the observed expression differences, or are working in concert with the changes to RNA folding energy reported in that study. RNA folding energy cannot serve as an explanation for our observations because the sequence of the transcript generated across all of the reporters considered in the present study is identical. Non-coding, low GC content can also cause silencing of a neutral-GC reporter in the chromosome and on a plasmid. This low GC content can also be evolutionarily favored, at least on a plasmid, by utilizing the more plentiful intracellular A+T nucleotide pools, independent of the specific plasmid sequence (49). Taken together, these findings suggest that low GC content DNA can be horizontally transferred and maintained at a relatively low cost and a low level of expression of potentially toxic gene products (even if short stretches of neutral GC content are included), which represents a reservoir of latent genetic material shared between compatible bacteria, as has been extensively discussed in the context of xenogeneic silencers (50). From an engineering perspective, it would be desirable to precisely predict and tune gene expression, taking genetic context into account. Indeed, the findings here warrant a greatly expanded survey of genetic context variants like the flanking regions investigated here (e.g. fine-tuning flanks between 10–90% GC),

and predictive modeling for synthetic biology applications on the chromosome and plasmids, similar to previous efforts for ribosome binding sites and promoters (36,51).

Genetic context exerts a strong influence on gene expression, irrespective of growth phase or condition

The effect of growth condition and phase on the global pattern of transcriptional propensity across the chromosome remains to be fully elucidated. However, results presented here show a clear and strong effect of different GC content genetic contexts, irrespective of growth phase or condition tested (Figure 3). Since GC content fraction is one of the most-well correlated features to transcriptional propensity, these results suggest that at least a major part of the position-dependent expression variation previously observed is not regulatory in nature (at least not across a broad range of growth rates tested here). This leaves open the question of the evolutionary significance of silenced low GC content regions, in particular. The enrichment of prophages in low transcriptional propensity regions suggests one possible explanation, but there are some exceptions, such as the GC-rich CP4-6 prophage region, where the *yagF* site is located. Low GC-content regions contain mobile genetic elements, which are capable of insertional inactivation of toxic genes and mediation of genomic rearrangements (52,53). Rare expression of mobile and prophage elements in combination with more strongly selective conditions than were tested here may be sufficient to explain the existence of stable low-GC regions that are silencing in several conditions (54,55).

Transcriptional propensity affects maximal induction but not induction kinetics

Although we observed a dramatic reduction in mNG level from reporters integrated into low GC content synthetic contexts, there were no major differences in the time required for maximal induction of the reporter (Figure 4). The silencing effect of low GC content flanks was also completely resistant to extended induction times. In other words, transcriptional activity did not cause further derepression of mNG expression in low GC contexts, unless over short periods of tens of minutes or less. Thus, it appears that the context effects that we have termed transcriptional propensity act mainly on the equilibrium rate of transcription achieved. The induction data also show that our previous *E. coli* random integration library cells, used to produce our high-resolution transcriptional propensity map, should have achieved the maximal expression level achievable given the genetic context of each reporter.

The high transcriptional propensity near ribosomal RNA operons is a genomic context feature that does not require the operon or its promoter

Ribosomal RNA operons (*rrn*) are another key structural element of the *E. coli* chromosome as they form some chromosome interacting domain boundaries (56). *rrn* can also form nucleolus-like zones in the nucleoid where they colocalize with foci of actively transcribing

