Supplemental text to accompany Wong et al. “The genome-wide transcriptional response to varying RpoS levels in *Escherichia coli* K-12”

Determination of most stably expressed control genes for QPCR

Gene expression differences identified via quantitative PCR require control genes that do not respond to the experimental conditions. In this case, we need genes that do not respond to *rpoS* levels. For this criterion, we chose genes that were not significantly differentially expressed between RpoS levels, and that showed a low Pearson's correlation coefficient between RpoS level and expression level in our RNAseq data set. Because we were initially interested in the interaction of DNA supercoiling levels and RpoS levels in influencing gene expression, we also chose genes that had a low correlation between gene expression and supercoiling levels (1). From this set of genes that were insensitive to RpoS and supercoiling levels, we selected 10 candidate genes that had wide range of mean expression levels in our RNAseq data set (Table S3).

qPCR primers for each of the 10 genes were designed with melting temperatures between 58-61 °C and amplicon lengths of 100-150 bp. Primer concentrations were optimized by selecting primer concentrations giving the greatest band intensity with all other PCR conditions being the same. Three different concentrations (100, 300, and 500 nM) were used for each primer, giving a total of nine different primer conditions. PCR products from all nine conditions were run on a gel, and the product giving the most intense band had the optimal primer concentrations. Because the intermediate concentration, 300 nM of each primer, was either the most intense or nearly the most intense for each trial, this concentration was selected for all future experiments.

A circular problem arises when trying to validate control genes because it is difficult to ensure that a gene is consistently expressed without the use of a control gene to normalize expression to. For this reason, Vandesompele et al. (2) devised a method that determines which genes out of a set of candidate genes are most consistently expressed. Briefly, this method involves using ratios of expression between two genes across samples to determine consistency of expression and reduce experimental noise. For example, if one gene is consistently expressed twice as much as another, even if the absolute values of expression changes (due to experimental variation), we would determine that these genes were consistently expressed. By pairwise comparing ratios of expression of one candidate gene to all the others across samples, we can determine which genes are most consistently expressed.

To measure which potential control genes were in fact most stably expressed, we used RT-qPCR to measure expression from 4 samples each from cultures grown with 0%, 26%, 89% and 100% of wild-type RpoS levels. A standard curve for each gene was made using the following amounts of gDNA: 200ng, 20ng, 2ng, 200pg, 20pg, and 2pg.

We then used the method of Vandesompele et al. (2) to select the most stably expressed genes. We calculated the stability measure M for each gene. By ranking genes by their M value, we determined which gene is least stable. We then removed this gene and repeated the analysis with the remaining genes. By repeatedly eliminating the least stable gene and recalculating the M values, we ultimately ended up with only

two most stable genes (Figure S1). It is important to note that the elimination of the least stable gene can in fact impact the gene rankings in the subsequent rounds: for example, *hemL* is ranked 2nd, 3rd, and 4th in different rounds.

After determining which genes are most consistently expressed, it is important to also determine how many genes are needed to ensure reliable quantification. Vandesompele et al. (2) suggests creating a normalization factor (NF*n*) to normalize qPCR results to by using the geometric mean of expression of the *n* most stable genes. Furthermore, they suggest starting with the three most stable control genes and repeatedly adding the next most stable gene until the normalization factors are not significantly altered, indicating that the additional gene would not provide significantly different results. Fortunately, it appears that the use of three control genes is sufficient as there is a high correlation (Spearman's rho = 0.97) between the normalization factor for three genes (NF3) and the normalization factor for four genes (NF4).

We conclude that the three most stably expressed genes across RpoS levels were *ftsZ, pgm,* and *hemL*. These were used as control genes for all subsequent experiments.

Literature cited

1. **Peter BJ**, **Arsuaga J**, **Breier AM**, **Khodursky AB**, **Brown PO**, **Cozzarelli NR**. 2004. Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. Genome Biol **5**:R87.

