A novel design of transcriptional factor-mediated dynamic control of DNA recombination.

Jiayang Li, Yihao Zhang, Yeqing Zong.

ABSTRACT

Genetic regulation is achieved by monitoring multiple levels of gene expression, from transcription to protein interaction. Unlike common temporary transcription regulation methods such as the use of inducible promoters, recombinases permanently edit DNA sequences. Recombinases, however, require especially strict regulation when implemented in synthetic genetic systems because of the irreversible result. Here we propose to improve the regulation of site-specific recombinase based genetic system by dynamically hiding one of the att sites that are essential for recombination with transcriptional factors. After effectively suppressing excessive recombination, we also validated the necessity of each of the essential components in our transcription factor-controlled recombination (TFCR) system. Our system applied transcription-level regulators directly on controlling the activity of existing non-transcription-level trans-factor proteins by inhibiting its binding to the cis-regulatory elements. We anticipate our results to provide greater stability for recombinase components to enable safer use in systems as well as be a starting point for future cross-expression level gene regulations.

Keywords: Synthetic biology, gene regulation, recombinase, transcriptional factor, dynamic control

Found in bacteriophages, site-specific recombinases are naturally used to insert phage DNA into bacteria. Capable of permanently excising or inversing DNA regions between two recognition sites attB and attP, recombinases have been applied in diverse biological systems. In classical molecular operations, recombinases were used to perform site-specific in vivo assembly of genes[7]. Researchers even showed the possibility to target att sites in human cells with recombinases and recombine them[1,2]. In the field of synthetic biology, its ability to irreversibly change the genetic circuit enabled researchers to permanently record events and the order in which they occur in a cell[8]. It has been found that some recombinases also possess high orthogonality, enabling them to be used together within a single complex system as biological registers to record over 1 byte of digital information and to create state machines in cells[9]. Other researches have also shown that the recombinase system can be used as a rewritable addressable data module to store digital information within chromosomes[5].

While utilizing recombinase's unique characteristics, one factor that cannot be ignored is the excessive expression of recombinase. Without proper control, recombination can irreversibly happen without induction, which reduces the controllability of the system and weakens precise regulation. The hyperactivity of recombinases can be suppressed by adjusting components at different levels of gene expression: transcription, translation, and protein modification. Possible solutions include finetuning the promoter and ribosome binding site (RBS). Previous systems implementing recombinases have applied fine regulation on the component through methods such as using tightly regulated promoters like pPhlF and pBAD[8] and decreasing the recombinase translation by switching start codon[5].

Here, we design an effective regulation of recombination on a different level by introducing engineered transcriptional factors with the ability of high-order oligomerization. According to former models, recombination does not occur without recognition of both att sites by recombinases. Taking advantage of this theory, we attempted to disable interactions between recombinases and att sites. Transcriptional factors such as repressors naturally have the ability to disable interactions between RNA polymerase and a DNA region; taking advantage of the DNA looping ability of transcriptional factors with oligomerize domain, we attempted to insulate att sites with the loop.

1 MATERIALS & METHODS

1.1 CONSTRUCTION OF RECOMBINASE PLASMIDS AND TRANSCRIPTIONAL FACTOR PLASMIDS

All recombinase plasmids and TF plasmids used in this study were assembled using restriction enzyme cloning, Golden Gate assembly, and GIBSON assembly. Correct colonies are selected using antibiotic selection and examined via DNA sequencing.

1.2 CONSTRUCTION OF REPORTER PLASMIDS

The vector for the reporter plasmids was assembled from a low-copy BAC vector derived from pcc1BAC. Superfold-GFP expression unit from the pPT plasmid[11] was added onto the vector through BioBrick Standard Assembly. The region with T7 promoter and PhiC31 attP site was added by PCR with primers carrying over 20 bp homologous arm to the vector as well as 20-30 bp overlapping the new region. The plasmid was transformed into competent *E. coli*, verified

through sequencing, and then extracted and purified. The region with the operators (or meaningless sequence with the same length) and the attB site was constructed from PCR with primers and PCR Phusion. PCR was carried with two primers with over 20 bp overlapping region without a plasmid template. The product of this PCR was then used in another PCR, amplified by the old reverse primer and a new forward primer with 20 bp overlapping region with the old forward primer to elongate the length of the previous PCR product. The final PCR product was designed to carry two restriction enzyme recognition sites with three protection base pairs at each of the two ends. These two restriction enzyme recognition sites enabled them to be inserted into the formerly constructed reporter plasmid at the appropriate location.

