Science Advances
AAAS

Supplementary Materials for

Single-cell analysis of multiple invertible promoters reveals differential inversion rates as a strong determinant of bacterial population heterogeneity

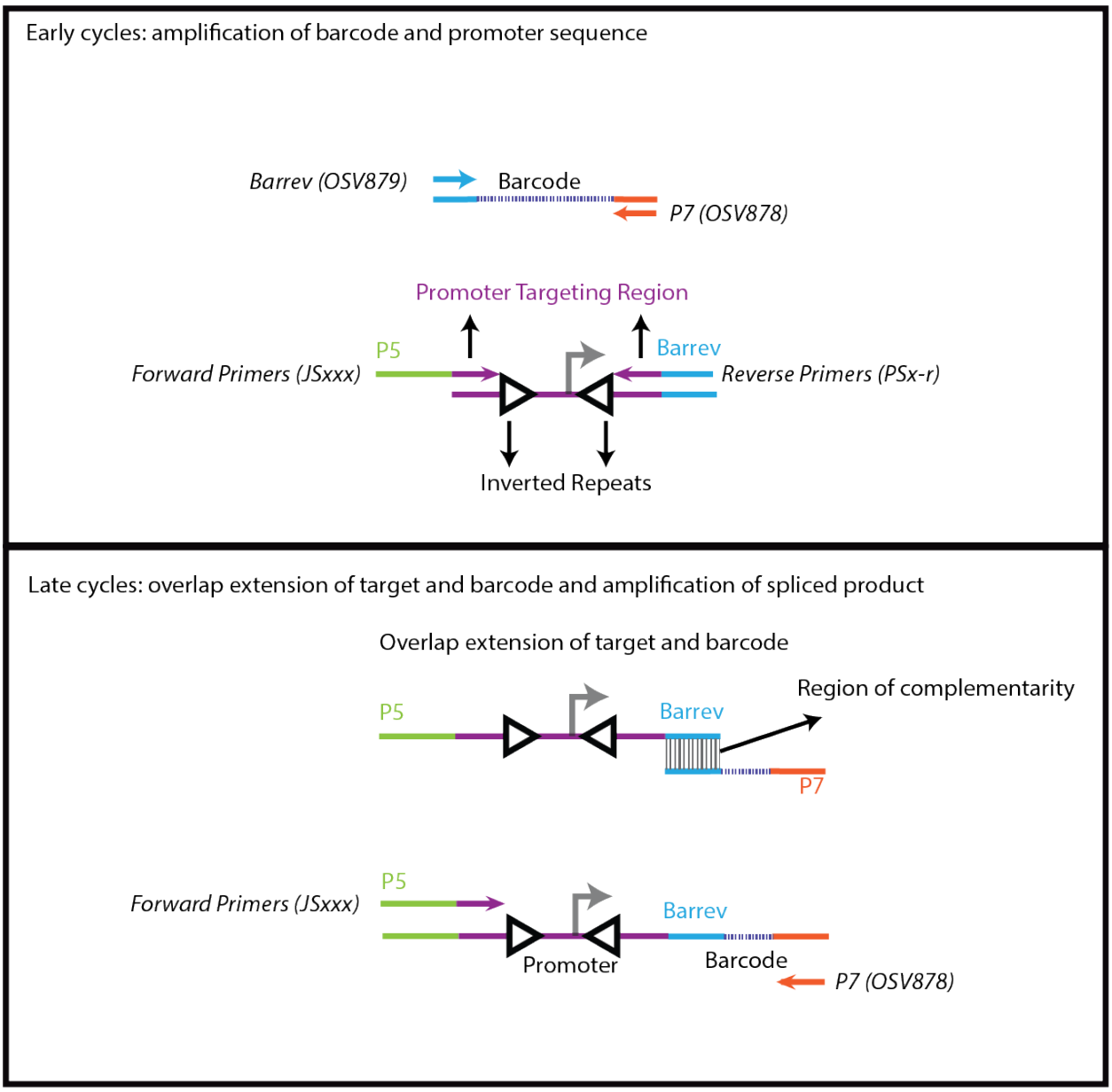
Freeman Lan *et al.*

Corresponding author: Ophelia S Venturelli, venturelli@wisc.edu; Robert Landick, landick@wisc.edu