RNAP ([24](),[57](),[58]()), with focus formation dependent upon the nucleoid-associated protein HU ([59]()). Directly around *rrn* there are broad transcriptional propensity peaks, suggesting that these chromosomal regions may be brought into or are affected by transcription factories ([1]()). In order to test the functional contribution of the *rrn* themselves to the high transcriptional propensity around them, we made targeted reporter integrations either upstream of, or replacing, the *rrn*. Replacement of the *rrn* did not cause a decrease in reporter fluorescence compared to the corresponding upstream integration (Figure [5]()). The *rrn* KO removes the entire transcription unit and regulatory features, including important features required for *rrn* colocalization, such as the promoters ([24]()). Thus, the *rrn* operons and their *cis* regulatory regions do not set the transcriptional propensity in the genomic region where these operons are found, but rather, the neighboring context must have evolved to enhance expression from the regions of *rrn*. Future work may determine the effect of physical localization of chromosomal regions and of potentially unidentified sequence features on local transcriptional propensity. Unexpectedly, some of the *rrn* replacements with the reporter construct in fact caused an increase in reporter fluorescence. The effect was especially pronounced for the *rrnD* replacement. We found that *rrnD* KO by a Kanamycin resistance cassette also caused an increase in reporter fluorescence that was integrated into other sites around the genome (Supplementary Figure S8). We could not attribute this change in reporter fluorescence to a change in cell size or growth rate (Supplementary Figure S7), nor was mCherry fluorescence from a gene present in our base strain affected by the *rrnD* KO (Supplementary Figure S9). A change in relative expression level for a cluster of translation factors proximal to *rrnD* is a possible explanation for these observations. Testing of additional reporter integration sites in the WT and *rrnD* KO context might determine whether some genomic regions are in competition with *rrnD* for limiting expression resources, such as RNAP.

Supercoiling interference extends past transcript boundaries and transcriptional terminators

Finally, we also confirmed interference effects that have previously been observed on gene expression when encoded downstream of an induced gene, regardless of the downstream gene orientation ([3]()). In order to test the interference of a strongly expressing gene on the mNG reporter, which is insulated by very strong well-characterized terminators ([32]()), we integrated the mNG reporter in all possible orientations around the highly expressed *rrnE* operon, without disrupting the promoter, associated Fis binding sites, or the operon terminators. Consistent with the previous studies, we find that the strongest interference effects on reporter expression when the reporter is integrated downstream of the strong transcriptional unit *rrnE,* especially in the convergent orientation (Figure [6]()). In general, our results regarding the effects of supercoiling interference for a reporter integrated into the genome are in agreement with previous results in which reporter genes were encoded on a plasmid ([26]()). We also note that differences in supercoiling were previously invoked as one possible explanation for the

pleiotropic and far-reaching effects of relocating the *dusB-fis* operon itself ([60]()). There may be a broader range of mNG fluorescence in strains with the reporter integrated in the convergent orientation (Supplementary Figure S10, Figure S4), which might be consistent with an observation of increased expression 'noise' for a reporter integrated into *rrn* operons in a previous study ([16]()). However, expression noise from a reporter integrated at a site within an *rrn* may also be due to reporter loss by recombination and a competitive advantage for cells with seven functional *rrn*, especially in rapid growth conditions.

Repressive effects have also been observed for genes in the divergent orientation even over relatively long (>2 kb) intervening lengths, but the length limit of this effect is not known ([30]()). *In vitro* evidence suggests that DNA gyrase can relieve the effects of divergent supercoiling competition ([26]()). In general, our findings are consistent with divergent supercoiling competition, but the effect size is small (Figure [6]()). Genetic interference effects can be further complicated by transcription units without completely insulating terminators ([61]()). Subtleties in the exact genetic construct design and spacing are likely responsible for variations in effect size reported between and within studies. Future work on a greater diversity of genetic construct designs using a similar induction matrix as Yeung et al. ([26]()) and single-cell analysis may be an important, albeit intensive, undertaking to fully understand competition between neighboring transcriptional units. Expanded mechanistic insights into context effects will allow utilization of genetic context as a modulating control mechanism in synthetic biology ([62]()). Regardless of the application, genetic context should be taken into consideration. This has been demonstrated by a recent effort to standardize genomic integration designs, where sequence-controlled expression level and integration frequency were variable depending on the landing pad position ([63]()). The findings presented here demonstrate not only the importance of such endeavors, but the long range of context effects that must be considered, and the extent to which synthetic flanking sequences can overcome natural context effects on the bacterial chromosome.