All plasmid maps are provided in supplementary data 6.1, and all plasmid files are provided in supplementary data 6.4.

1.3 Strain Preparation

Reporter plasmids (pBCL0, pBCL1, pBCL2, pBCP0, pBCP2) and TF plasmids (pUAP, pUAL) were constructed and extracted separately. For each combination, a reporter plasmid and a TF plasmid are co-transformed into *E. coli* DH10B in which a constitutively expressing T7 RNA polymerase module is integrated into the genome[11]. The strains were then made competent and the recombinase plasmid was transformed into these competent strains carrying both a reporter plasmid and a TF plasmid. The final *E. coli* strains with three plasmids were grown overnight on LB agar plates with antibiotic selection of Kanamycin A, Ampicillin, and Chloramphenicol. Regions or colonies of *E. coli* were selected from the agar plate and inoculated into liquid M9 medium with antibiotic selection with the Kanamycin A, Ampicillin, and Chloramphenicol as well. The incubation temperature is 37ºC for both LB agar plates and liquid M9 medium.

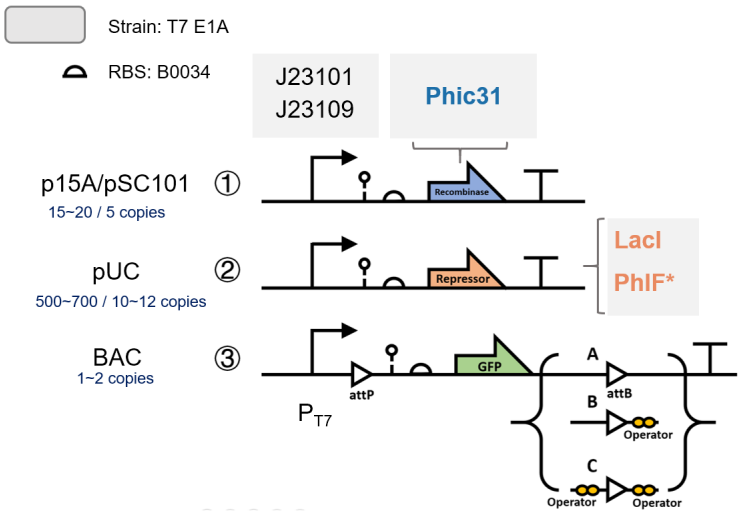


Figure 1. A schematic of plasmids used in this study.

1.4 Fluorescence Measurement

Fluorescence of groups in the experiments was both observed with the naked eye or fluorescent microscope and quantitatively measured through flow cytometry.

For flow cytometry measurement, the detailed procedures are as follows:

1. Co-transfer the recombinase-expressing plasmid, TF plasmid, and reporter plasmid into the host *E. coli* strain.

2. Inoculate monoclonal colony in the M9 supplemented medium.

3. The cell cultures were diluted 1000-fold with M9 supplemented medium of various inducer (IPTG or cumate, depending on the experiment) concentrations, and were incubated for 20h.

4. 3-μl samples of each culture were transferred to a new plate containing 200 μl per well of PBS supplemented with 2 mg/μL kanamycin to terminate protein expression. The fluorescence distribution of each sample was assayed using a flow cytometer with appropriate voltage settings. The ratio of fluorescent cells and the arithmetical mean of each sample were determined using FlowJo software.

2 RESULTS

2.1 PLASMID DESIGN AND DETERMINATION OF REPRESSION EFFICIENCY

The trans-factor-controlled recombination (TFCR) system we designed contains three plasmids: a recombinase plasmid that expresses recombinase PhiC31, a TF plasmid that results in induced or constitutive TF protein expression, and a reporter plasmid that indicates recombinant reaction quantitatively.