2. **Vandesompele J**, **Preter KD**, **Pattyn F**, **Poppe B**, **Roy NV**, **Paepe AD**, **Speleman F**. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol **3**:research0034.

3. **Haldimann A**, **Wanner BL**. 2001. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J Bacteriol **183**:6384–93.

4. **Khlebnikov A**, **Datsenko KA**, **Skaug T**, **Wanner BL**, **Keasling JD**. 2001. Homogeneous expression of the PBAD promoter in Escherichia coli by constitutive

expression of the low-affinity high-capacity AraE transporter. Microbiology **147**:3241–3247.

5. **Baba T**, **Ara T**, **Hasegawa M**, **Takai Y**, **Okumura Y**, **Baba M**, **Datsenko KA**, **Tomita M**, **Wanner BL**, **Mori H**. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol **2**:2006.0008.

6. **Butland G**, **Peregrín-Alvarez JM**, **Li J**, **Yang W**, **Yang X**, **Canadien V**, **Starostine A**, **Richards D**, **Beattie B**, **Krogan N**, **Davey M**, **Parkinson J**, **Greenblatt J**, **Emili A**. 2005. Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature **433**:531–537.

7. **Datsenko KA**, **Wanner BL**. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A **97**:6640–6645.

8. **Stoebel DM**, **Hokamp K**, **Last MS**, **Dorman CJ**. 2009. Compensatory evolution of gene regulation in response to stress by Escherichia coli lacking RpoS. PLoS Genet **5**:e1000671.

9. **Edwards AN**, **Patterson-Fortin LM**, **Vakulskas CA**, **Mercante JW**, **Potrykus K**, **Vinella D**, **Camacho MI**, **Fields JA**, **Thompson SA**, **Georgellis D**, **Cashel M**, **Babitzke P**, **Romeo T**. 2011. Circuitry Linking the Csr and Stringent Response Global Regulatory Systems. Mol Microbiol **80**:1561–1580.

Table S1 Strains and plasmids used in this study

Strain	Genotype	Source	Citation
Strains			
BW23473	F-, Δ(argF-lac)169, ΔuidA3::pir+, recA1, rpoS396(Am)?, endA9(del-ins)::FRT, rph-1, hsdR514, rob-1, creC510	CGSC	(3)
BW27786	Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), Δ(araH-araF)570(::FRT), ΔaraEp-532::FRT, φPcp13araE534, Δ(rhaD-rhaB)568, hsdR514	CGSC	(4)
JW5437	Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), ΔrpoS746::kan, Δ(rhaD-rhaB)568, hsdR514	CGSC	(5)
DMS2545	BW27786, with ΔrpoS746::kan from JW5437	This study	
DMS2564	BW27786 with ΔnlpD::kan-P_{araB}, so that RpoS is under the control of P_{araB}	This study	
DMS2686	*astA-lacZ* transcriptional fusion in pLFX, chromosomally integrated into DMS2564	This study	
DMS2687	*astA* core promoter-*lacZ* transcriptional fusion in pLFX, chromosomally integrated into DMS2564	This study	
DMS2689	*gadC-lacZ* transcriptional fusion in pLFX, chromosomally integrated into DMS2564	This study	
DMS2690	*gadC* core promoter-*lacZ* transcriptional fusion in pLFX, chromosomally integrated into DMS2564	This study	
RPB104	rpoS-SPA, in MG1655 background. Tagged allele moved from tagged DY330 strain by P1 transduction.	This study	(6)
Plasmids			
pKD46	Plasmid for recombination of PCR products; Amp^R, temperature-sensitive	CGSC	(7)
pDMS123	*otsB-gfp* fusion, Cm^R	Charles Dorman	(8)
pLFX	Vector for *lacZ* transcriptional fusions, integrates at **λ** attachment site; Amp^R	CGSC	(9)
pPFINT	Vector for integration; Tet^R, Amp^R, ts	CGSC	(9)
pAH150	Plasmid source of P_{araB} promoter	CGSC	(3)
pCP20	Encodes flp recombinase; ts, Amp^R, Cm^R		(7)

Table S2 Primers used in this study. Underlined bases indicate restriction enzyme recognition sites.