CONCLUSION

Local and regional genetic contexts have a strong influence on expression from a transcriptionally insulated reporter expressed from both strong and weak promoters. These context effects are key considerations that should be taken in parallel with other genetic construct design choices like promoter, RBS and terminator for plasmid cloning and for integration into genomes. Genetic context manipulation is also a unique tool in the genetic toolbox for tuning gene expression without altering the coding or regulatory element sequence for genes in synthetic metabolic pathways or genetic circuits.

DATA AVAILABILITY

Flow cytometry data from this study have been deposited at FlowRepository (https://flowrepository.org/) under repository ID FR-FCM-Z529.

REFERENCES

1. Scholz,S.A., Diao,R., Wolfe,M.B., Fivenson,E.M., Lin,X.N. and Freddolino,P.L. (2019) High-Resolution mapping of the *Escherichia coli* chromosome reveals positions of high and low transcription. *Cell Syst.*, **8**, 212–225.
2. Block,D.H.S., Hussein,R., Liang,L.W. and Lim,H.N. (2012) Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Res.*, **40**, 8979–8992.
3. Bryant,J.A., Sellars,L.E., Busby,S.J.W. and Lee,D.J. (2014) Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res.*, **42**, 11383–11392.
4. Brambilla,E. and Sclavi,B. (2015) Gene regulation by H-NS as a function of growth conditions depends on chromosomal position in *Escherichia coli*. *G3*, **5**, 605–614.
5. Berger,M., Gerganova,V., Berger,P., Rapiteanu,R., Lisicovas,V. and Dobrindt,U. (2016) Genes on a wire: the nucleoid-associated protein HU insulates transcription units in *Escherichia coli*. *Sci. Rep.*, **6**, 31512.
6. Urtecho,G., Tripp,A.D., Insigne,K.D., Kim,H. and Kosuri,S. (2019) Systematic dissection of sequence elements controlling σ70 promoters using a genomically encoded multiplexed reporter assay in *Escherichia coli*. *Biochemistry*, **58**, 1539–1551.
7. Shen,B.A. and Landick,R. (2019) Transcription of bacterial chromatin. *J. Mol. Biol.*, **431**, 4040–4066.
8. Dame,R.T., Rashid,F.-Z.M. and Grainger,D.C. (2020) Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat. Rev. Genet.*, **21**, 227–242.
9. Goormans,A.R., Snoeck,N., Decadt,H., Vermeulen,K., Peters,G., Coussement,P., Van Herpe,D., Beauprez,J.J., De Maeseneire,S.L. and Soetaert,W.K. (2020) Comprehensive study on *Escherichia coli* genomic expression: does position really matter? *Metab. Eng.*, **62**, 10–19.
10. Jeong,D.-E., So,Y., Park,S.-Y., Park,S.-H. and Choi,S.-K. (2018) Random knock-in expression system for high yield production of heterologous protein in *Bacillus subtilis*. *J. Biotechnol.*, **266**, 50–58.
11. Saleski,T.E., Chung,M.T., Carruthers,D.N., Khasbaatar,A., Kurabayashi,K. and Lin,X.N. (2021) Optimized gene expression from bacterial chromosome by high-throughput integration and screening. *Sci. Adv.*, **7**, eabe1767.
12. Carruthers,D.N., Saleski,T.E., Scholz,S.A. and Lin,X.N. (2020) Random chromosomal integration and screening yields K-12 derivatives capable of efficient sucrose utilization. *ACS Synth. Biol.*, **9**, 3311–3321.
13. Loeschcke,A., Markert,A., Wilhelm,S., Wirtz,A., Rosenau,F., Jaeger,K.-E. and Drepper,T. (2013) TREX: a universal tool for the transfer and expression of biosynthetic pathways in bacteria. *ACS Synth. Biol.*, **2**, 22–33.
14. Park,Y., Espah Borujeni,A., Gorochowski,T.E., Shin,J. and Voigt,C.A. (2020) Precision design of stable genetic circuits carried in highly-insulated *E. coli* genomic landing pads. *Mol. Syst. Biol.*, **16**, e9584.
15. Englaender,J.A., Andrew Jones,J., Cress,B.F., Kuhlman,T.E., Linhardt,R.J. and Koffas,M.A.G. (2017) Effect of genomic integration location on heterologous protein expression and metabolic engineering in *E. coli*. *ACS Synth. Biol.*, **6**, 710–720.