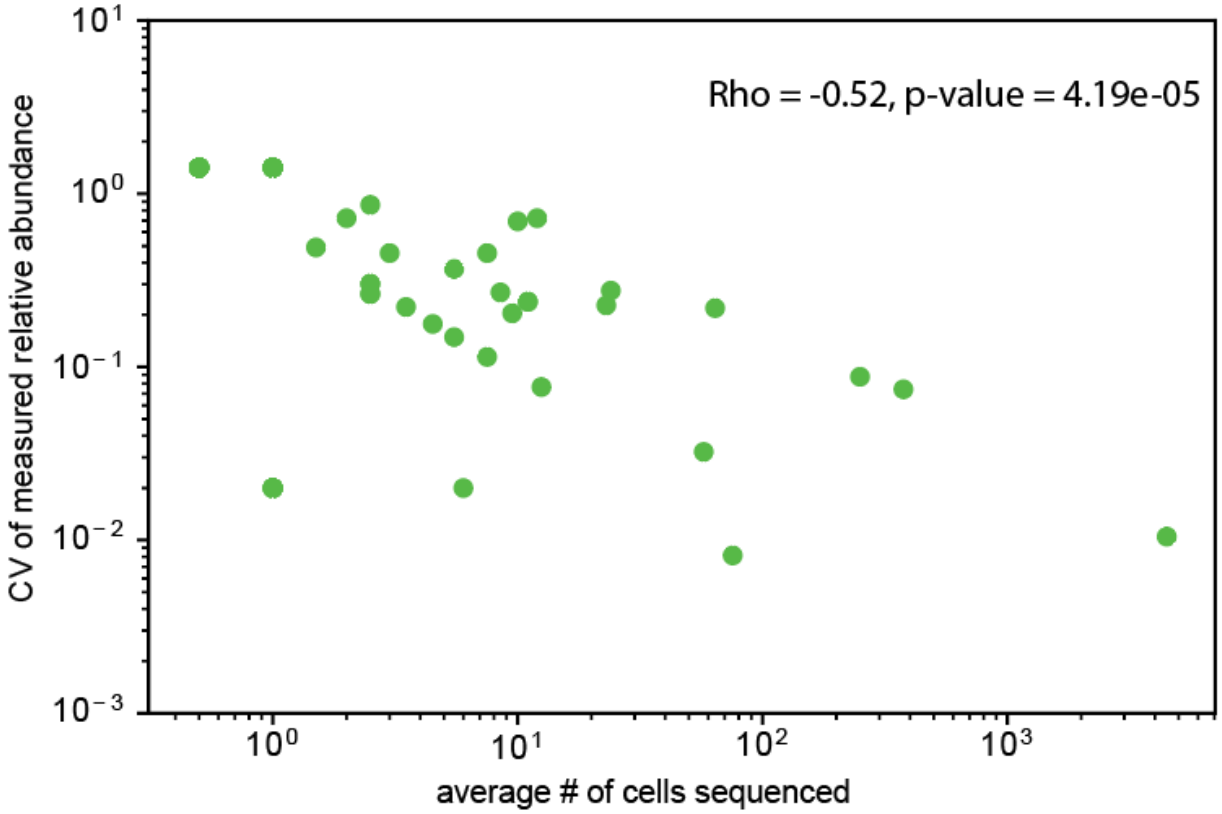
Sci. Adv. **9**, eadg5476 (2023)
DOI: 10.1126/sciadv.adg5476

The PDF file includes:

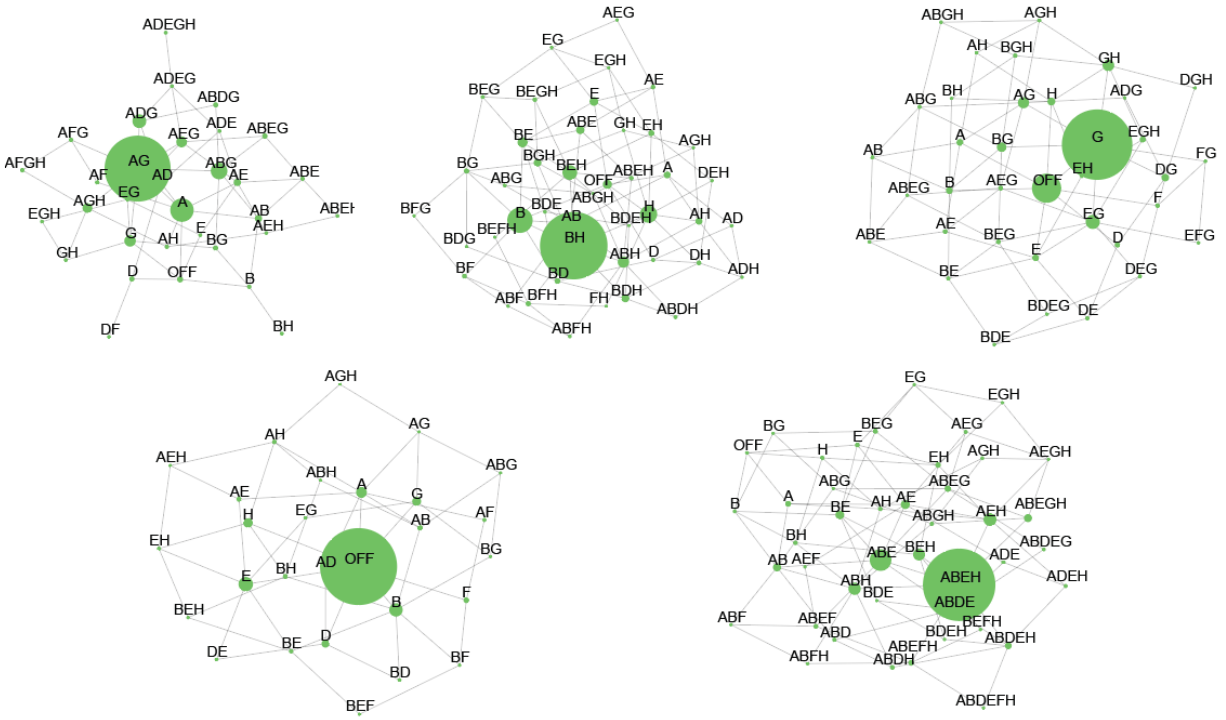
Other Supplementary Material for this manuscript includes the following:



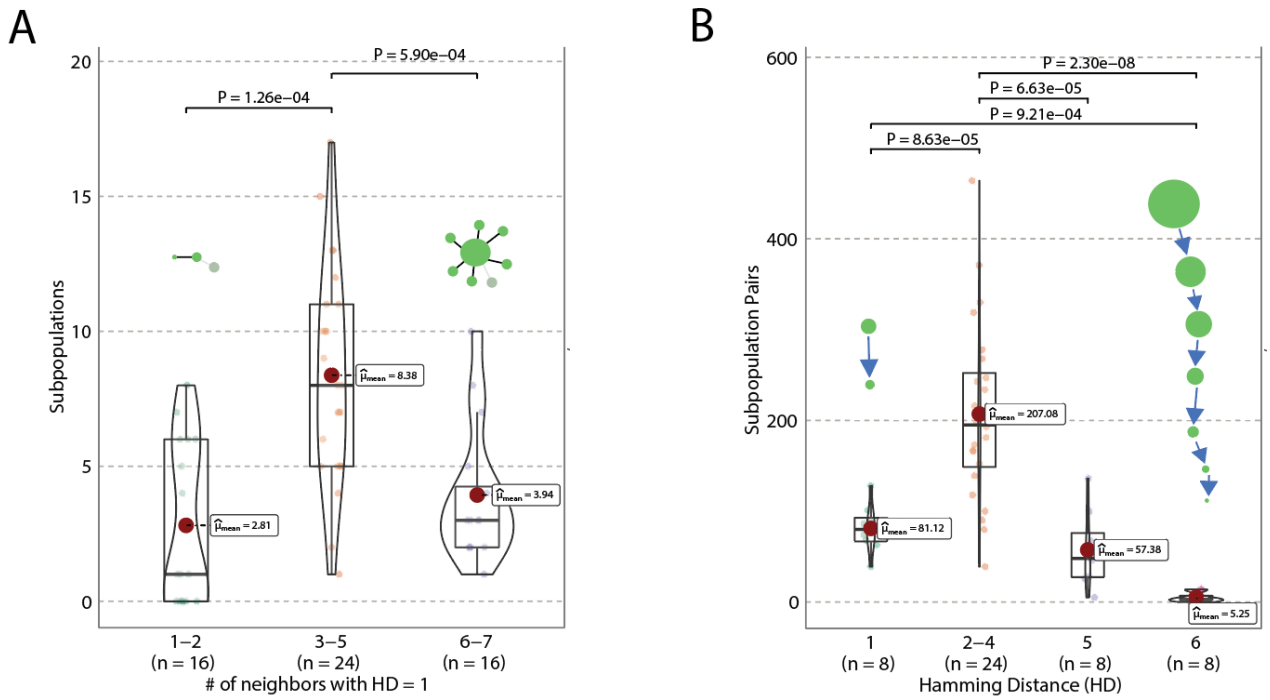
Supplementary Figure 1. Detailed description of the PCR reaction happening inside each droplet. PCR primers are labelled in italics with oligo number in parentheses (corresponding to supplementary table 4). Specific DNA regions are labelled in regular font and color coded. Upper: During the early cycles of PCR, PCR primers amplify both the unique DNA barcode and promoter regions of the genome. Lower: During the later cycles of PCR, the promoter and barcode amplicons overlap through a region of complementarity to produce single fused amplicon, which are further amplified by the PCR primers.



Supplementary Figure 2. Coefficient of variation (CV) of the measured relative abundance is negatively correlated with the number of cells sequenced for each subpopulation. Rho represents the Spearman's correlation coefficient for the dataset.