Primer name	Sequence
ParaBRpoSRecomb-F	CCTGGATTTTTCCTGGTTATTTTGCCGCAGGTCAGCGTATGGTATCGAACCCCAGAGTCC
ParaBRpoSRecomb-R	ATCTTCATTTAAATCATGAACTTTCAGCGTATTCTGACTCATATGGCTTGACTCCGTACA
astApromoter+	TTTGGTACCGTGCTTTTCGACGATGGCTT
astApromoter-	TTTGAATTCCTCTTTCCCCTGCTGATCCC
gadCpromoter+	AAAGGTACCGGTAGAGAATGTTAACACTGCCG
gadCpromoter-	AAAGAATTCCGTCATTGATAATCTGGAATGCGA
astA-35to+1for	TTTGGTACCCTGGCTGGCACGAACCCTGCAATCTACATTTACAGCGGAATTCTTT
astA-35to+1rev	AAAGAATTCCGCTGTAAATGTAGATTGCAGGGTTCGTGCCAGCCAGGGTACCAAA
gadC-35to+1for	CCCGGTACCCTTGCTTACTTTATCGATAAATCCTACTTTTTTAATGCGAATTCCCC
gadC-35to+1rev	GGGGAATTCGCATTAAAAAAGTAGGATTTATCGATAAAGTAAGCAAGGGTACCGGG
hdeAHiFi+	TCTAGAGGATCCCCGGGTACACAGCAGCACGGCAATAA
hdeAHiFi-	TCGGTTGTCGGATCCCCGGGCTGGCAGAAGAAGCAGAC
gadAHiFi+	TCTAGAGGATCCCCGGGTACCTGAACGGGGTGTACACG
gadAHiFi-	TCGGTTGTCGGATCCCCGGGCGGAAATCGTTTTGACTC
hdeAcoreHiFi	TCGACTCTAGAGGATCCCCGGGTACTGACATATACAGAAAACCAGGTTATAACCTCAGTGAATTCCCGGGGATCCGACAACCGAT
gadAcoreHiFi	TCGACTCTAGAGGATCCCCGGGTACGCCTTGCTTCCATTGCGGATAAATCCTACTTTTTTATTGCAATTCCCGGGGATCCGACAACCGAT

Table S3. Candidate control genes, primer sequences, and information used in the selection of the genes.

Gene	Primer sequences	Supercoiling correlation	RpoS correlation	Mean expression across RpoS levels (AU)
gatY	AATGCGAGGCGTCAATCATG AAAATCTGTTGGCGCTGGTC	0.14	0.15	7917
talB	AAACAGCAGAGCAACGATCG AAGACGCGCATCAACTTCAG	0.05	0.15	215
ynaJ	TCGGCGGCGTTATAGAACAG GCCAGCAACACAGTAAAGACC	0.03	0.19	487
crp	TGCGTTTGTCTGCACAGATG TAGCGTCTGGTTGTTTTGCC	0.10	0.18	4008
thrS	TTGGTATGCGGCCAAAGTTG AAAACGACGCACAACTGTCG	0.26	0.14	521
ftsZ	TGCATTTGCTTCCGACAACG ACGTTTGTCCATGCCGATAC	0.003	0.18	399
pgm	TGATTAACGTCGCCCAACTG AATGTGCGGCTCGTTAAAGC	0.06	0.18	59
speD	ACGTTCACATCCACCATGTC TTTACCCGCGACATTAACGG	0.02	0.08	168
hemL	TTCACCATGCGCACCATATC AATCCGCAATGCCGTGATTG	0.01	0.19	85
tehB	TGTGGCAATGGTCGTAACAG TTTAATGCGCTCGACGTTGG	0.06	0.07	145

