eLIFE elifesciences.org

RESEARCH ARTICLE

Figures and figure supplements

Regulatory network structure determines patterns of intermolecular epistasis

Mato Lagator *et al*

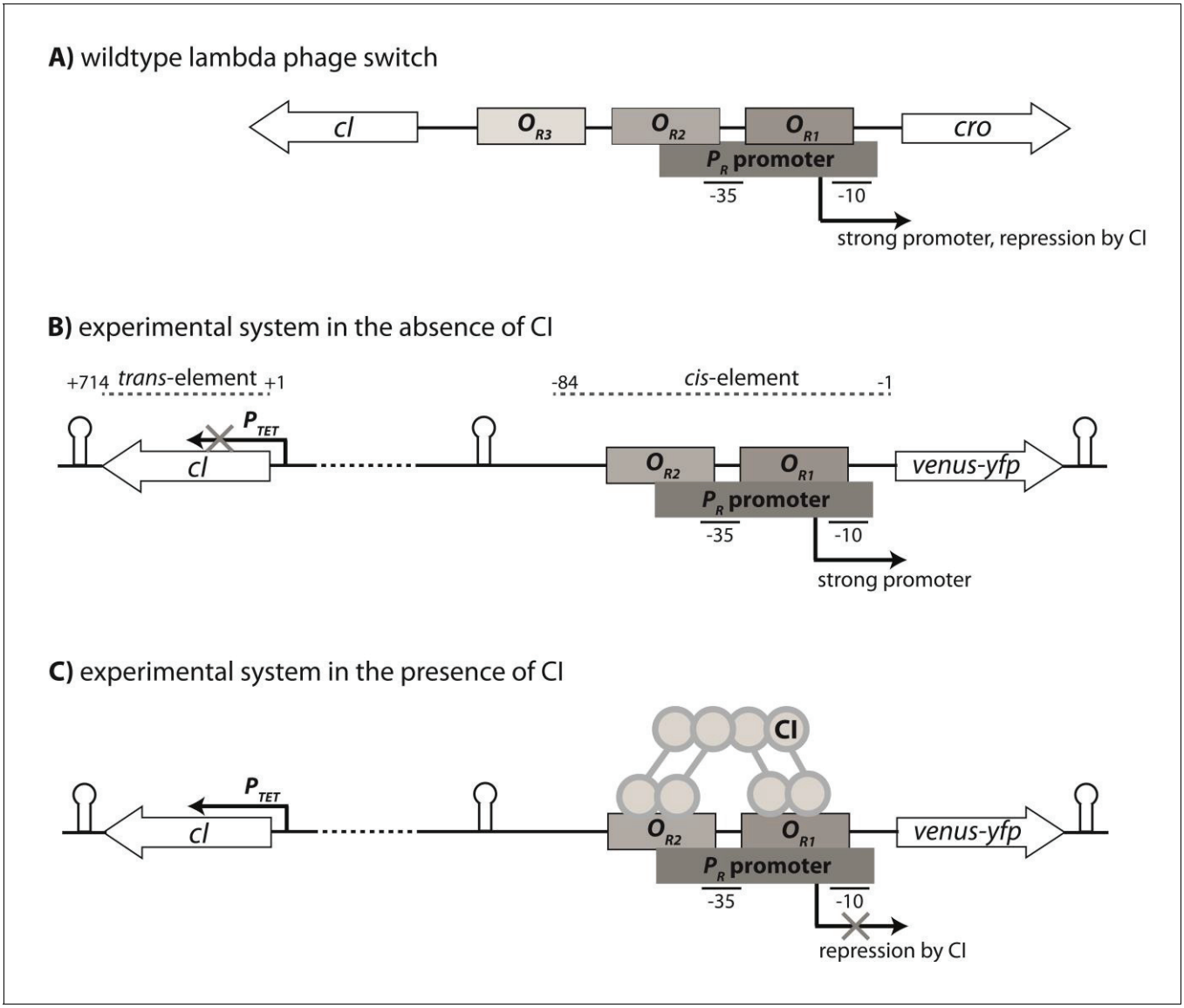

Figure 1. Genetic organization of the Lambda phage switch and the experimental system. (**A**) The Lambda phage switch consists of two transcription factors - CI and Cro; two promoters - a strong promoter P_R and a weak promoter P_{RM} (not shown); and three operator sites - O_{R1}, O_{R2}, and O_{R3}. (**B**) The experimental synthetic system, where *cro* was substituted with the fluorescence marker gene (*venus-yfp*) under control of P_R. The P_{RM} promoter was removed by the removal of parts of O_{R3}. Located 500 bp away on the reverse strand and separated by a terminator sequence is the *cI* gene under control of an inducible P_{TET} promoter. CI, the *trans*-element, is 714 bp; the *cis*-element is 84 bp long. (**C**) CI dimers bind cooperatively to O_{R1} and O_{R2}, leading to repression of the P_R promoter.

DOI: https://doi.org/10.7554/eLife.28921.002

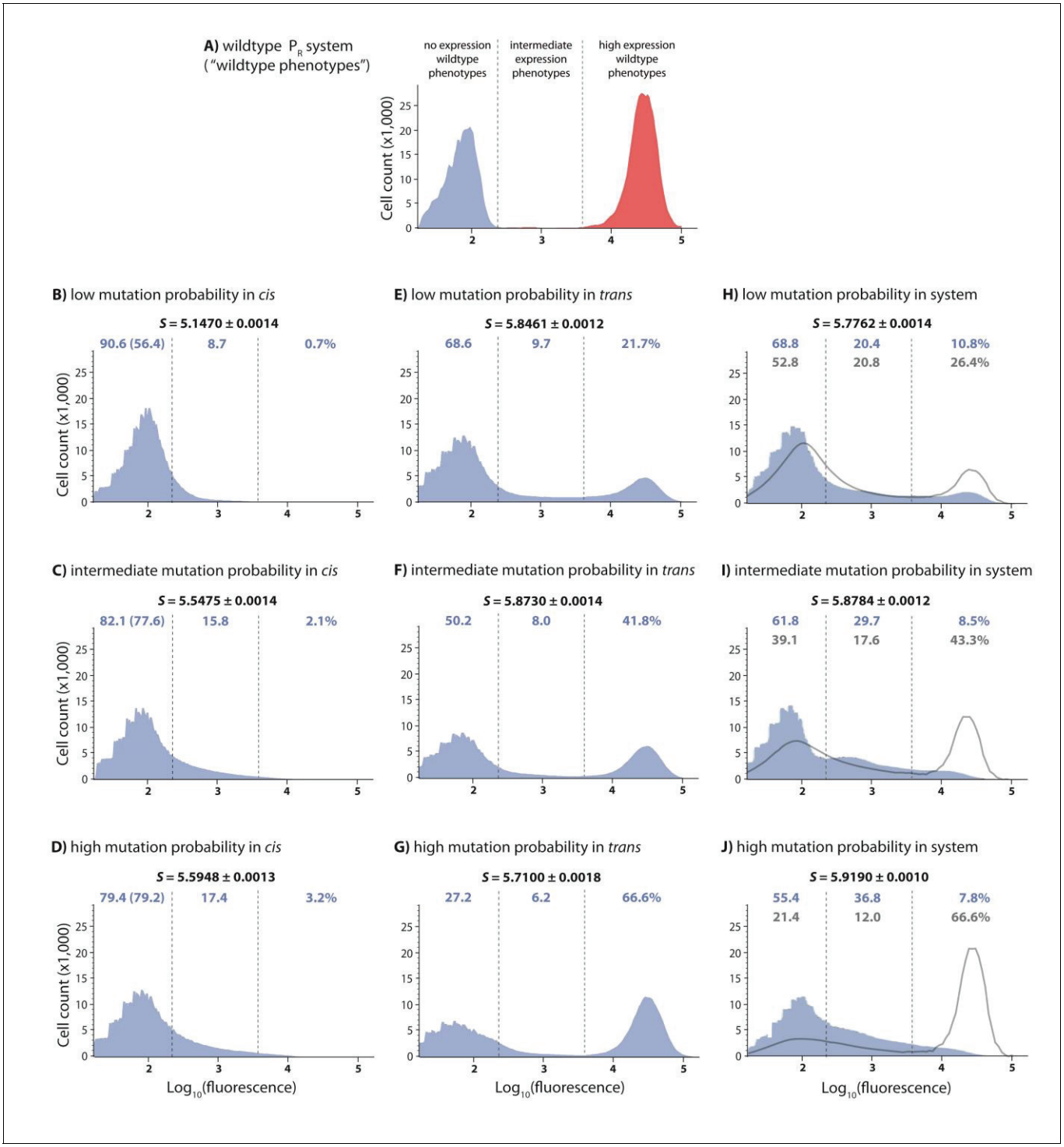

Figure 2. DMEs of the whole system are more evenly distributed than the individual component DMEs. In the experimental system, CI acts as a tight repressor. The distributions of fluorescence are shown in the absence of CI (red) and in the presence of CI (blue). Each distribution was obtained by measuring fluorescence of two independent measurements of 500,000 cells by flow cytometry, which were then pooled together. The dashed lines separate three categories of phenotypes – 'no expression' phenotypes (corresponding to repressed wildtype); 'high expression' phenotypes (corresponding to the wildtype in the absence of CI); and 'intermediate' phenotypes. No expression and high expression categories are defined to include >99.9% of the wildtype fluorescence distribution in the presence and in the absence of CI, respectively. The Shannon entropy (S) is used to estimate how uniform each distribution is across the entire range of possible expression levels. The associated standard deviation (±) is given for each *S* value. Blue numbers are percentage of counts in each category in the presence of CI. Numbers in parentheses are percentage of counts excluding the

Figure 2 continued on next page

Figure 2 continued

estimated percentages of uniquely transformed individuals carrying the wildtype genotype (see Materials and methods). The naïve additive convolution prediction for each system library and the associated predictions for the frequency of mutants in each category are shown in grey. Pearson's Chi-squared test was used to assess the difference between the observed and the convolution-predicted frequency of mutants in each category (low: $\chi^2_{(2)}$=8.20; p<0.05; intermediate: $\chi^2_{(2)}$=32.26; p<0.0001; and high mutation frequency library: $\chi^2_{(2)}$=74.51; p<0.0001). The distributions of the effects of mutations for the *cis*-element, the *trans*-element, and the whole system in the absence of CI are shown in ***Figure 2—figure supplement 1***. ***Figure 2—figure supplement 2*** shows distributions of the effects of 150 single point mutations in the *cis*- and the *trans*-elements. Statistical significance of the differences in entropy values between the mutant libraries is shown in ***Figure 2—source data 3***. Flow cytometry measurements of 20 individual isolates from each library are shown in ***Figure 2—figure supplements 3***, ***4*** and ***5***, the analysis of which was used to demonstrate that gene expression noise is constant (***Figure 2—source data 1***). Convolutions for each mutation probability performed with the knowledge of the genetic regulatory structure of the system are shown in ***Figure 2—figure supplement 6***, while ***Figure 2—figure supplement 7*** provides an explanation of how convolutions were performed. The outcome of the test for how sensitive the shapes of distributions are to the number of sampled individuals is shown in ***Figure 2—source data 4***, while the confirmation that the mutagenesis protocol resulted in expected distributions of the number of mutations are shown in ***Figure 2—source data 2***.