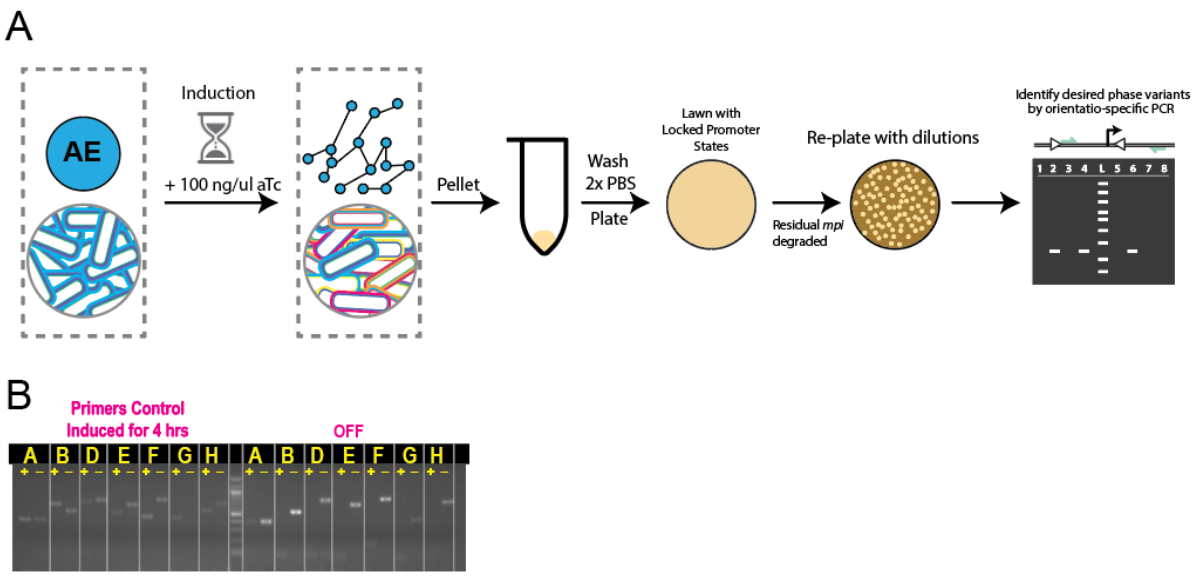


Supplementary Figure 3. Network graphs representing the population structures of different colonies grown, as determined by DoTA-seq. Each node represents a different subpopulation with the label indicating the combination of promoters turned ON. The “OFF” label represents subpopulations with all invertible promoters turned OFF. Node sizes represent the relative population abundances and edges connect nodes that are one promoter flip away.