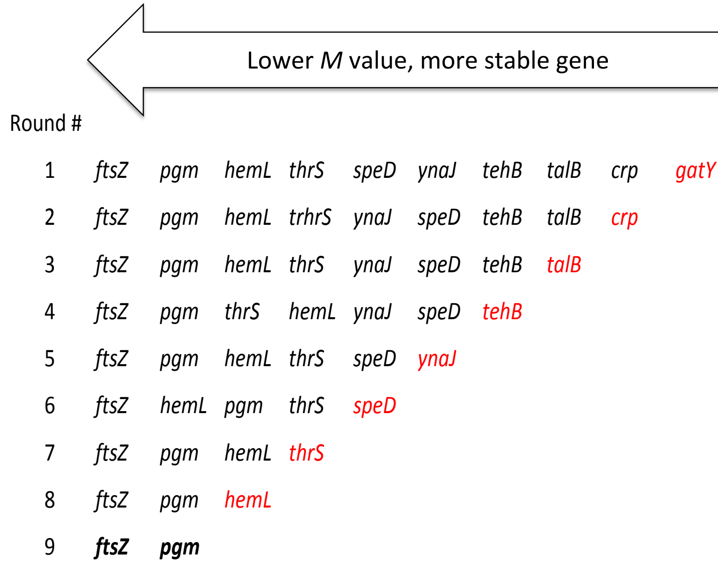

Figure S1. A ranking of the ten candidate control genes. In each round, the genes were ranked with the most stable (lowest *M* value) gene on the left and the least stable gene (highest *M* value) in red. After each round, the least stable gene was removed and the calculation and ranking by *M* value repeated. From this, *ftsZ* and *pgm* are the two most stable genes.

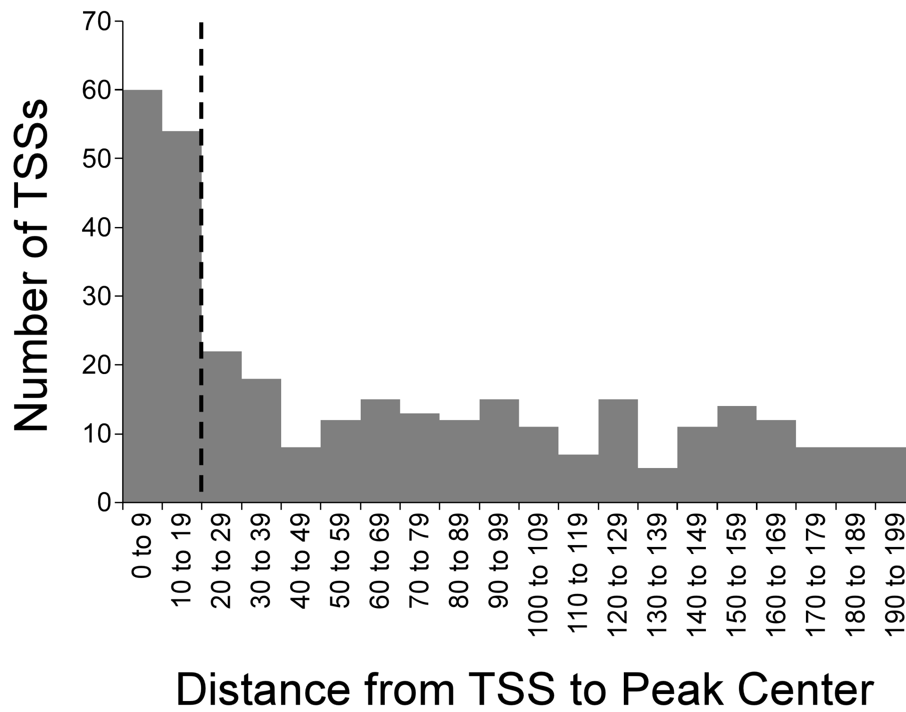

Figure S2. Histogram showing the frequency distribution of distances between ChIP-seq peak centers and TSSs identified during stationary phase growth in minimal medium. The black, vertical, dashed line indicates the cut-off used for calling TSSs as being associated with an RpoS ChIP-seq peak.