DOI: https://doi.org/10.7554/eLife.28921.003

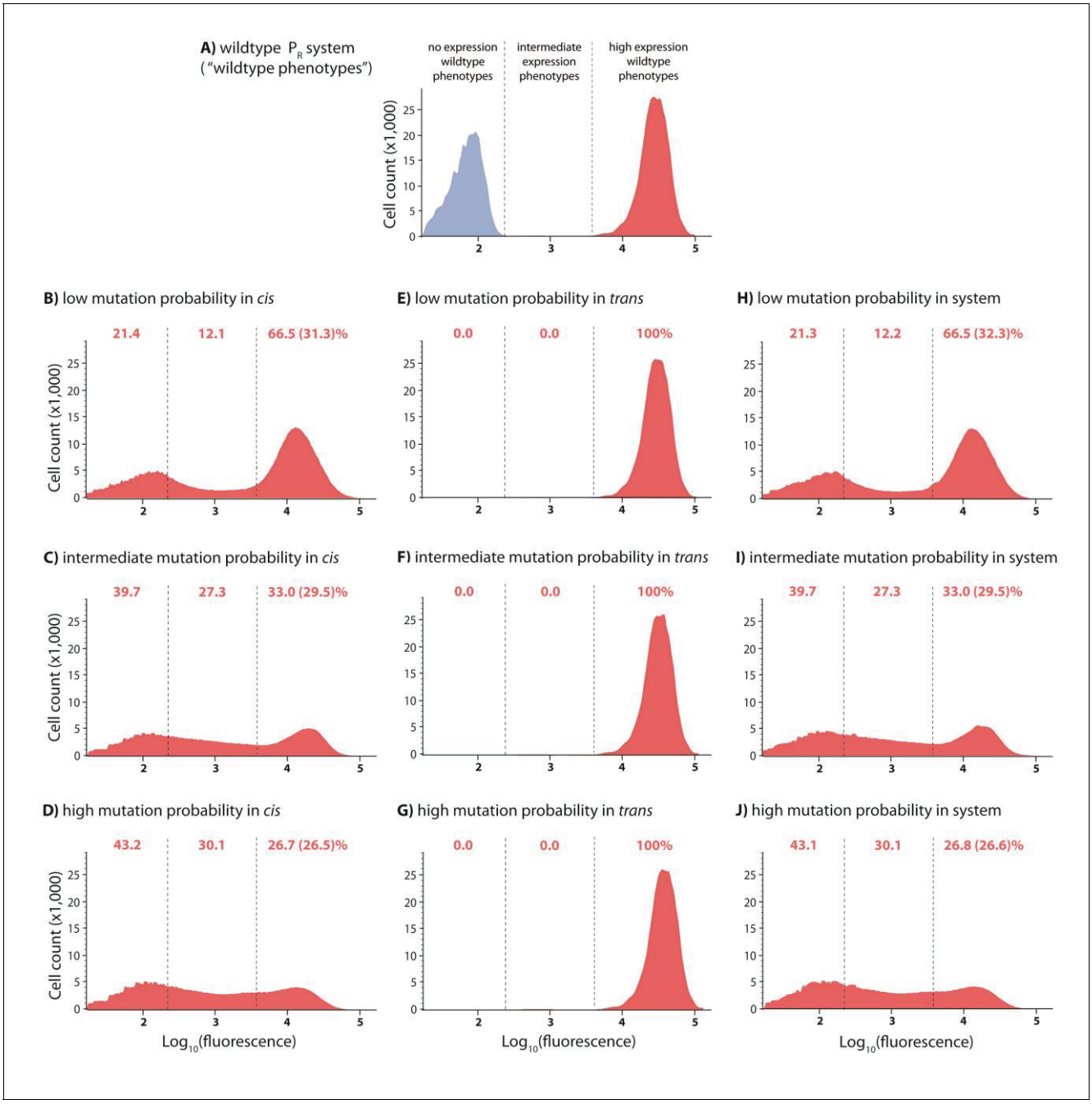

Figure 2—figure supplement 1. DMEs for *cis*-element, *trans*-element, and system libraries in the absence of CI. In the experimental system, CI acts as a tight repressor. Each distribution was obtained by measuring fluorescence of two independent measurements of 500,000 cells by flow cytometry, which were then pooled together. The dashed lines separate three categories of phenotypes. 'No expression' and 'high expression' categories are defined to include >99.9% of the wildtype fluorescence distribution in the presence and in the absence of CI, respectively. Red numbers are percentage of counts in each category in the absence of CI. Numbers in brackets are percentage of counts excluding the estimated percentage of uniquely transformed individuals carrying the wildtype genotype. DMEs of the *trans*-element library in the absence of CI are the same as the wildtype. DMEs of the *cis*-element and system libraries with equivalent mutation probability are not different from each other in the absence of CI.
DOI: https://doi.org/10.7554/eLife.28921.004

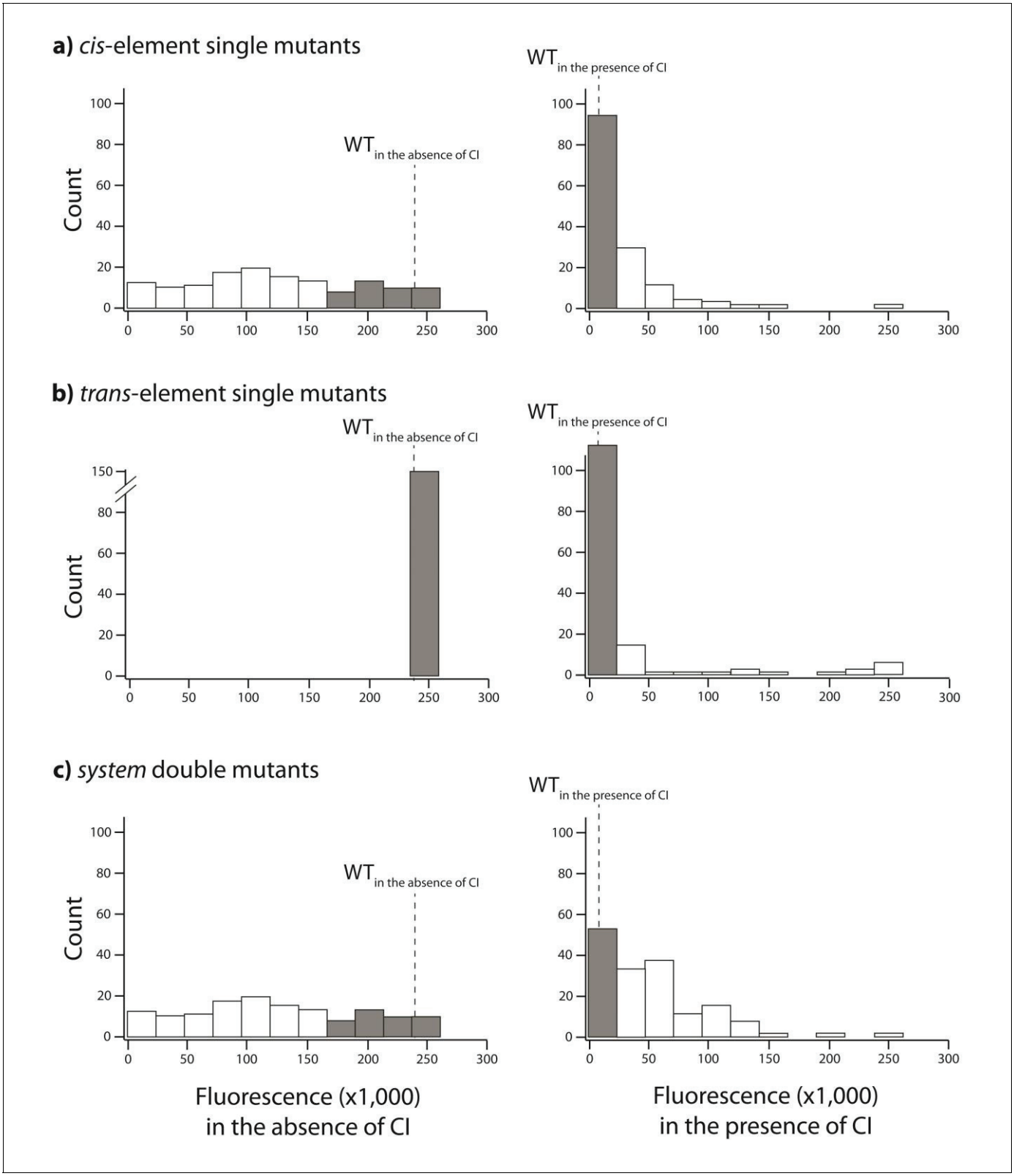

Figure 2—figure supplement 2. Distribution of single mutation effects in 150 random system double mutants and their corresponding single mutants. We created 150 random unique double mutants, with one point mutation in the *cis*- and the other in the *trans*-element. We measured gene expression of each mutant at a population level in a plate reader. Histogram of expression levels in the absence and in the presence of CI are shown for: (**A**) point mutations in *cis*; (**B**) point mutations in *trans*; (**C**) double mutants in the system. Dotted line represents mean wildtype fluorescence in the corresponding

Figure 2—figure supplement 2 continued on next page

Figure 2—figure supplement 2 continued

environment. Six replicates of each mutant were measured. Grey bars indicate mutants that were not significantly different from the wildtype. The data underlying this figure are shown in ***Figure 3—source data 2***. The data from this library are used to calculate epistasis shown in ***Figure 3***.