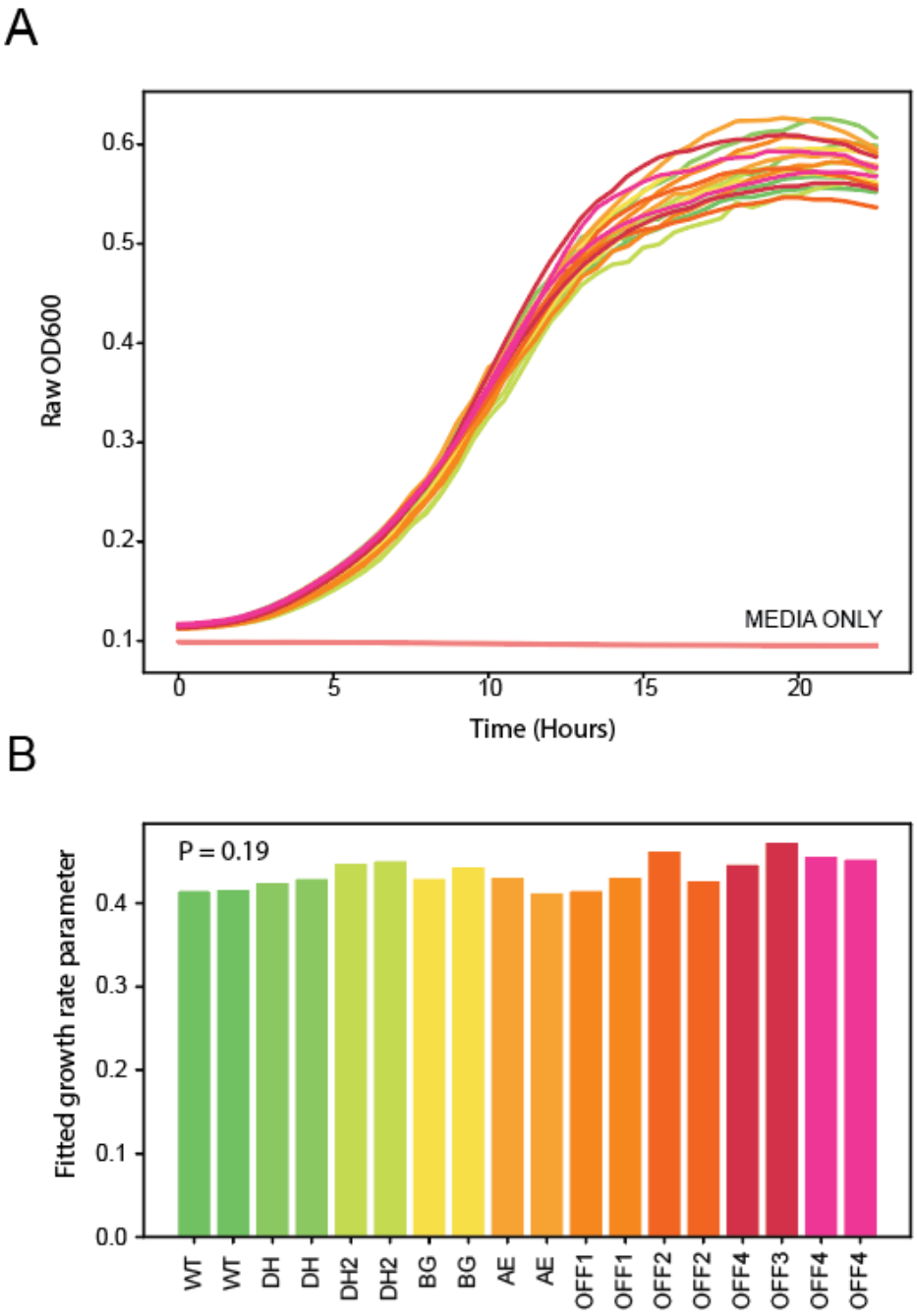


Supplementary Figure 4. Summary statistics for all colonies.

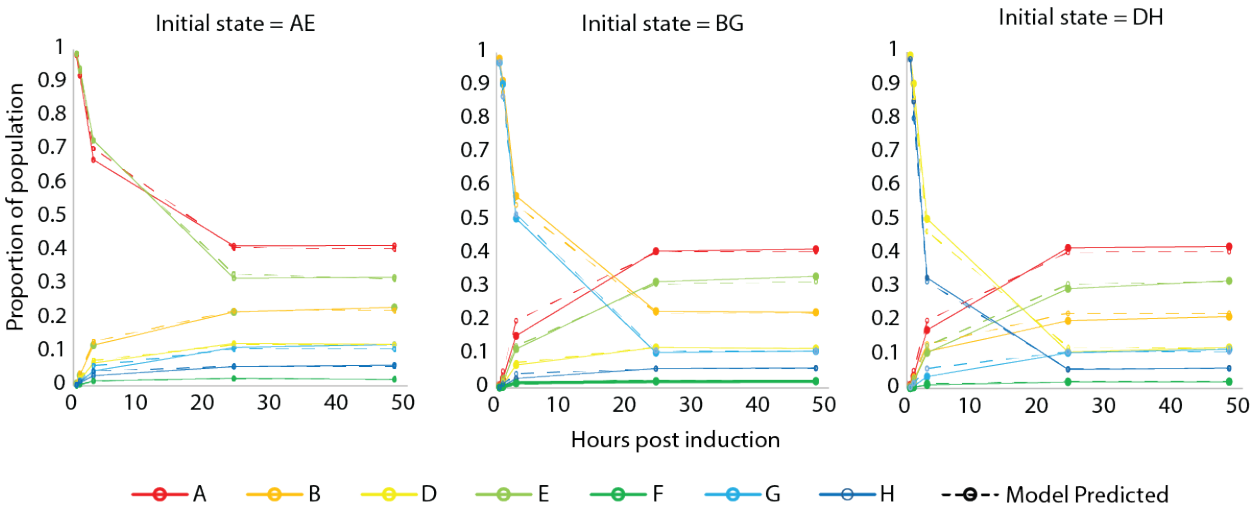
(a) Frequency distributions as violin plot of immediate (Hamming Distance (HD) = 1) neighbor count for all subpopulations within their respective colonies. A boxplot is overlayed and the mean highlighted as a red dot. Welch's ANOVA was performed and supplemented by post-hoc pairwise Games-Howell tests. Holm-Bonferroni-corrected p-values are reported for pairwise comparisons. (b) Frequency distribution as violin plot of HDs for all possible pairs of subpopulations within their respective colonies. A boxplot is overlayed and the mean highlighted as a red dot. Welch's ANOVA was performed and supplemented by post-hoc pairwise Games-Howell tests. Holm-Bonferroni-corrected p-values are reported for pairwise comparisons. Horizontal brackets represent statistical significant pairwise comparisons.