The basic structure of the reporter plasmid consists of a T7 promoter, an attP site, GFP, an attB site parallel to the attP site, and a terminator. We attempted to repress recombinase activity by putting operator sites around the attB site which is necessary for recombination according to former models[5,6]. Three types of reporter plasmids are set up, each containing none, one or two operator sites around the attB site (Figure 2). For the groups with two operators, one operator is placed on each side of the attB site. The spacing between two operators was set around a multiple of 10 bp to make sure that both operators were located on the same side of the DNA double helix. For the groups with zero or one operator sites, meaningless sequences with the length of the operators are added at the location of the operators to maintain equal distances between the two att sites in each group. We assume that when the TFs bind to the operator sites around the attB region, the recombinase becomes no longer able to recognize both att sites, which prevents unwanted recombination.

Without the interference of TFs, the two parallel att sites are exposed to recombinase to enable recombination, resulting in excision of the region of DNA between the two sites containing the Green Fluorescent Protein expressing module (Figure 2). This would result in an increasing ratio of low-fluorescence E. coli in the population. Therefore, in flow cytometry measurement, a higher percentage of cells at low fluorescence is considered as the identification of high recombinase activity, and vice versa.

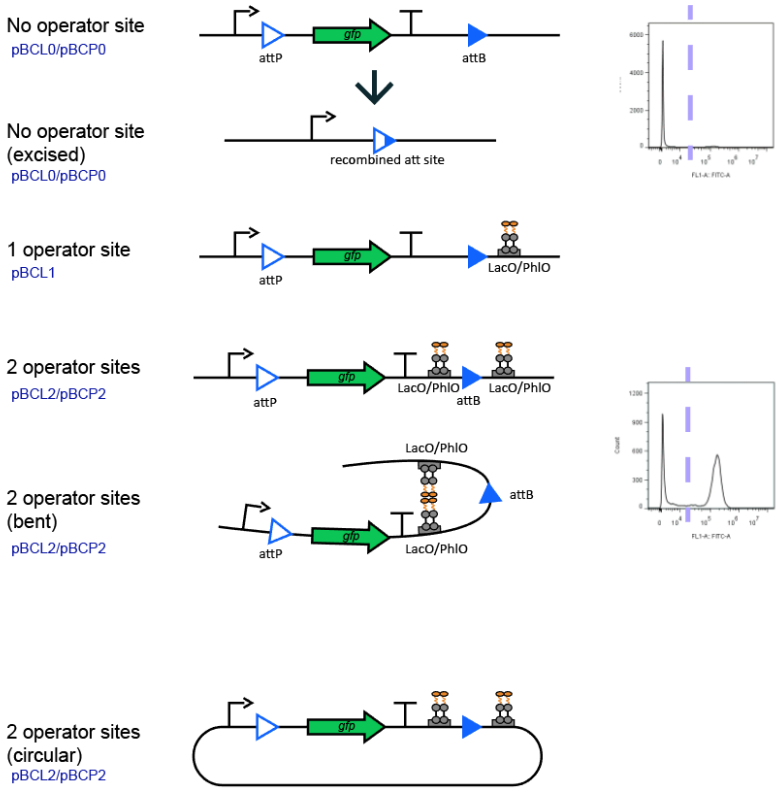


Figure 2 Design of the reporter plasmid. In the reporter plasmid with no operator sites, recombinase can recognize the attP and attB sites and recombine them, excising the DNA region between the two sites. The circuit after excision is shown as no operator site (excised). The diagram for 2 operator sites (bent) illustrates the proposed model of TFs with oligomerization domains in which TFs oligomerize and bend the region of attB between them, disabling recognition and excision by recombinases. Diagrams on the right represent expected shapes of flow cytometry under conditions on the left. The light purple dotted line represents the border between non-fluorescent and fluorescent cells. The percentage of fluorescent cells is calculated from the count of cells on the right side of the line divided by the total count.

2.2 Preliminary results from the initial design of the TFCR system

According to the hypothesized DNA bending mechanism of these TFs, the full repression function requires both the operator sites to bind with each other and hide the att site in between. To test the plausibility of our TFCR system, we compared the performance of systems with one, or two operator sites around the attB site. We chose the widely applied lacI TF and PhiC31 recombinase for the preliminary experiment. Due to the uncertain effect on recombination rate by the difference on the DNA sequence of trans-regulatory element, we focused on comparing results from groups with and without induction of the TF at different recombinase induction levels consisted of identical plasmids.