DOI: https://doi.org/10.7554/eLife.28921.005

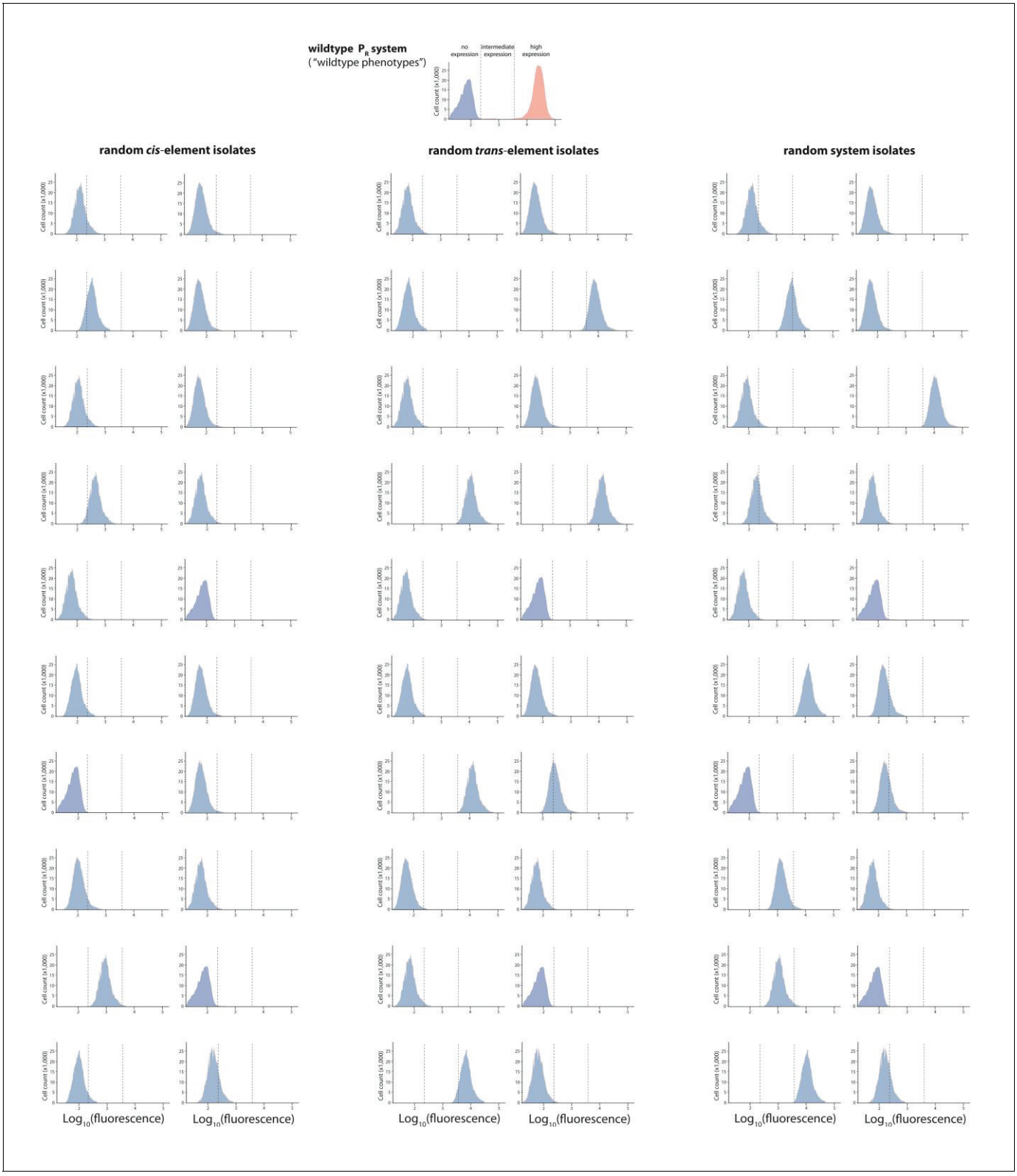

Figure 2—figure supplement 3. Mutant isolates from the low mutation probability libraries. Twenty mutants were arbitrarily isolated from the *cis*, the *trans*, and the system library, and the expression of two replicates of each isolate in the presence of CI was measured for 100,000 individuals in the flow cytometer. From these measurements, we calculated the gene expression noise for each isolate (***Figure 2—source data 1***). Each histogram shows the

Figure 2—figure supplement 3 continued on next page

Figure 2—figure supplement 3 continued

combined distribution based on two replicate measurements, for each mutant isolate. The top histogram shows the monoclonal wildtype distribution in the absence (red) and in the presence (blue) of CI.

DOI: https://doi.org/10.7554/eLife.28921.006

Figure 2—figure supplement 4. Mutant isolates from the intermediate mutation probability libraries. Twenty mutants were arbitrarily isolated from the *cis*, the *trans*, and the system library, and the expression of two replicates of each isolate in the presence of CI was measured for 100,000 individuals in the flow cytometer. From these measurements, we calculated the gene expression noise for each isolate (***Figure 2—source data 1***). Each histogram shows the combined distribution based on two replicate measurements, for each mutant isolate. The top histogram shows the monoclonal wildtype distribution in the absence (red) and in the presence (blue) of CI.

DOI: https://doi.org/10.7554/eLife.28921.007

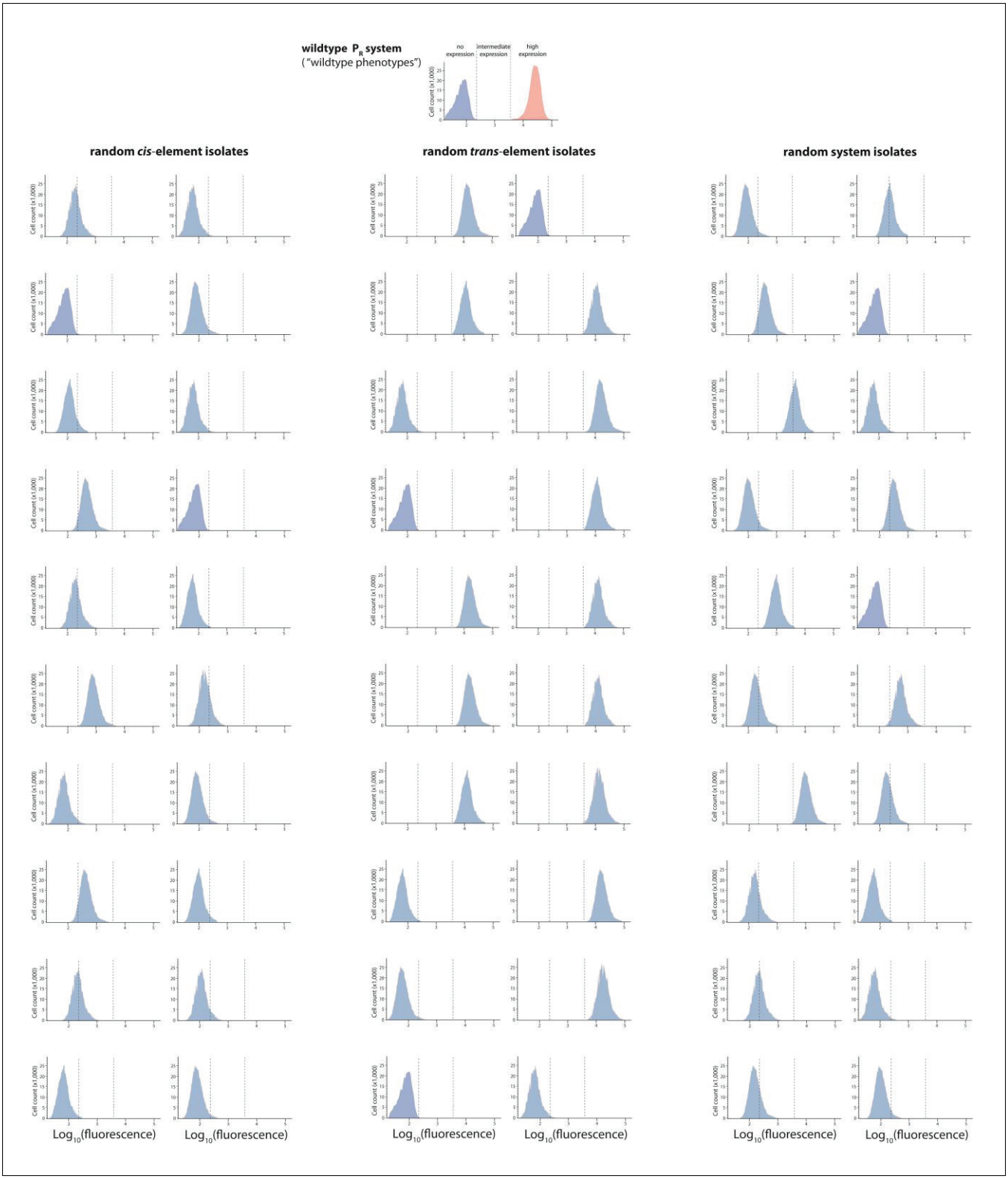

Figure 2—figure supplement 5. Mutant isolates from the high mutation probability libraries. Twenty mutants were arbitrarily isolated from the *cis*, the *trans*, and the system library, and the expression of two replicates of each isolate in the presence of CI was measured for 100,000 individuals in the flow cytometer. From these measurements, we calculated the gene expression noise for each isolate (***Figure 2—source data 1***). Each histogram shows the combined distribution based on two replicate measurements, for each mutant isolate. The top histogram shows the monoclonal wildtype distribution in the absence (red) and in the presence (blue) of CI.

DOI: https://doi.org/10.7554/eLife.28921.008

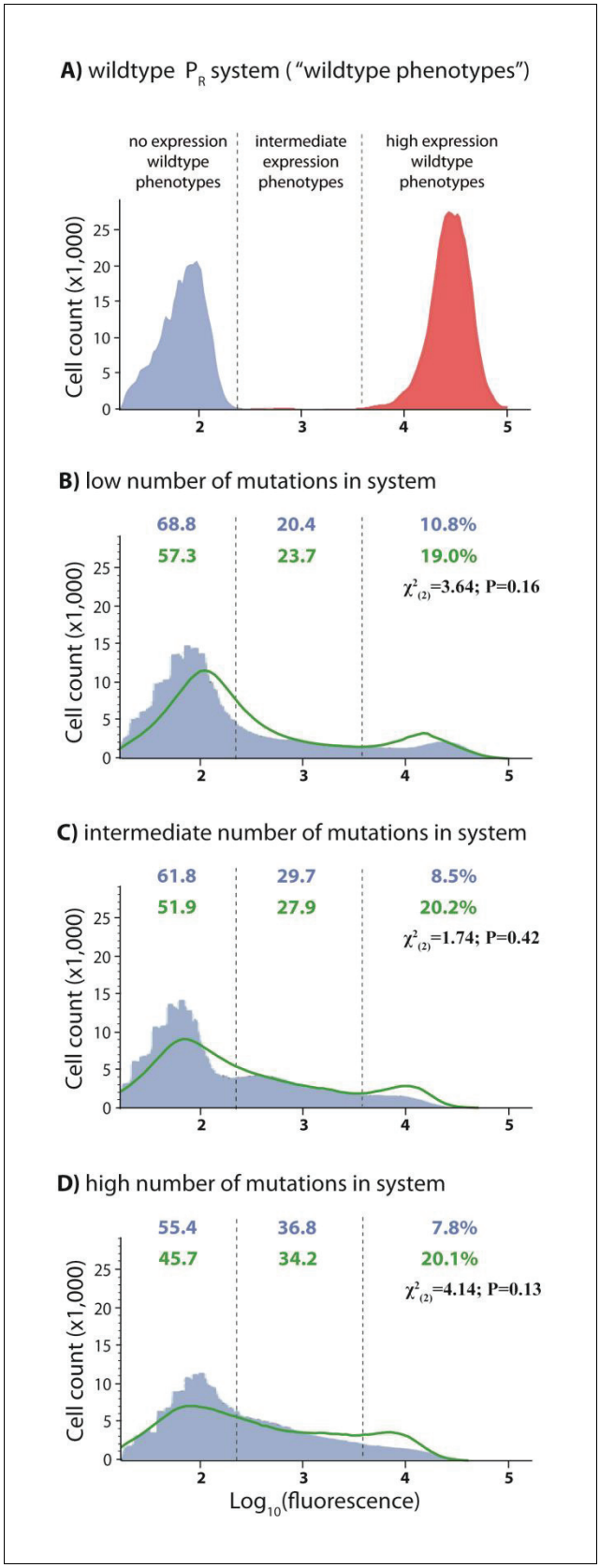

Figure 2—figure supplement 6. Mathematical predictions that account for the genetic regulatory structure accurately describe the system DME. (**A**) Distribution of the monoclonal wildtype population in the absence (red)

Figure 2—figure supplement 6 continued on next page

Figure 2—figure supplement 6 continued

and in the presence of CI (blue). (**B**) Low, (**C**) Intermediate, and (**D**) High mutation probability system mutant libraries. Shown in green are the convolutions that accounted for the effects of *cis* mutations in the absence of CI (***Figure 2—figure supplement 1***). Frequencies of mutants in the three categories ('no', 'intermediate', and 'high' expression phenotypes) for the observed system DME are shown in blue, while those from the convolution are shown in green. Pearson's Chi-squared statistic and the associated p value used to test for the differences between the observed and the predicted frequencies of mutants in the three categories are shown for each mutations probability.

