HHS Public Access

Author manuscript

Nat Methods. Author manuscript; available in PMC 2015 March 01.

Published in final edited form as:

Nat Methods. 2014 September ; 11(9): 966–970. doi:10.1038/nmeth.3030.

Isolation of rare recombinants without using selectable markers for one-step seamless BAC mutagenesis

George T. Lyozin[1], **Paul C. Bressloff**[#2], **Amit Kumar**[#3], **Yasuhiro Kosaka**[1], **Bradley L. Demarest**[4], **H. Joseph Yost**[4,5], **Michael R. Kuehn**[3], and **Luca Brunelli**[1]

[1]Department of Pediatrics (Neonatology), The University of Utah School of Medicine, Salt Lake City, Utah, USA [2]Department of Mathematics, The University of Utah, Salt Lake City, Utah, USA [3]Laboratory of Protein Dynamics and Signaling, National Cancer Institute, Frederick, Maryland, USA [4]Department of Neurobiology and Anatomy, and [5]Department of Pediatrics, The University of Utah School of Medicine, Salt Lake City, Utah, USA

[#] These authors contributed equally to this work.

Abstract

Current laboratory methods to isolate rare (1:10,000 to 1:100,000) bacterial artificial chromosome (BAC) recombinants require selectable markers. Seamless BAC mutagenesis needs two steps: isolation of rare recombinants using selectable markers, followed by marker removal through counterselection. Here we illustrate founder principle-driven enrichment (FPE), a simple method developed to rapidly isolate rare recombinants without using selectable markers, allowing one-step seamless BAC mutagenesis. As proof-of-principle, we isolated 1:100,000 seamless fluorescent protein-modified *Nodal* BACs *via* FPE and confirmed BAC functionality by generating fluorescent reporter mice. We also isolated small indel P1-phage derived artificial chromosome (PAC) and BAC recombinants. Statistical analysis revealed that 1:100,000 recombinants can be isolated running <40 PCRs and we developed a web-based calculator to optimize FPE. By eliminating the need for selection-counterselection, this work highlights a straightforward and low-cost approach to BAC mutagenesis, providing a tool for seamless recombineering pipelines in functional genomics.

Recombination-mediated genetic engineering (recombineering) is a technique to manipulate DNA in *Escherichia coli* (*E. coli*) using short homology sequences[1-3]. Bacterial artificial chromosome (BAC) recombineering enables to accurately study gene expression and function[4], but it needs to be seamless to avoid any unnecessary sequence in modified BACs.

Correspondence should be addressed to L.B. (luca.brunelli@genetics.utah.edu).

AUTHOR CONTRIBUTIONS

G.T.L. and L.B. developed the project; G.T.L. and Y. K. performed experiments; P.C.B. carried out the mathematical analysis; A.K. analyzed mice; M.R.K. directed mouse research and provided editing; B.L.D. developed the web calculator; H.J.Y. provided research support, resources and editing; L.B. and G.T.L. wrote the manuscript; L.B. directed overall project.

COMPETING FINANCIAL INTERESTS

The authors have no competing financial interests to disclose.

Currently, seamless mutagenesis requires two steps: isolation of mutants using selectable markers followed by marker removal through counterselection (**Fig. 1**)[5]. This approach has substantial drawbacks because counterselection is relatively complex, prone to false-positives as well as resource and time consuming[6].

Single-stranded oligonucleotide targeting vectors can produce seamless single-base alterations, deletions and short insertions in BACs. Occasionally, these changes can occur at such high frequencies (1:90-260, up to 70%) that recombinants can be isolated without selection[7, 8]. However, these frequencies are not routinely achieved[9], and some DNA inserts, such as those encoding fluorescent protein reporters[10], are too large to be encompassed in commercially available oligonucleotides. Seamless fluorescent reporters are less likely to interfere with transcriptionally active regions, including those located in introns as it is for *Nodal*[11], a critical developmental gene belonging to TGF-β superfamily[12]. The recombination frequency for large inserts can be between one recombinant BAC per 10,000-100,000 electroporated cells (1:10,000-100,000). Current laboratory methods do not allow isolation of these rare recombinants without using selection. Techniques such as clone pooling or cloning by limiting dilution (LDC)[13] can only be used for isolating high-frequency (1:90-260) recombinants[8].