16. Yousuf,M., Iuliani,I., Veetil,R.T., Seshasayee,A.S.N., Sclavi,B. and Lagomarsino,M.C. (2020) Early fate of exogenous promoters in *E. coli*. *Nucleic Acids Res.*, **48**, 2348–2356.
17. Kahramanoglou,C., Seshasayee,A.S.N., Prieto,A.I., Ibberson,D., Schmidt,S., Zimmermann,J., Benes,V., Fraser,G.M. and Luscombe,N.M. (2011) Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res.*, **39**, 2073–2091.
18. Lucchini,S., Rowley,G., Goldberg,M.D., Hurd,D., Harrison,M. and Hinton,J.C.D. (2006) H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog.*, **2**, e81.
19. Grainger,D.C., Hurd,D., Goldberg,M.D. and Busby,S.J.W. (2006) Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res.*, **34**, 4642–4652.
20. Oshima,T., Ishikawa,S., Kurokawa,K., Aiba,H. and Ogasawara,N. (2006) Escherichia coli histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Res.*, **13**, 141–153.
21. Boudreau,B.A., Hron,D.R., Qin,L., van der Valk,R.A., Kotlajich,M.V., Dame,R.T. and Landick,R. (2018) StpA and Hha stimulate pausing by RNA polymerase by promoting DNA–DNA bridging of H-NS filaments. *Nucleic Acids Res.*, **46**, 5525–5546.
22. French,S.L. and Miller,O.L. Jr (1989) transcription mapping of the *Escherichia coli* chromosome by electron microscopy. *J. Bacteriol.*, **171**, 4207–4216.
23. Paul,B.J., Ross,W., Gaal,T. and Gourse,R.L. (2004) rRNA transcription in Escherichia coli. *Annu. Rev. Genet.*, **38**, 749–770.
24. Gaal,T., Bratton,B.P., Sanchez-Vazquez,P., Sliwicki,A., Sliwicki,K., Vegel,A., Pannu,R. and Gourse,R.L. (2016) Colocalization of distant chromosomal loci in space in *E. coli*: a bacterial nucleolus. *Genes Dev.*, **30**, 2272–2285.
25. Walker,D.M., Freddolino,P.L. and Harshey,R.M. (2020) A well-mixed E. coli genome: widespread contacts revealed by tracking mu transposition. *Cell*, **180**, 703–716.
26. Yeung,E., Dy,A.J., Martin,K.B., Ng,A.H., Del Vecchio,D., Beck,J.L., Collins,J.J. and Murray,R.M. (2017) Biophysical constraints arising from compositional context in synthetic gene networks. *Cell Syst.*, **5**, 11–24.
27. Bordoy,A.E., Varanasi,U.S., Courtney,C.M. and Chatterjee,A. (2016) Transcriptional interference in convergent promoters as a means for tunable gene expression. *ACS Synth. Biol.*, **5**, 1331–1341.
28. Brophy,J.A.N. and Voigt,C.A. (2016) Antisense transcription as a tool to tune gene expression. *Mol. Syst. Biol.*, **12**, 854.
29. O'Connor,N.J., Bordoy,A.E. and Chatterjee,A. (2021) Engineering transcriptional interference through RNA polymerase processivity control. *ACS Synth. Biol.*, **10**, 737–748.
30. Kim,S., Beltran,B., Irnov,I. and Jacobs-Wagner,C. (2019) Long-Distance cooperative and antagonistic RNA polymerase dynamics via DNA supercoiling. *Cell*, **179**, 106–119.
31. Neidhardt,F.C., Bloch,P.L. and Smith,D.F. (1974) Culture Medium for Enterobacteria. *J. Bacteriol.*, **119**, 736–747.
32. Chen,Y.-J., Liu,P., Nielsen,A.A.K., Brophy,J.A.N., Clancy,K., Peterson,T. and Voigt,C.A. (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods*, **10**, 659–664.
33. 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.
34. Sousa,C., de Lorenzo,V. and Cebolla,A. (1997) Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology*, **143**, 2071–2078.
35. Schmid,M.B. and Roth,J.R. (1987) Gene location affects expression level in *Salmonella typhimurium*. *J. Bacteriol.*, **169**, 2872–2875.
36. Fleur,T.L., La Fleur,T., Hossain,A. and Salis,H.M. (2022) Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. *Nat. Comm.*, **13**, 5159.
37. Mejía-Almonte,C., Busby,S.J.W., Wade,J.T., van Helden,J., Arkin,A.P., Stormo,G.D., Eilbeck,K., Palsson,B.O., Galagan,J.E.