DOI: https://doi.org/10.7554/eLife.28921.009

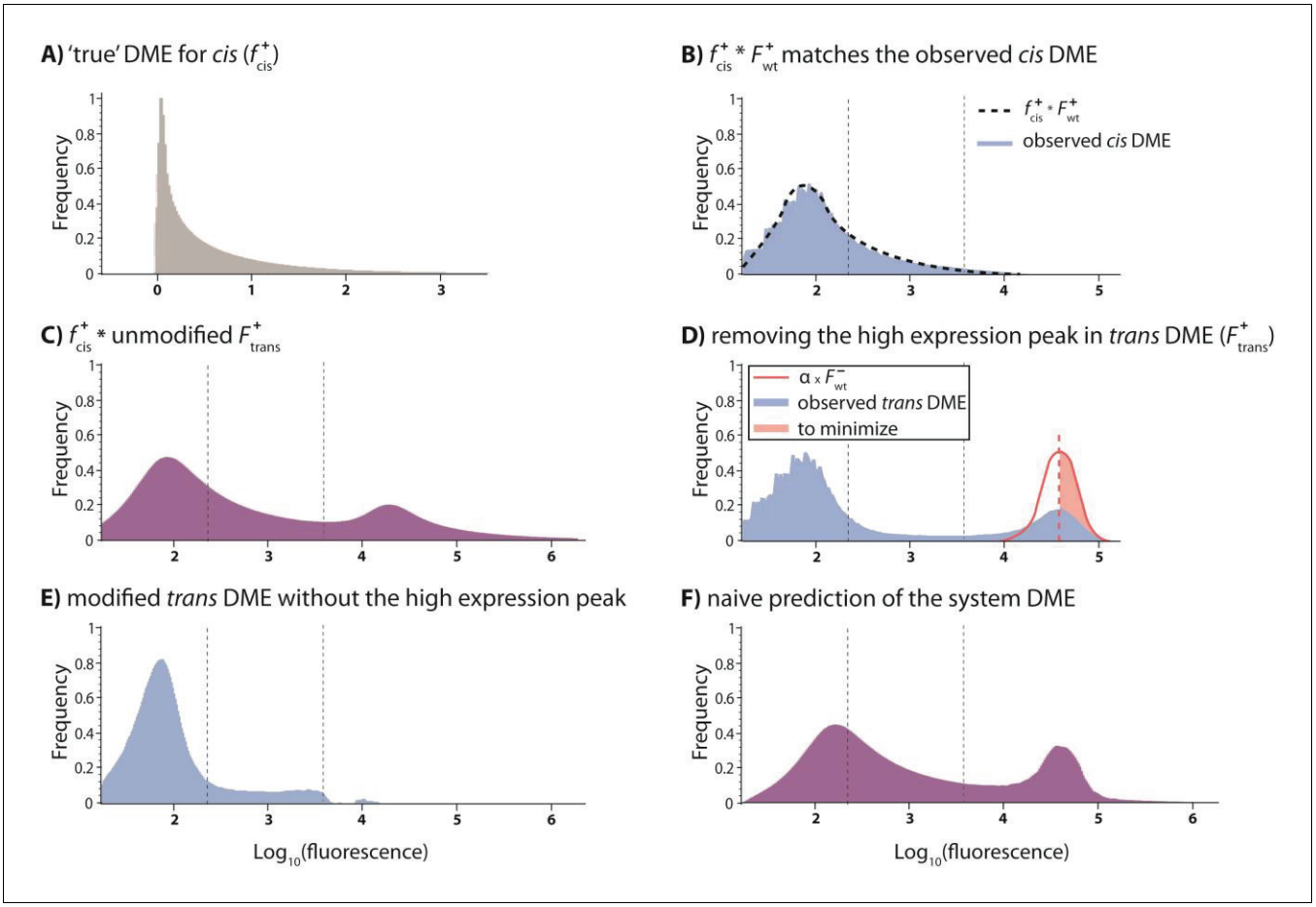

Figure 2—figure supplement 7. Predicting the system DME based on convolving component DMEs. We attempted to predict the DME for the system by convolving the corresponding DMEs of *cis* and *trans* components. To do this, we convolved the 'true' *cis* distribution (f_{cis}) with the observed *trans* distribution (F_{trans}). Here, we use low mutation probability *trans* and high mutation probability *cis* libraries to illustrate the procedure. (**A**) The 'reverse engineered' true *cis* distribution (f_{cis}), shows how the mutations in the *cis*-element alter wildtype expression levels. (**B**) The specific f_{cis} was chosen so to minimize the difference between the observed *cis*-element DME and the convolution between f_{cis} and the observed wildtype distribution in the presence of CI (F^+_{wt}). (**C**) Convolving f_{cis} with unmodified F_{trans} gives a predicted system DME with values outside the biologically meaningful range. (**D**) In order to impose a biological limit to high expression, which is given by the wildtype expression in the absence of CI, we removed the high expression peak in the observed *trans* DME. This was done by fitting a fraction of the wildtype distribution in the absence of CI (α) to minimize its (square) difference to the right-hand part of the high-expression *trans* peak (red shaded area). (**E**) Modified *trans* DME after removing high expression phenotypes. (**F**) The naïve convolution prediction for the system DME is obtained by convolving f_{cis} with the modified *trans* DME, and then adding back a corresponding amount of the high-expression wildtype distribution in the absence of CI. When convolving with the knowledge of the underlying genetic regulatory structure, we add back the *cis* DME in the absence of CI (modified by the scaling factor α). Here, the 'naïve' prediction is shown.
DOI: https://doi.org/10.7554/eLife.28921.010

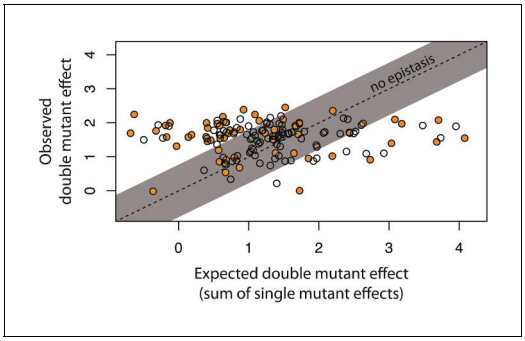

Figure 3. Epistasis in 150 random system double mutants. We created 150 double mutants with one unique random point mutation in the *cis*- and the other in the *trans*-element (***Figure 3—figure supplement 1***; ***Figure 3—source data 1***). In a plate reader, we measured expression levels of monoclonal populations of each double mutant and its constitutive single mutants in the presence of CI. Epistasis was estimated as the deviation of the observed double mutant effect from the additive expectation based on single mutant effects (shown for each double mutant in ***Figure 3—figure supplement 2***). The grey bar indicates measurements that are not in significant epistasis. Double mutants above the 'no epistasis' line are in positive epistasis (observed double mutant effect is greater than the additive expectation), and those below are in negative epistasis. Distributions of mutational effects for single *cis* and *trans* mutants, as well as for the double system mutants are shown in ***Figure 2—figure supplement 2***, while the data for epistasis calculations are shown in ***Figure 3—source data 2***. By measuring expression levels of 60 random double mutants (shown in orange) and their constitutive single mutants in the flow cytometer, we confirmed that estimates of epistasis using the plate reader correspond to those based on flow cytometry measurements (***Figure 3—figure supplement 6***; Source Data 3), and that the noise of expression is constant for all mutants (***Figure 3—figure supplements 3***, ***4*** and ***5***; Source Data 4).