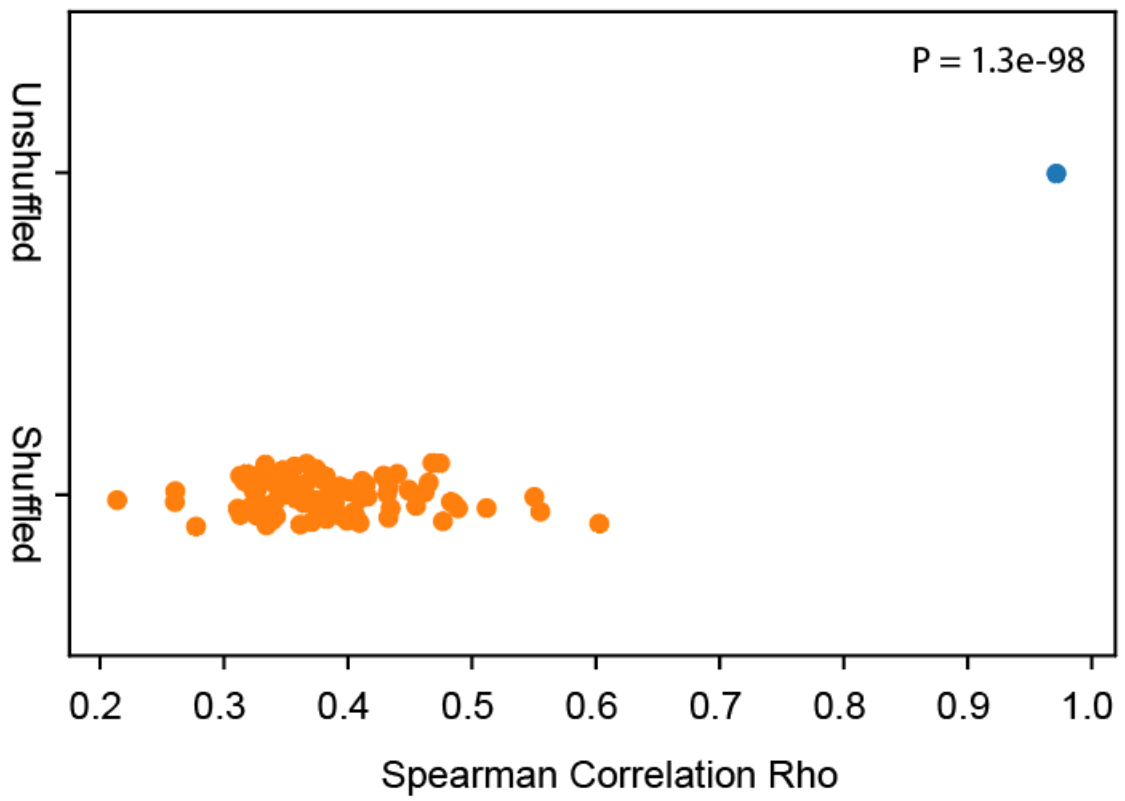
Supplementary Figure 5. Workflow for rapidly isolating and characterizing phase-locked variants. (a) Schematic of workflow for generating different phase-locked variants. Starting with an inducible invertase strain, after 4 hours of induction, a population with diverse combinations of promoter orientations was generated. Cells were washed twice to remove residual amounts of inducer, then plated for overnight growth and gradual dilution and degradation of residual invertase. The population is re-streaked to new plates to isolate single colonies largely fixed in promoter orientations. (b) Representative results of orientation-specific PCR using previously published primers (Coyne et al., 2003), which is used to identify the promoter orientations of different locked colonies. The '+' signifies a lane in which PCR was done using primers to detect an on-oriented promoter, while conversely the '-' indicates a lane in which PCR was done using primers to detect an off-oriented promoter. The extreme sensitivity of orientation-specific PCR suggests a small fraction of PSA on, and this sensitivity was noted previously with these published primer sets (Coyne et al 2003).