The group with one operator site demonstrated little difference on recombination rate between groups with and without TF plasmid induction at high recombinase plasmid induction levels, showing minimal repression activity by the LacI TF. In the group with two operator sites, at high recombinase plasmid induction levels, the group without TF decoupling from the operators by IPTG was measured to have lower overall recombination rates, showing over 20% lower

recombination rate than the group with IPTG interference at 40 μM cumate induction level. This demonstrated the plausibility of reducing recombinase activity with the TFCR system.

Since the group with two operators demonstrated significantly better repression at 50 μM cumate than the group with one operator, we continued to use the two-operator set in the following experiments.

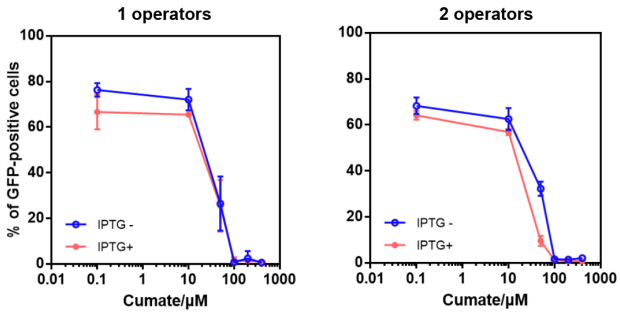


Figure 3. Results of the preliminary experiment. Two systems are set up, one with one Lac operator and the other with 2 Lac operators. Recombination levels were measured at 0, 10, 50, 100, 200, 400 μM cumamte induction levels. The group with 2 operators IPTG- showed notable reduction of recombination level at 50 μM cumate than the group with 2 operators IPTG+.

2.3 IMPROVED TFCR SYSTEM REDUCED RECOMBINATION

Although the considerable difference was shown in the preliminary experiment, the recombination difference between the two-operator strain and the zero-operator strain was not ideal. We suspected that the small difference between zero, one, and two sites in the preliminary experiment may be due to the relatively weak repression ability of lacI. Thus, we replaced lacI with PhlF TF fused with oligomerization domains from another TF protein cI434 (PhlF*) to increase repression level[10]. The new TF PhlF* plasmid also implements an inducible Ptac promoter to enable convenient control of TF expression. With the PhlF* TF, three groups of experiments were set up to validate the necessity of each component in our improved TFCR system.

We first tested the effect of the PhlF* TF within the system. The experimental group consists of a reporter plasmid with two operators, the TF plasmid and the recombinase plasmid. This group is treated with 0.5 mM IPTG to initiate the expression of the TF plasmid. The second group contains the same plasmids but is not treated with IPTG. Results from flow cytometry showed that at a representative recombinase induction level, the group without IPTG resulted in over 25% of decrease in the percentage of highly fluorescent cells from the experimental group (Figure 4A). We therefore concluded from the significant increases in recombinase activity without the presence of the PhlF* TFs that the TFs are essential to the repression of recombinase activity.

The necessary role of PhlO operators in the repression system was also verified. Two groups with all three plasmids are constructed, one with two operator sites on the reporter plasmid and the other with none. With IPTG induction of the TF plasmid, the group with two operators showed over 50% decrease in the percentage of high-fluorescence cells than the group with no operators at recombinase induction levels over 50 μM cumate (Figure 4B). Thus, we confirmed the necessity of the two operator sites to achieve an effective decrease in recombination.

In addition, we validated our suspect of the artificially appended oligomerization domain as an essential condition for repression of recombinase. A group with induced wild type PhlF TF replacing the engineered PhlF* TF in an otherwise identical system was set up and compared with the group with the PhlF* system (Figure 4C). With induction levels past 50μM cumate, the group with the wild type PhlF demonstrated significantly higher recombination rates that were similar to those from the PhlF* system without TF induction. The results here matched the model of TFs with oligomerization domains that these domains are indispensable for proper repression.