DOI: https://doi.org/10.7554/eLife.28921.016

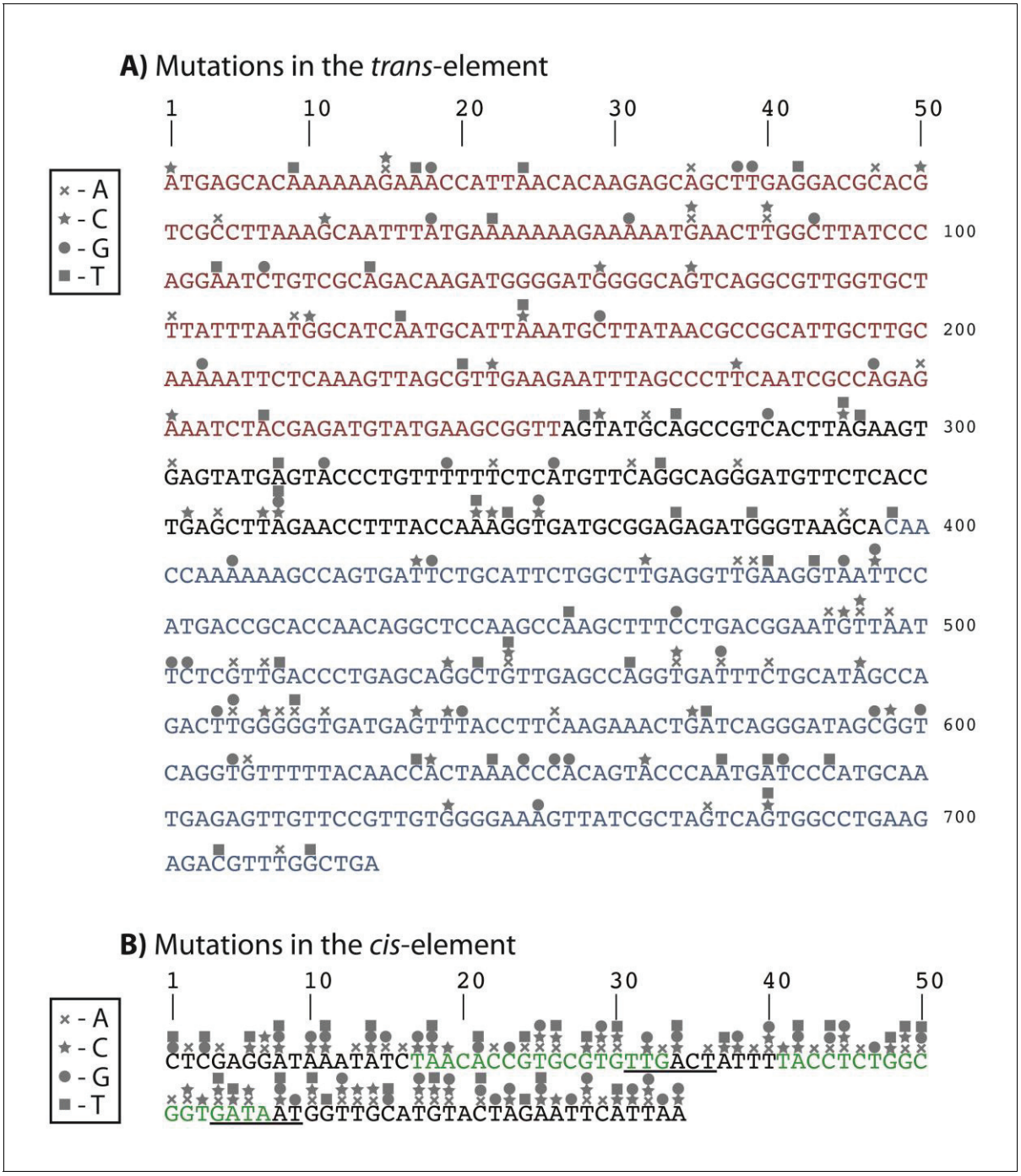

Figure 3—figure supplement 1. Identity and location of mutations in the 150 random double mutant library. (**A**) DNA sequence of the *trans*-element showing the N-Terminal Domain (red), C-Terminal Domain (blue), and the linker region connecting the two domains (black). (**B**) DNA sequence of the *cis*-element showing the −35 and −10 RNAP recognition sites (from left to right, underlined), and the two CI repressor operators (green). Every point mutation in cis was randomly paired with each mutation in trans, to give rise to 150 unique double mutants.

DOI: https://doi.org/10.7554/eLife.28921.017

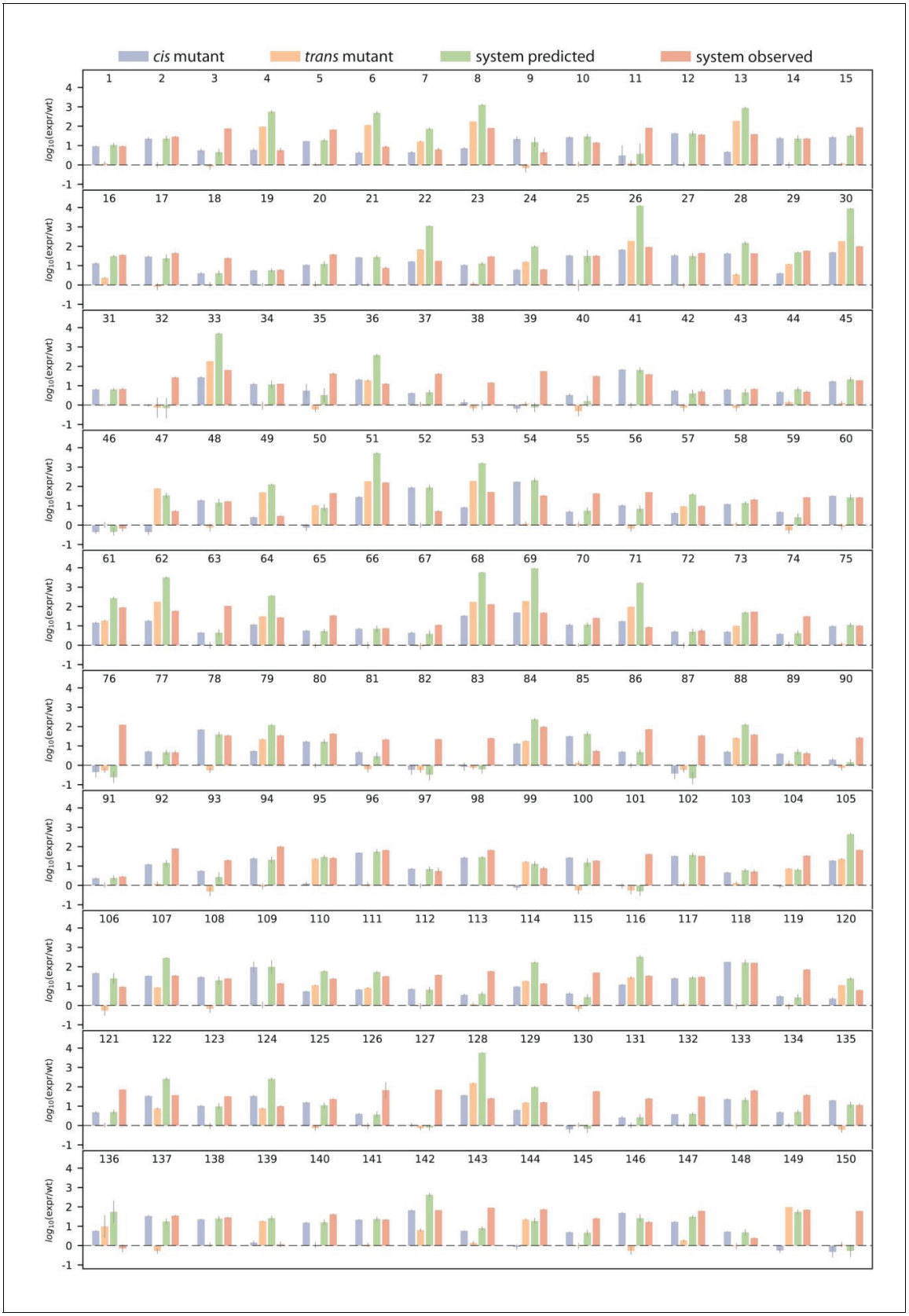

Figure 3—figure supplement 2. Single mutant effects, as well as predicted and observed double mutant effects. Bar charts show $\log_{10}$ of wildtype-normalized expression levels for (i) measured single mutant effect in the *cis*- (blue) and (ii) the *trans*-element (orange); (iii) predicted double system

Figure 3—figure supplement 2 continued on next page

Figure 3—figure supplement 2 continued

mutant expression based on single mutant effects (green), and (iv) observed double system mutant effect, for each double mutant. Numbers above the bar charts indicate double mutant identity, which correspond to ***Figure 3—source data 1***. Dashed line indicates same expression as the wildtype.

DOI: https://doi.org/10.7554/eLife.28921.018

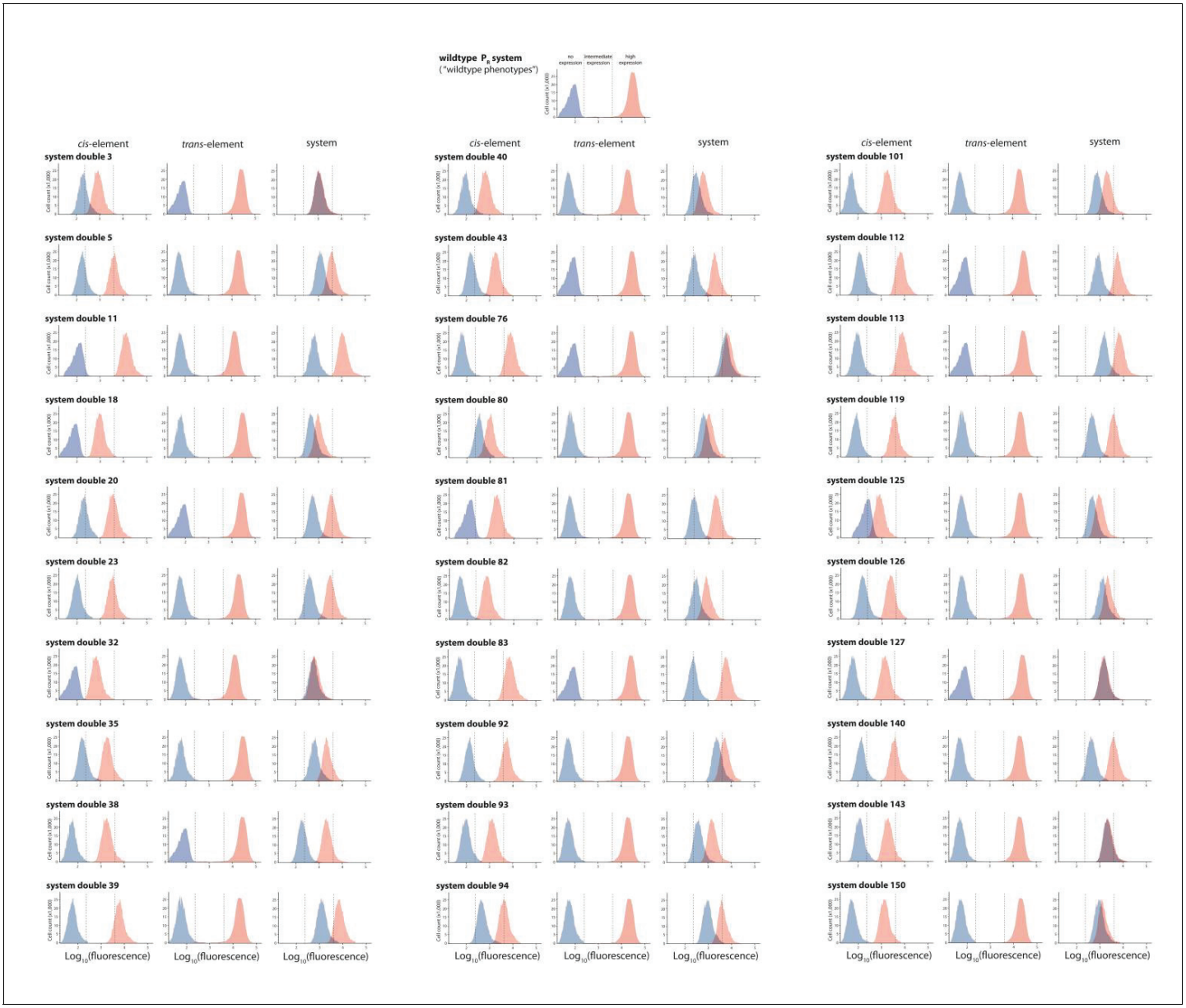

Figure 3—figure supplement 3. 30 double mutants with their corresponding single mutants, which are in significant positive epistasis. From the library of 150 mutants shown in ***Figures 3*** and 30 double mutants that were in significant positive epistasis were randomly selected. Expression levels of two replicates of each double mutant, as well as its corresponding single mutants, in the absence and in the presence of CI were measured for 100,000 individuals in the flow cytometer. Each panel shows the combined distribution based on two replicate measurements, for each mutant, in the absence (red) and in the presence (blue) of CI. From these measurements, we calculated gene expression noise for each isolate (***Figure 3—source data 3***), as well as epistasis (***Figure 3—figure supplement 2***; ***Figure 3—source data 4***). The top histogram shows the monoclonal wildtype distribution in the absence (red) and in the presence (blue) of CI.