and Collado-Vides,J. (2020) Redefining fundamental concepts of transcription initiation in bacteria. *Nat. Rev. Genet.*, **21**, 699–714.

38. Kosuri,S., Goodman,D.B., Cambray,G., Mutalik,V.K., Gao,Y., Arkin,A.P., Endy,D. and Church,G.M. (2013) Composability of regulatory sequences controlling transcription and translation in *Escherichia coli. Proc. Natl. Acad. Sci. U.S.A.*, **110**, 14024–14029.

39. Johns,N.I., Gomes,A.L.C., Yim,S.S., Yang,A., Blazejewski,T., Smillie,C.S., Smith,M.B., Alm,E.J., Kosuri,S. and Wang,H.H. (2018) Metagenomic mining of regulatory elements enables programmable species-selective gene expression. *Nat. Methods*, **15**, 323–329.

40. Klein,C.A., Teufel,M., Weile,C.J. and Sobetzko,P. (2021) The bacterial promoter spacer modulates promoter strength and timing by length, TG-motifs and DNA supercoiling sensitivity. *Sci. Rep.*, **11**, 24399.

41. Landick,R., Wade,J.T. and Grainger,D.C. (2015) H-NS and RNA polymerase: a love-hate relationship? *Curr. Opin. Microbiol.*, **24**, 53–59.

42. Haas,B.J., Chin,M., Nusbaum,C., Birren,B.W. and Livny,J. (2012) How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics*, **13**, 734.

43. Urtecho,G., Insigne,K.D., Tripp,A.D., Brinck,M., Lubock,N.B., Kim,H., Chan,T. and Kosuri,S. (2020) Genome-wide Functional Characterization of *Escherichia coli* Promoters and Regulatory Elements Responsible for their Function. bioRxiv doi: https://doi.org/10.1101/2020.01.04.894907, 06 January 2020, preprint: not peer reviewed.

44. Dornenburg,J.E., Devita,A.M., Palumbo,M.J. and Wade,J.T. (2010) Widespread antisense transcription in *Escherichia coli. MBio*, **1**, e00024-10.

45. Raghavan,R., Sloan,D.B. and Ochman,H. (2012) Antisense transcription is pervasive but rarely conserved in enteric bacteria. *MBio*, **3**, e00156-12.

46. Gao,Y., Foo,Y.H., Winardhi,R.S., Tang,Q., Yan,J. and Kenney,L.J. (2017) Charged residues in the H-NS linker drive DNA binding and gene silencing in single cells. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 12560–12565.

47. Kotlajich,M.V., Hron,D.R., Boudreau,B.A., Sun,Z., Lyubchenko,Y.L. and Landick,R. (2015) Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *Elife*, **4**, e04970.

48. Wu,F., Zhang,Q. and Wang,X. (2018) Design of adjacent transcriptional regions to tune gene expression and facilitate circuit construction. *Cell Syst.*, **6**, 206–215.

49. Dietel,A.-K., Merker,H., Kaltenpoth,M. and Kost,C. (2019) Selective advantages favour high genomic AT-contents in intracellular elements. *PLos Genet.*, **15**, e1007778.

50. Singh,K., Milstein,J.N. and Navarre,W.W. (2016) Xenogeneic silencing and its impact on bacterial genomes. *Annu. Rev. Microbiol.*, **70**, 199–213.