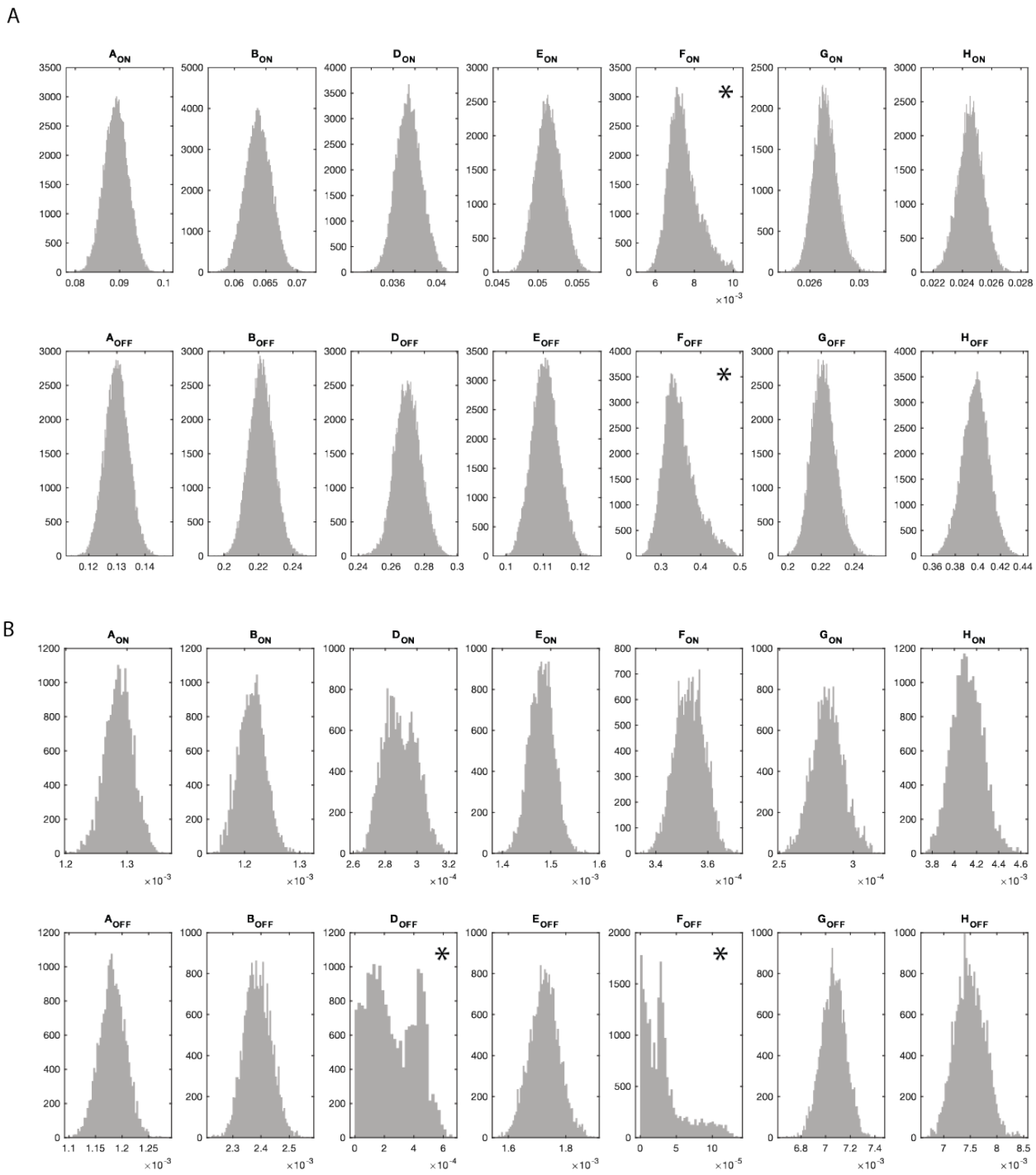
Supplementary Figure 6. Growth rates are similar between different CPS locked strains. (a) Raw OD600 growth curves of wildtype and different CPS locked strains grown in media. (b) Growth rate parameters of wildtype (WT) and strains locked in different CPS promoter orientations as determined by growth curve fitting to a logistic growth model. Each strain (unique label on the X axis) is grown in technical duplicates. One-factor ANOVA between the different promoter orientations shows no statistically significant difference (P-value 0.19) between the growth rates among the different CPS promoter orientations. Colors indicate different strains tested and are consistent across A and B.