DOI: https://doi.org/10.7554/eLife.28921.019

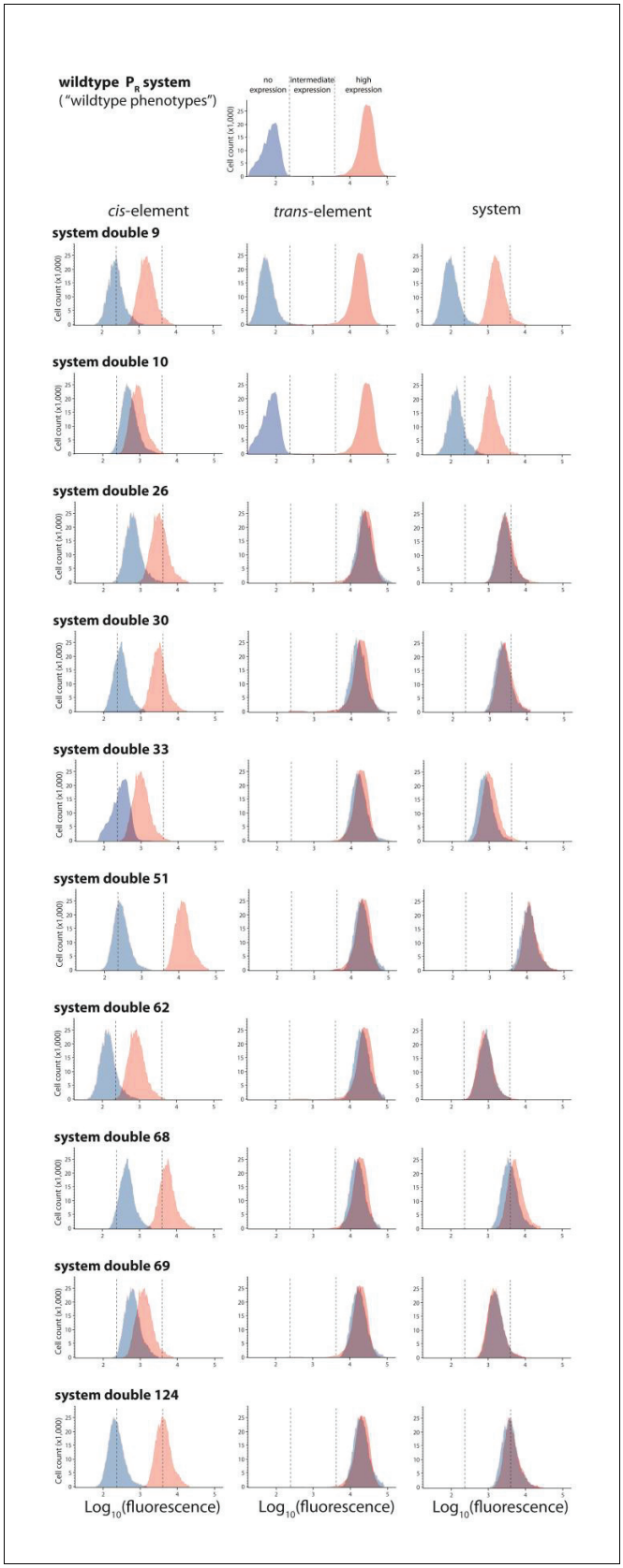

Figure 3—figure supplement 4. Ten double mutants with their corresponding single mutants, which are in significant negative epistasis. From the library of 150 mutants shown in ***Figures 3*** and 10 double mutants that

Figure 3—figure supplement 4 continued on next page

Figure 3—figure supplement 4 continued

were in significant negative epistasis were randomly selected. Expression levels of two replicates of each double mutant, as well as its corresponding single mutants, in the absence and in the presence of CI were measured for 100,000 individuals in the flow cytometer. Each panel shows the combined distribution based on two replicate measurements, for each mutant, in the absence (red) and in the presence (blue) of CI. From these measurements, we calculated gene expression noise for each isolate (***Figure 3—source data 3***), as well as epistasis (***Figure 3—figure supplement 2***; ***Figure 3—source data 4***). The top histogram shows the monoclonal wildtype distribution in the absence (red) and in the presence (blue) of CI.

DOI: https://doi.org/10.7554/eLife.28921.020

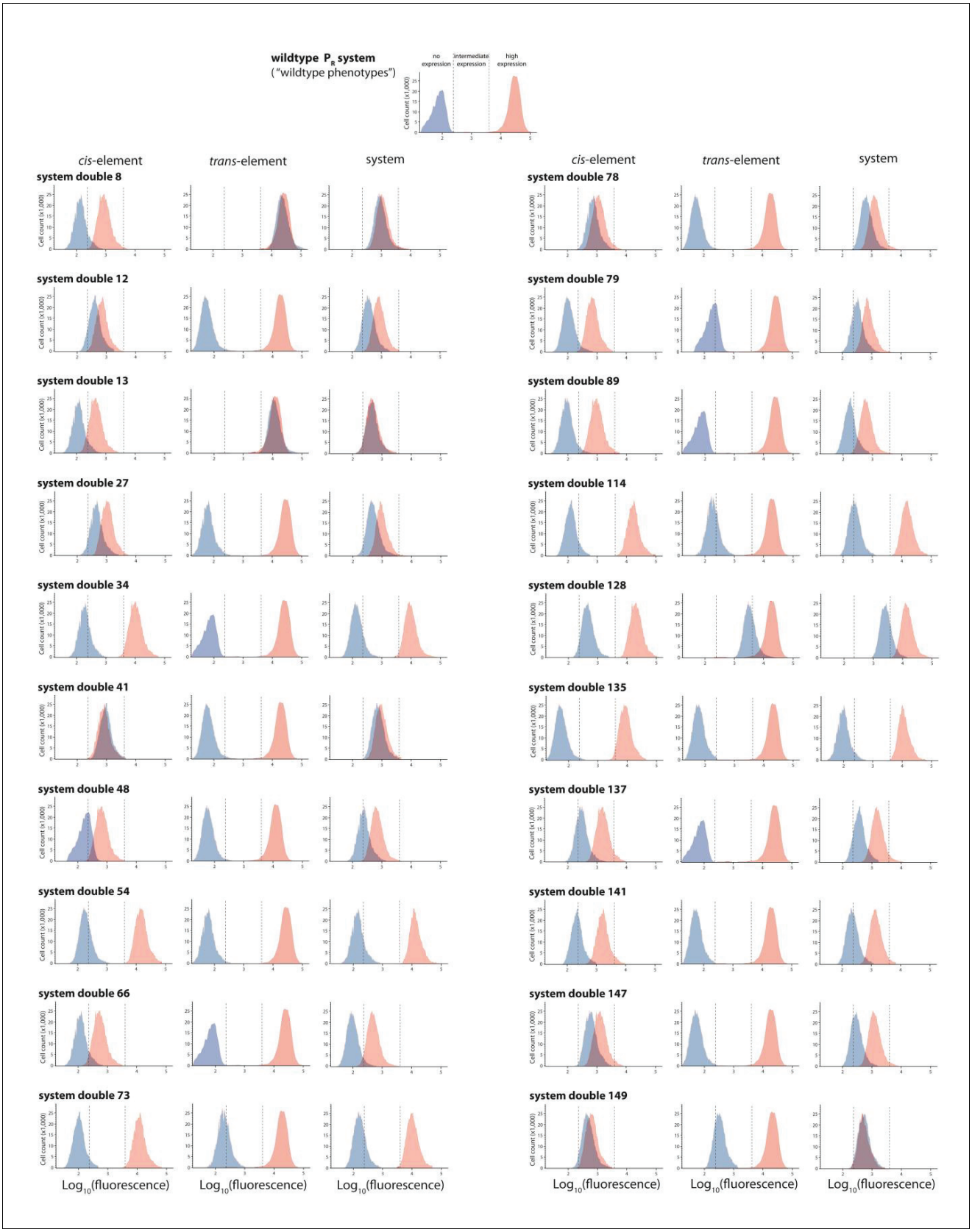

Figure 3—figure supplement 5. Twenty double mutants with their corresponding single mutants, which are not in significant epistasis. From the library of 150 mutants shown in ***Figures 3*** and 20 double mutants that were not in significant epistasis were randomly selected. Expression levels of two

Figure 3—figure supplement 5 continued on next page

Figure 3—figure supplement 5 continued

replicates of each double mutant, as well as its corresponding single mutants, in the absence and in the presence of CI were measured for 100,000 individuals in the flow cytometer. Each panel shows the combined distribution based on two replicate measurements, for each mutant, in the absence (red) and in the presence (blue) of CI. From these measurements, we calculated gene expression noise for each isolate (***Figure 3—source data 3***), as well as epistasis (***Figure 3—figure supplement 2***; ***Figure 3—source data 4***). The top histogram shows the monoclonal wildtype distribution in the absence (red) and in the presence (blue) of CI.