51. Reis,A.C. and Salis,H.M. (2020) An automated model test system for systematic development and improvement of gene expression models. *ACS Synth. Biol.*, **9**, 3145–3156.

52. Umenhoffer,K., Fehér,T., Balikó,G., Ayaydin,F., Pósfai,J., Blattner,F.R. and Pósfai,G. (2010) Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microb. Cell Fact.*, **9**, 38.

53. Pósfai,G., Plunkett,G., Fehér,T., Frisch,D., Keil,G.M., Umenhoffer,K., Kolisnychenko,V., Stahl,B., Sharma,S.S., de Arruda,M. *et al.* (2006) Emergent properties of reduced-genome *Escherichia coli. Science*, **312**, 1044–1046.

54. Reynolds,A.E., Felton,J. and Wright,A. (1981) Insertion of DNA activates the cryptic *bgl* operon in *E. coli* K12. *Nature*, **293**, 625–629.

55. Kitamura,K., Torii,Y., Matsuoka,C. and Yamamoto,K. (1995) DNA sequence changes in mutations in the *tonB* gene on the chromosome of *Escherichia coli* K12: insertion elements dominate the spontaneous spectra. *Jpn. J. Genet.*, **70**, 35–46.

56. Lioy,V.S., Cournac,A., Marbouty,M., Duigou,S., Mozziconacci,J., Espéli,O., Boccard,F. and Koszul,R. (2018) Multiscale structuring of the *E. coli* chromosome by nucleoid-associated and condensin proteins. *Cell*, **172**, 771–783.

57. Jin,D.J. and Cabrera,J.E. (2006) Coupling the distribution of RNA polymerase to global gene regulation and the dynamic structure of the bacterial nucleoid in *Escherichia coli. J. Struct. Biol.*, **156**, 284–291.

58. Weng,X., Bohrer,C.H., Bettridge,K., Lagda,A.C., Cagliero,C., Jin,D.J. and Xiao,J. (2019) Spatial organization of RNA polymerase and its relationship with transcription in *Escherichia coli. Proc. Natl. Acad. Sci. U.S.A.*, **116**, 20115–20123.

59. Berger,M., Farcas,A., Geertz,M., Zhelyazkova,P., Brix,K., Travers,A. and Muskhelishvili,G. (2010) Coordination of genomic structure and transcription by the main bacterial nucleoid-associated protein HU. *EMBO Rep.*, **11**, 59–64.

60. Gerganova,V., Berger,M., Zaldastanishvili,E., Sobetzko,P., Lafon,C., Mourez,M., Travers,A. and Muskhelishvili,G. (2015) Chromosomal position shift of a regulatory gene alters the bacterial phenotype. *Nucleic. Acids. Res.*, **43**, 8215–8226.

61. Nagy-Staron,A., Tomasek,K., Caruso Carter,C., Sonnleitner,E., Kavčič,B., Paixão,T. and Guet,C.C. (2021) Local genetic context shapes the function of a gene regulatory network. *Elife*, **10**, e65993.

62. Tas,H., Grozinger,L., Stoof,R., de Lorenzo,V. and Goñi-Moreno,Á. (2021) Contextual dependencies expand the re-usability of genetic inverters. *Nat. Commun.*, **12**, 355.

63. Bayer,C.N., Rennig,M., Ehrmann,A.K. and Nørholm,M.H.H. (2021) A standardized genome architecture for bacterial synthetic biology (SEGA). *Nat. Commun.*, **12**, 5876.

64. Kroner,G.M., Wolfe,M.B. and Freddolino,P.L. (2019) *Escherichia coli* Lrp regulates one-third of the genome via direct, cooperative, and indirect routes. *J. Bacteriol.*, **201**, e00411-18.

65. Gama-Castro,S., Salgado,H., Santos-Zavaleta,A., Ledezma-Tejeida,D., Muñiz-Rascado,L., García-Sotelo,J.S., Alquicira-Hernández,K., Martínez-Flores,I., Pannier,L., Castro-Mondragón,J.A. *et al.* (2016) RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic. Acids. Res.*, **44**, D133–D43.