In population genetics, the founder principle is “the establishment of a new population by a few original founders”[14]. In small populations, the proportional decrease of parental alleles is disrupted, resulting in the random elimination or fixation of some alleles[15]. Applying this principle to recombineering, we reasoned that sampling a number of cells below a certain threshold from an *E. coli* parental culture disrupts the proportional decrease of rare recombinant and prevailing non-recombinant BACs. This threshold is the reciprocal of the frequency of recombinant BACs in the parental culture. Under these conditions, we can establish founder cultures with either no recombinants or a higher frequency of recombinants, i.e. enrichment, which can be detected by PCR.

Applying these notions, we developed FPE, a remarkably simple, low-cost and efficient method to isolate rare BAC recombinants without using selectable markers (**Fig. 1**). This concurrently enabled one-step seamless BAC mutagenesis. We isolated seamless fluorescent protein-modified *Nodal* BACs and confirmed their functionality by generating reporter mice. We also successfully used FPE to isolate small indel PAC and BAC recombinants, and we optimized FPE through statistical modeling.

RESULTS

Isolation of rare modified *Nodal* BACs using FPE

Using markerless targeting vectors (**Supplementary Fig. 1**), we generated rare mutant BACs in which the first exon downstream of the *Nodal* start methionine was partially replaced by mStrawberry or EYFP (Methods). To estimate the recombinant frequency (*f*) in recombineering mixes and establish FPE parental cultures, we dispensed 10^5 total cells/well in a single 96-well plate (**Fig. 2**, **Supplementary Fig. 2**). In this plate, PCR detected recombinant-containing wells, identifying positive parental saturated cultures (SCs).

To start enrichment cycles, positive SCs were diluted to establish a diluted culture (DC) containing 10^5 total cells and, presumably, at least one recombinant (**Fig. 2**, **Supplementary Figs. 2** and **3**). 10^5 cells is the FPE threshold for a recombinant frequency of 10^{-5}. Then, we equally divided the DC into a single 96-well plate to establish founder cultures, i.e. diluted and divided cultures (DDCs), with a ~1,000 total cells, i.e. a hundred times below the FPE threshold. After overnight growth, positive SCs were again identified by PCR. The frequency of mutant BACs in PCR-positive wells was at least 1:1,000 (10^{-3}).

FPE was again performed on positive cultures, and at the end of the second cycle the frequency of mutant BACs was increased to ~10^{-1} (**Supplementary Fig. 4**). At this point, ~20-30 PCRs were needed to detect a mutant BAC clone by screening individual colonies on agar plates. Alternatively, a similar number of PCRs was needed for the third enrichment cycle to reach *f* of ~1/1. The first approach was used to isolate mStrawberry and the second one for EYFP, with two-three FPE cycles being completed in about two to four days. These data show that FPE enables the isolation of mutant BACs as rare as 1:100,000 without using selectable markers.

Applicability of FPE

In genetic engineering, it is often necessary to insert or delete *LoxP* and *FRT* sites on/from BAC vector backbones or cloned inserts[16]. Thus, we tested whether FPE can isolate PAC and BAC recombinants generated by single-stranded oligonucleotide-directed recombineering. We started with a mini-P1 phage vector derived from pCYPAC3 (ref. 17) in which a fragment of plasmid PL253 (ref. 18) containing a 20 nucleotide repeat replaced the mini-P1 vector sequence containing *LoxP* and I-SceI. We deleted the 20 nucleotide repeat and re-inserted *LoxP* and I-SceI in different sites of the mini-P1 vector, using FPE to isolate seamless PAC recombinants (**Supplementary Table 1**).