DOI: https://doi.org/10.7554/eLife.28921.021

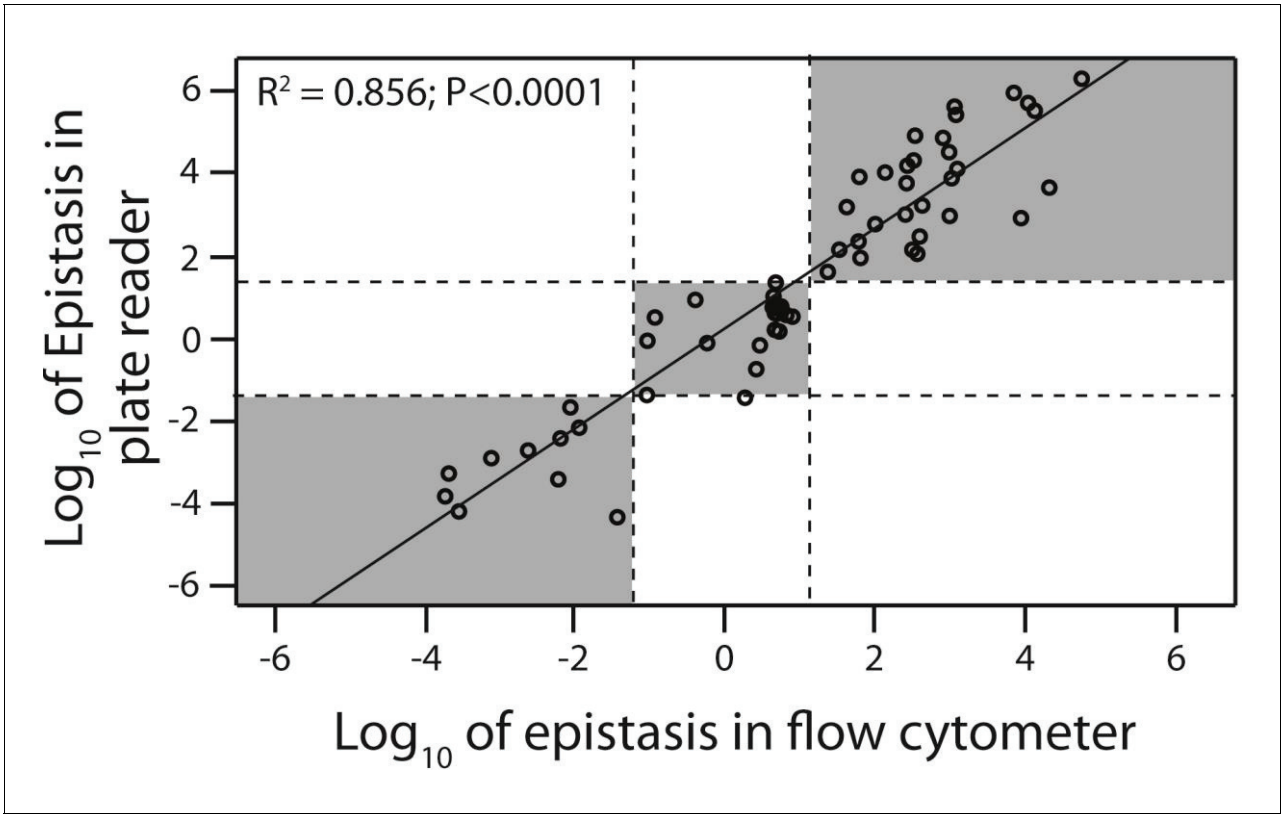

Figure 3—figure supplement 6. Flow cytometer and plate reader measurements give equivalent estimates of epistasis. Epistasis was calculated as the deviation of the double mutant effect from the additive prediction based on single mutant effects. From the original 150 double mutants measured in the plate reader, 30 that were in significant positive, 10 in significant negative, and 20 that were not in significant epistasis were selected (shown in orange in ***Figure 3***). Epistasis for these 60 mutants was calculated based on mean fluorescence and standard deviation (gene expression noise) obtained through flow cytometry measurements of 200,000 individuals (***Figure 3—figure supplement 3***;4;5). Linear regression was performed to test for correlation between two experimental approaches to calculating epistasis. Dotted lines are the 95% confidence intervals around the no-epistasis null model (Log_{10}(epistasis) = 0). Grey areas indicate that the two experimental methods give the same qualitative description of epistasis (positive, negative, or no significant epistasis). The data underlying this figure are shown in ***Figure 3—source data 2***;4.

DOI: https://doi.org/10.7554/eLife.28921.022

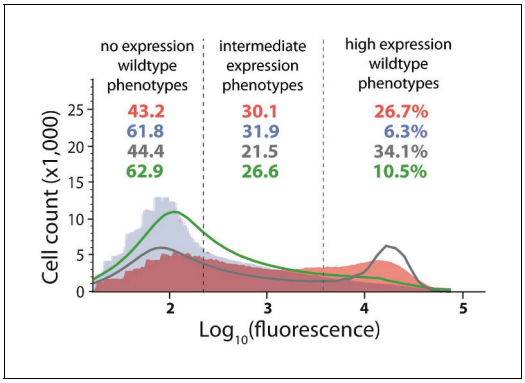

Figure 4. Distribution of expression phenotypes for the system with high mutation probability in the *cis*-element and low mutation probability in the *trans*-element. The distributions of fluorescence are shown in the absence of CI (shown in red) and in the presence of CI (shown in blue). Each distribution was obtained by pooling two independent measurements of 500,000 cells. The dashed lines separate three categories of phenotypes – no expression phenotypes (corresponding to repressed wildtype); intermediate expression phenotypes; and high expression phenotypes (corresponding to the wildtype in the absence of CI). Numbers are percentage of counts in each category, in the absence (red) and in the presence (blue) of CI. The naïve convolution-predicted DME in the presence of CI, performed in the absence of any knowledge of the genetic regulatory structure of the system, is shown in grey, together with the corresponding frequencies of mutants in each category. The convolution prediction that accounted for the regulatory structure is shown in dark green. Results of the Pearson's Chi-squared test for the differences between the observed and both types of convolution-predicted DMEs in the presence of CI are shown in ***Table 2***.

DOI: https://doi.org/10.7554/eLife.28921.028

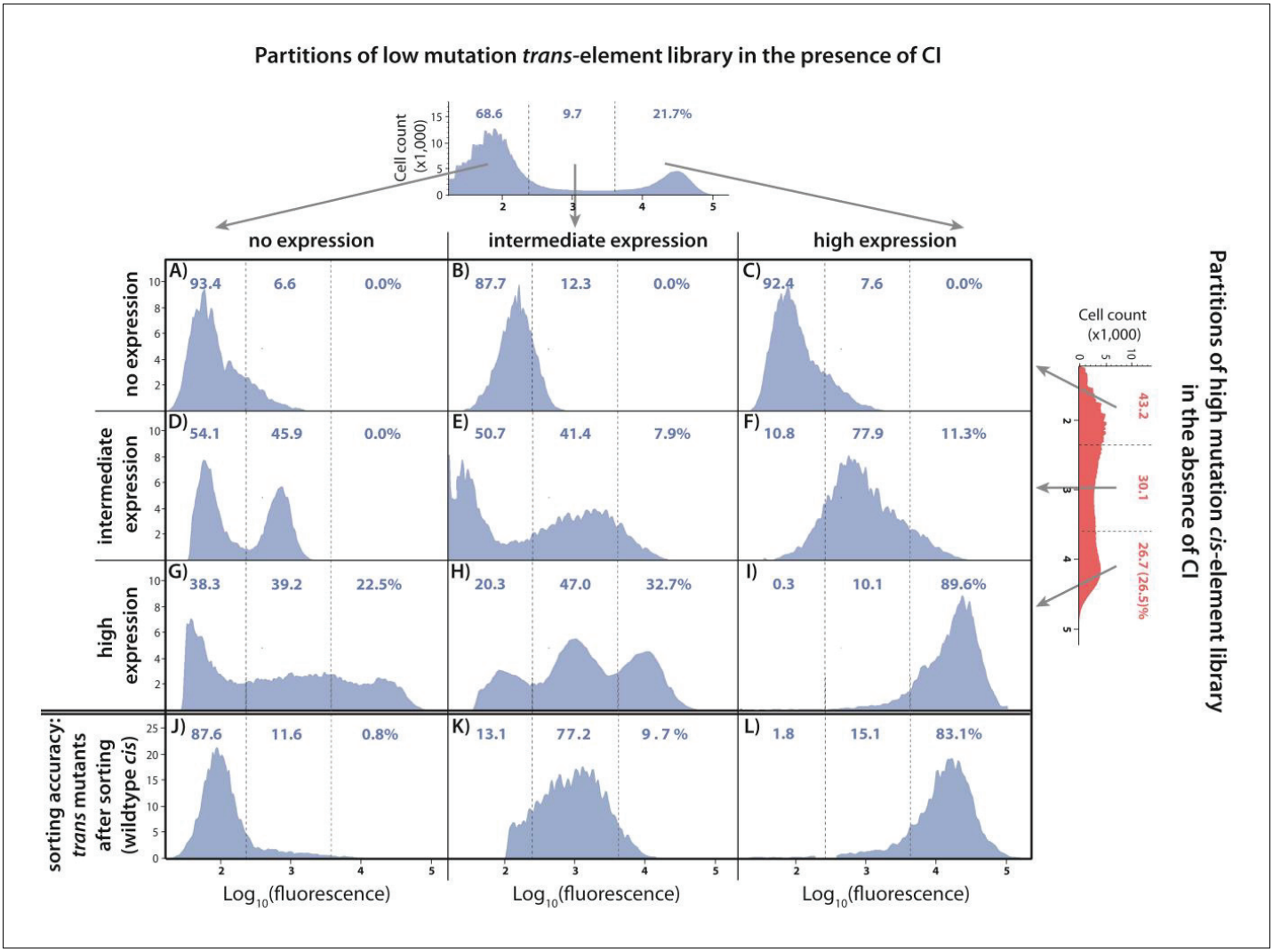

Figure 5. Understanding the interactions between mutations in *cis* and *trans* by accounting for the genetic regulatory structure of the system. Three partitions, obtained by FACS, of the low mutation probability *trans*-element library (corresponding to no expression, intermediate, and high expression phenotypes) were combined with the three equivalent partitions of the high mutation probability *cis*-element library in the absence of CI. The *cis*-element library in the absence of CI shows only the effect of mutations on RNAP binding. The DME of the original *trans*-element library, corresponding to ***Figure 2E***, is shown on top. The DME of the original *cis*-element library, corresponding to ***Figure 2—figure supplement 1D***, is shown on the right. The arrows illustrate from which category of the original DME (either *trans* or *cis*) were the sorted mutants used to make the particular combined library. DMEs of all nine partition-combination libraries were estimated using flow cytometry, with each distribution obtained by pooling two independent measurements of 500,000 cells. Also shown is the sorting accuracy, obtained as the repeated DME measurement of each *trans* partition following the original FACS sorting (panels J,K,L). The distributions of fluorescence are shown in the presence of CI. The dashed lines separate three categories of phenotypes – 'no expression', 'intermediate', and 'high expression' phenotypes. Numbers are percentage of counts in each category. At least in part, intermolecular epistasis can be qualitatively explained by considering the genetic regulatory structure of the system, as follows. Panels A), (**B**), and C): no expression *cis*-element mutants do not bind RNAP sufficiently to lead to expression, so that system mutants containing them remain in the 'no expression' bin irrespective of the effect of mutations in trans. (**D**) No expression *trans*-element mutants fully repress on wildtype *cis*. When *cis*-element mutations of intermediate expression are introduced, some still bind the functional CI mutants leading to repression, while others carry mutations that prevent CI binding, resulting in intermediate expression system mutants. (**E**) Intermediate expression CI mutants only partially repress on wildtype *cis*, but fully repress some *cis*-element mutants that have lowered RNAP binding. Other intermediate expression CI mutants do not bind a mutated *cis*-element background. (**F**) High expression CI mutants cannot bind wildtype *cis*. Similarly, they do not bind intermediate expression *cis*-element mutants, resulting in intermediate expression in the system. (**G**), (**H**): Some high expression *cis*-element mutants can fully bind mutated but functional CI mutants, others can only partially bind them, while some maintain full RNAP binding while losing all CI binding. (**I**) non-functional CI mutants do not repress on wildtype *cis*, hence also not on mutated *cis*-element backgrounds.