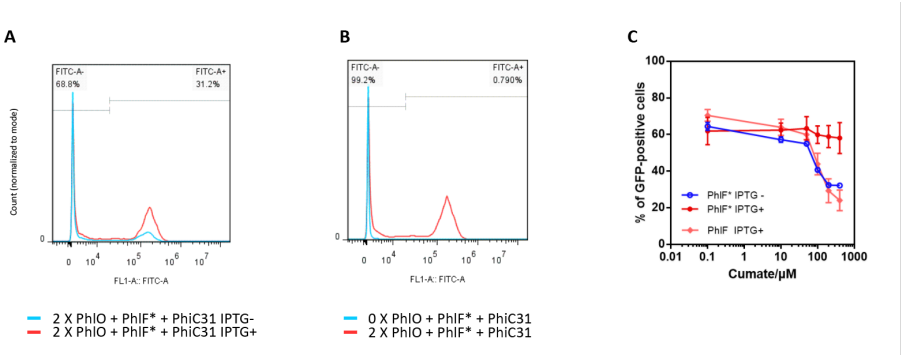


Figure 4. Flow cytometry results. (A) Testing the necessity of PhlF* TF in the reduction of recombinase activity by altering IPTG induction of 0 or 500 μM, with 400 μM cumate. (B) Testing the necessity of PhlO operators in the reduction of recombinase activity at 200μM cumate. (C) Comparing results from systems with induced PhlF*, uninduced PhlF* and induced wild type PhlF to test the necessity of the oligomerization domain for reduction of recombinase activity. The group with induced wild type PhlF showed similar recombinase activity levels as the uninduced PhlF* group beyond 50μM cumate induction level.

3 DISCUSSION

4.1 SIGNIFICANCE OF THE PROJECT

In former studies, recombinase activity was controlled at transcriptional and translational levels with methods such as changing promoter strength, finetuning RBS, and using tightly regulated promoters[5,8]. In this project, we suggested the repression of att sites with operators as a new level of recombinase regulation. This level is orthogonal to the other regulation methods, adding a new possible approach to be used with the above-mentioned methods in the same system. With the addition of each level of regulation, the efficiency of regulation multiplies, reducing the possibility and extent of expression leakage. Therefore, through increasing robustness and controllability of recombinases, this project offers them greater potential to be applied for gene regulation in engineered biological systems.

Our design enables synthetic biological systems to apply controllable removal of DNA sequences. In a patented biological manufacturing system, recombinases can be programmed to delete the patented gene components when the cells leave the desired conditions and prevent the theft of intellectual properties. A similar system can be applied to pollutive biological manufacturing factories; with gene contents which express pollutive chemicals removed by

recombinase when the cells leave ideal incubation conditions, biochemical pollution to natural water sources can be reduced efficiently.

On the other hand, we explored more potentials of TFs, which are commonly used only for the regulation of promoters, operators, and repressors that control downstream transcription. Formerly used only for direct regulation of transcription, TFs were now found the possibility to regulate trans-factor protein activity by interacting with the DNA region containing cis-regulatory elements. Its function of regulating recombinase activity demonstrates its potential of interacting with other cis-regulatory elements to enable the locking of DNA regions to prevent unwanted reactions with substances in the environment, allowing further complex signal processing through the addition of a new controllable component.

Apart from this, the results of this project were consistent with the model established by previous researches on recombinase[5,6]. According to the model of recombination by large serine recombinases, two recombinase dimers each need to bind with an att site to perform recombination. In the experiments of our project, recombination ceased after disabling the recognition of one of the att sites. This result satisfies the former established model by verifying the necessity of the recombinase and both att sites in recombination.

Furthermore, we took advantage of independent parts artificially modified by synthetic biologists. Replacing the natural LacI repressor with the engineered PhlF* repressors with oligomerization domains, we achieved a more rigorous repression activity. Utilizing artificially modified or assembled vectors with standard BioBrick components, we conveniently constructed our plasmids with appropriate antibiotic resistance and number of copies. Building off the database with works by former researchers, we added the new recombinase system to provide another option for signal processing for future works.