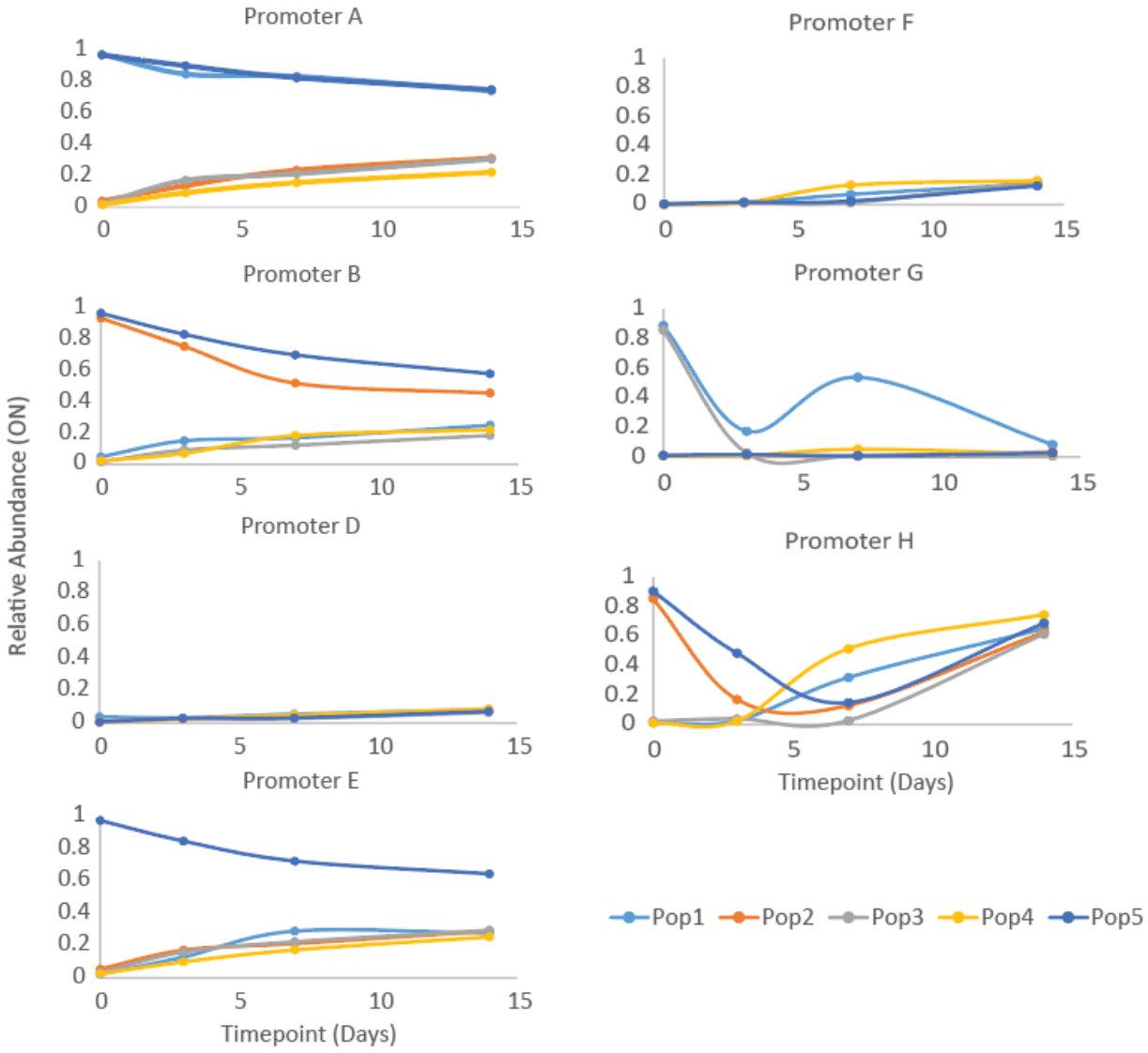
Supplementary Figure 7. Experimental (solid lines) and model simulated (dashed lines) data for population compositions of the synthetic strains post induction of *mpi*. Each color represents the proportion of a specific promoter turned on within the population. Each graph represents a different population starting with different initial states of promoters turned on.



Supplementary Figure 8. Spearman correlation between experimental data and model predictions for models generated using unshuffled and label shuffled experimental data from the induced invertase timeseries. Experimental data with unshuffled and 100 iterations of randomly shuffled labels are fit to the promoter inversion model using Markov-Chain Monte Carlo (see Methods). Spearman's correlation between the model predictions and the data used to fit the models is used as a measure of goodness-of-fit. Unshuffled data obtains significantly higher correlation (Rho = 0.98) compared to shuffled data controls (Rho = 0.38 +/- 0.06) showing that the model data agreement would not arise from random data.



Supplementary Figure 9. Parameter distributions obtained from Markov Chain Monte Carlo fitting of the stochastic inversion model on induced population timeseries data (a) and wildtype population timeseries data (b). Asterisk (*) indicates parameters where the coefficient of variation is larger than 5%, indicating possible insufficient constraint on those parameters.



Supplementary Figure 10. Diversification trajectories of each promoter within the 5 wild-type populations. Promoters A-E appear to converge at a slower rate than promoters F-H.

Supplementary Auxiliary Materials

Droplet-maker-mask.dwg: Autocad mask design for microfluidic droplet maker used in DoTA-seq.

Inversion-modelling.pdf: Detailed description of CME model for promoter inversion dynamics

Supplementary Table 1: Read filtering statistics and metadata for all DoTA-seq sequencing libraries.

Supplementary Table 2: Population compositions (obtained from DoTA-seq pipeline) for all populations used in this manuscript.

Supplementary Table 3: Parameter distributions and constraints from Markov-Chain Monte-Carlo fitting of the CME model.

Supplementary Table 4: Oligonucleotides used in DoTA-seq