DOI: https://doi.org/10.7554/eLife.28921.029

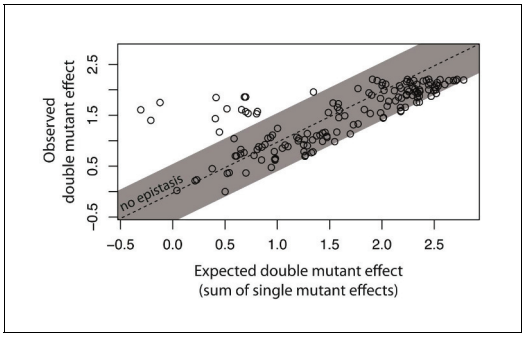

Figure 6. Not all intermolecular epistasis can be explained by accounting for the underlying genetic regulatory structure of the system. We created 150 double mutants with single mutation combinations corresponding to ***Figure 5G***. A double mutant in this library would not be in epistasis unless a mutation in trans binds the *cis* mutant differently than the wildtype *trans* does. In a plate reader, we measured expression levels of monoclonal populations of each double mutant and its constitutive single mutants in the presence of CI. Epistasis was estimated as the deviation of the observed double mutant effect from the additive expectation based on single mutant effects. The grey bar indicates measurements that are not in significant epistasis. The effects of single mutations in *cis* and in *trans*, as well as the double mutant effects, are shown in ***Figure 6—figure supplement 1***, while the underlying data are shown in ***Figure 6—source data 1***. The location of point mutations is shown in ***Figure 6—figure supplement 2*** and ***Figure 6—source data 2***.

DOI: https://doi.org/10.7554/eLife.28921.030

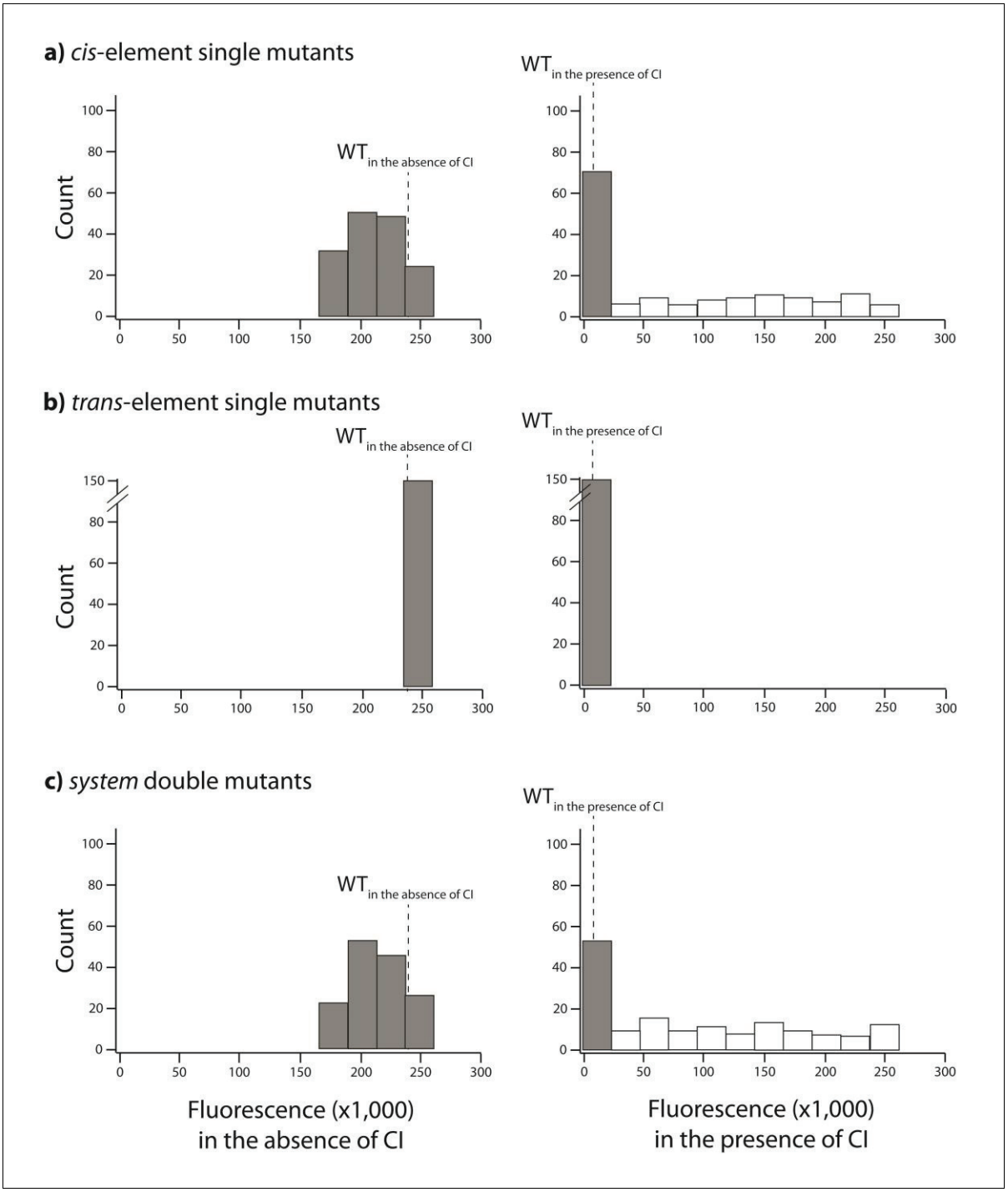

Figure 6—figure supplement 1. Distribution of single mutation effects in 150 system double mutants and their corresponding single mutants. We created 150 unique double mutants, with one point mutation in *cis* that had high expression in the absence of CI, and the other in trans that exhibited no expression in the presence of CI. We measured gene expression of each mutant at a population level in a plate reader. Histograms of expression levels in the absence and in the presence of CI are shown for: (**A**) point mutations in *cis*; (**B**) point mutations in trans; (**C**) double mutants in the system. Dotted line represents mean wildtype fluorescence in the corresponding environment. Six replicates of each mutant were measured. Grey bars indicate mutants that were not significantly different from the wildtype. The data underlying this figure are shown in ***Figure 6—source data 1***.

DOI: https://doi.org/10.7554/eLife.28921.031

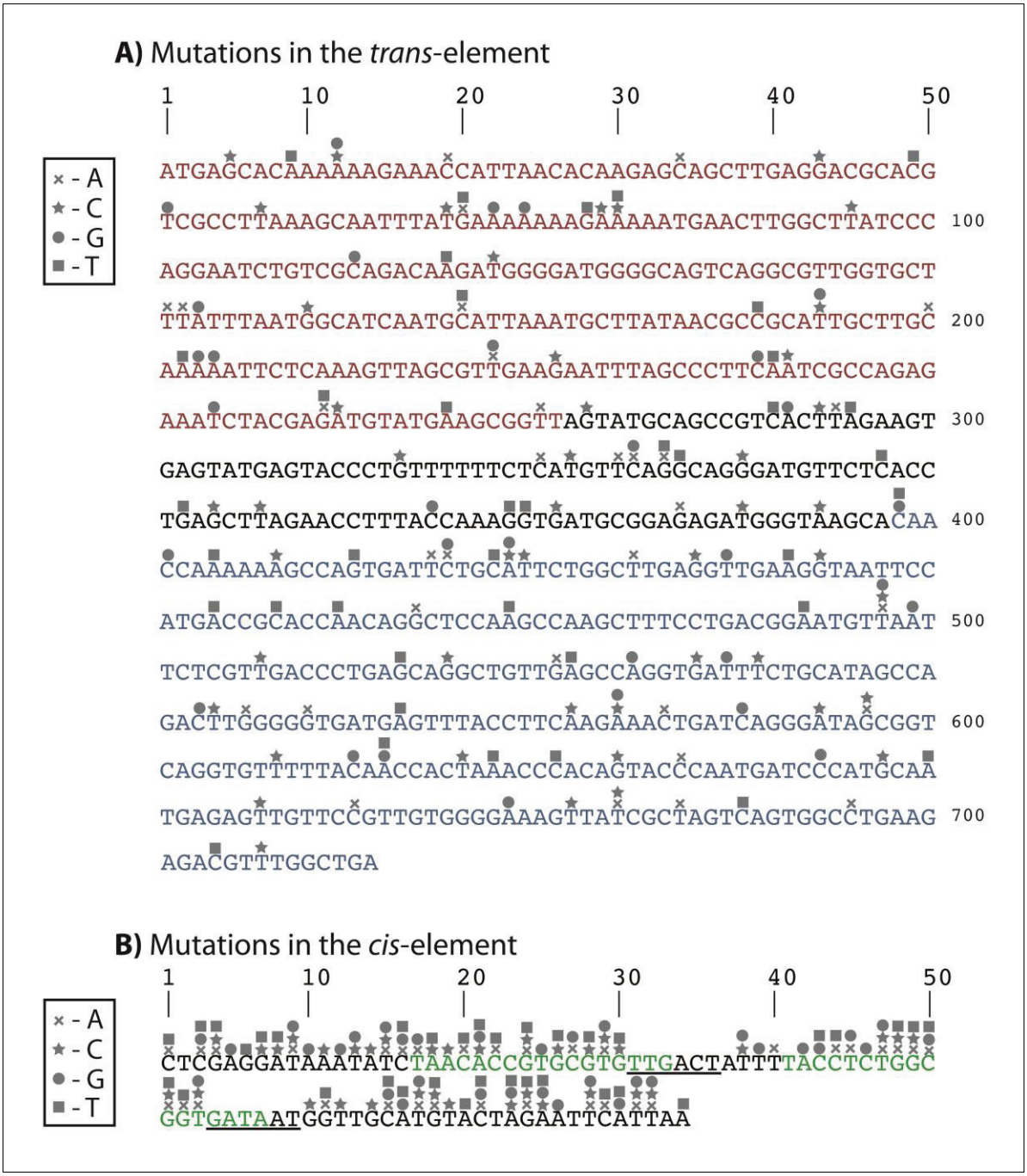

Figure 6—figure supplement 2. Identity and location of mutations in the double mutant library with a *trans* mutation that has no expression in the presence of CI, and a *cis* mutation with high expression in the absence of CI. (**A**) DNA sequence of the *trans*-element showing the N-terminal domain (red), C-terminal domain (blue), and the linker region connecting the two domains (black). (**B**) DNA sequence of the *cis*-element showing the −35 and −10 RNAP recognition sites (from left to right, underlined), and the two CI repressor operators (green). Unlike the double mutant library shown in ***Figure 3***, where each double mutant had a unique mutation in *cis* and *trans*, in this library some *cis* mutations had to be repeated, as we could not identify 150 mutations in *cis* that had high expression in the absence of CI.

DOI: https://doi.org/10.7554/eLife.28921.032