To test FPE in BACs, we deleted *LoxP* from nine distinct BACs, and then isolated recombinant clones using FPE (**Supplementary Table 2**). Notably, eight BACs were simultaneously isolated by one person. Finally, we inserted a 74 nucleotide array consisting of *FRT*, *LoxP* and a restriction endonuclease recognition site into three BACs modified in the previous step, followed by their rapid isolation *via* FPE (**Supplementary Table 2**). At the reported electroporation and recombineering efficiencies, the recombination frequency in these insertion experiments ranged from 3×10^{-2} to 6×10^{-5} (1:30-17,000). In all cases, FPE successfully isolated PAC and BAC recombinants and their structure was confirmed by DNA sequencing.

In few instances we had to perform additional enrichment cycles to isolate recombinants. This did not alter the overall success of FPE. To determine whether the need for additional enrichment cycles could be due to differential growth of recombinants vs. wild types, we reconstituted cell mixes with known frequencies of EYFP mutant BACs (**Supplementary Fig. 5**). We re-isolated recombinant BACs from these mixes without observing any deviation from the expectations of our FPE model. Additionally, there was overlap of the DDC and SC recombinant frequencies in the final enrichment cycle of the experiments reported in Supplementary Tables 1 and 2 (**Supplementary Fig. 6**), and we observed similar

growth rates in cells harboring recombinant and non-recombinant BACs (**Supplementary Fig. 7**). Thus, we speculate that the need to perform additional enrichment cycles could be due to instability of some newly generated recombinants. This is consistent with a report showing that integration of a chloramphenicol marker into *LacZ* can generate both white (Lac^-) and unstable, blue (Lac^+) chloramphenicol-resistant colonies[19]. Overall, these data show that FPE can isolate PAC and BAC indels within a relatively broad range of frequencies and in a variety of DNA engineering strategies.

Testing the functionality of modified *Nodal* BACs

To confirm functional integrity of modified BACs, we generated fluorescent reporter *Nodal* mice. The mStrawberry and EYFP *Nodal* BACs were purified and sequenced to ensure generation of correct structures (Methods). However, since recombineering can generate repeats[19], deletions[6], and plasmid concatenates[20], there could be structural modifications elsewhere in the BAC leading to altered expression of the reporter. Thus, we used the modified BACs isolated *via* FPE to generate transgenic reporter mice and found that mStrawberry and EYFP recapitulate the endogenous *Nodal* expression at embryonic day (e) 6.5 (**Fig. 3a,b**). These experiments show that seamless fluorescent protein-modified *Nodal* BACs generated without selection-counterselection have the correct structure and function.

Computational modeling of FPE

We built a mathematical model to elucidate some FPE features, such as the inclusion of multiple cycles, and how to manage the risk of failure. FPE has three parameters: *r*, the enrichment rate; a, the dilution number; and *b*, the division number (Methods and Supplementary Note), with any two of them being independent. As shown in (7) (all formulas are in Supplementary Note), any recombinant frequency could be obtained in one enrichment cycle by properly selecting the product $a \times b$. However, since *b* was equal to the number of DDCs to be screened by PCR, there was a practical limit of divisions (*B*) that an operator could perform in a day. In addition, attempting higher enrichment by overly increasing *a*, could produce DCs with a low probability of containing mutants. Thus, a was also limited. With both a and *b* limited, their product was often not high enough to achieve the desired recombinant frequency in one cycle, showing that FPE needs to be performed in repetitive cycles.

The number of cells containing mutant BACs followed a binomial distribution in both DCs and DDCs (Supplementary Note). This allowed to calculating the risk of failure (16), which can be managed by changing *b/r*, i.e. the average number of recombinants sampled in DCs (Methods). We derived formulas for the expected number of cycles, and hence number of PCRs to complete FPE, for any given risk of failure (18). For example, when attempting to establish in an enrichment cycle a DC containing one recombinant ($b/r = 1$), the risk of failure was 0.37, and the expected number of cycles to complete this task was 1.6 (1/(1-0.37)). With $b/r = 3$, the risk was 0.05, and the expected number of cycles was 1.05. When $b/r = 3$, we had to screen three times more cultures compared to $b/r = 1$ to achieve the same enrichment, but we ran 1.52 (1.6/1.05) times fewer cycles. Overall, this strategy was less risky, with fewer enrichment cycles compared to sampling one recombinant, although the number of PCRs was double (3/1.52).