4.2 FUTURE PERSPECTIVES

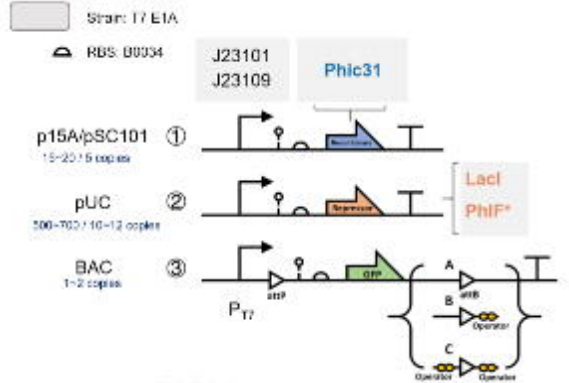
Previous researches have proposed the model of cooperation between TFs with oligomerization domains[10]. According to this theory, the presence of at least two operators enables them to oligomerize and bend the DNA region between them to disable actions to the region. Therefore, it is reasonable that one TF would have significantly less effect than two TFs would. However, in the preliminary experiment, little difference was observed between strains with only one and two operator sites. Thus for future perspectives, it is valuable to find out if the effect of the oligomerization of TFs is crucial for effective control of recombination.

4 REFERENCE

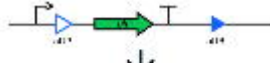
1. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403,** 339–342 (2000).
2. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352,** aac7341–aac7341 (2016).
3. Stanton, B. C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nature Chemical Biology* **10,** 99–105 (2013).
4. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
5. Bonnet, J., Subsoontorn, P. & Endy, D. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U.S.A.* **109,**

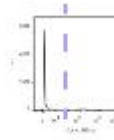
8884–8889 (2012).

6. Smith, M. C. M., Brown, W. R. A., McEwan, A. R. & Rowley, P. A. Site-specific recombination by φC31 integrase and other large serine recombinases. *Biochemical Society Transactions* **38,** 388–394 (2010).
7. Colloms, S. D. *et al.* Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Research* **42,** e23–e23 (2013).
8. Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S. & Lu, T. K. Synthetic recombinase-based state machines in living cells. *Science* **353,** aad8559 (2016).
9. Yang, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nature Methods* **11,** 1261–1266 (2014).
10. Hou, J. *et al.* Engineering the Ultrasensitive Transcription Factors by Fusing a Modular Oligomerization Domain. *ACS Synth Biol* **7,** 1188–1194 (2018).
11. Zong, Y. *et al.* Insulated transcriptional elements enable precise design of genetic circuits. *Nat Commun* **8,** 52 (2017).
12. Groth, A. C., Olivares, E. C., Thyagarajan, B. & Calos, M. P. A phage integrase directs efficient site-specific integration in human cells. *Proceedings of the National Academy of Sciences* **97,** 5995–6000 (2000).

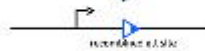

Strain: T7 E1A
RBS: B0034
J23101
J23109
Phlc31
p15A/pSC101 ①
15~20 / 5 copies
pUC ②
500~700 / 10~12 copies
BAC ③
1~2 copies
LacI
PhlF*
P_{T7}
A
B
C
Operator
Operator
Operator

No operator site



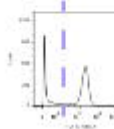


No operator site (excised)

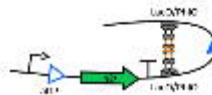


1 operator site

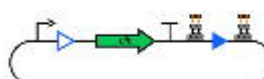
2 operator sites

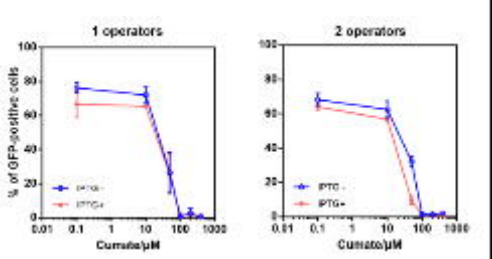


2 operator sites (bent)



2 operator sites (circular)




1 operators
2 operators
% of GFP-positive cells
Cumate/μM
IPTG -
IPTG+
0.01
0.1
1
10
100
1000
0
20
40
60
80
100

A
B
C
2 X PND + PNP+ + PNC3L IPTG-
2 X PND + PNP+ + PNC3L IPTG+
0 X PND + PNP+ + PNC3L
2 X PND + PNP+ + PNC3L
% of CFP+ cells
Concentration