Optimizing FPE for high-throughput applications

To make FPE suitable for high-throughput applications, we optimized it by building a FPE cost function. Optimal FPE parameters minimize the relative expense of resources (number of PCRs), α, and time (number of cycles), $1-\alpha$. From (21), the cost function equaled the mean number of cycles to complete FPE when $\alpha = 0$ (time-saving strategy) (19), and the mean number of PCRs to complete FPE when $\alpha = 1$ (resource-saving strategy). At intermediate numbers, i.e. $0<\alpha<1$, the cost function was the sum of the expected number of cycles and PCRs multiplied by $1-\alpha$ and α, respectively.

When time-saving was a priority, $\alpha = 0$, the cost function decreased with increasing *b* (**Fig. 4a)**. When *b* reached its practical limit *B*, there was an optimal enrichment rate minimizing cost function. When *B* = 100, the V shaped cost function curve indicated that sampling two to three recombinants in DCs at each cycle (*b*/*r* = 100/~45 = 2-3) was highly advantageous (**Fig. 4a**). To a lesser degree, this was true for other *B* values up to 400. Thus, when an operator can perform 100-400 PCRs per day, the optimal scenario is sampling two to three recombinant BAC-containing cells in DCs at each cycle.

When resource-saving was a priority, $\alpha = 1$, minimization of the cost function was reached at $r_{min} = \exp(1) = e \approx 2.72$ (see (22)) regardless of *b*, implying that the enrichment rate was constant (~2.72) at any division number, *b*. For example, to increase mutant BAC *f* from 10^{-5} to 10^{-1} using the minimal number of PCRs, the expected number of cycles when sampling one mutant BAC-containing cell in DCs was 14.61 (9.21/1-0.37), and 9.7 (9.21/1-0.05) when sampling three mutants ((18) in which 9.21 is calculated using (10)). The expected minimal number of PCRs was 2.72×1×14.61 = 39.71 and 2.72×3×9.71 = 79.23, respectively. Thus, PCRs could be as low as ~40 for isolating 1:100,000 mutant BACs.

We optimized the cost function for any given α because both time and resources are important for high-throughput applications ($0<\alpha<1$). For this interval, there was a unique pair of FPE parameters ($b_{\min}(\alpha)$, $r_{\min}(\alpha)$) that minimized the cost function (**Fig. 4b,c**). As expected, with α closer to zero, i.e. time-saving strategy, FPE optimization led to increasing the division number *b*; with α closer to one, i.e. resource-saving strategy, it led to an enrichment rate *r* approaching *e* (~2.72) (**Fig. 4b**,**c**). To facilitate determination of the optimal FPE parameters in any working condition, we developed a web calculator (http://yost.genetics.utah.edu/software.php). This mathematical analysis makes useful predictions for industrial level FPE scaling-up, and suggests that FPE optimization can enable efficient genome-wide seamless BAC recombineering pipelines.

DISCUSSION

FPE has a number of important features. First, bacterial cultures at consecutive enrichment cycles are established and grown in such a way that the frequency of recombinants progressively increases in a controllable and predictable manner. Secondly, screening is done on replicas so that the culture never contacts the reagents or factors used for screening[21]. This avoids any disturbance on cell metabolism, and isolated recombinants retain the genuine structures generated by recombineering. Thirdly, FPE does not result in

false positives because PCR can be designed to detect the expected recombineering product. Finally, FPE *per se* is a liquid-dispensing procedure that can be easily robotized.

The isolation of unselected recombinants is a classical problem in recombinant DNA technology[2, 22, 23]. The classical solution has been to link a recombining DNA sequence of interest to selectable markers. Besides the additional engineering work, we and others have shown that selection itself can change the rate of genetic variants[24, 25]. In addition, recent improvements in selection-counterselection for seamless BAC recombineering still require two targeting vectors, five days for each recombineering step, and counterselection using numerous reagents[6]. Instead, FPE can achieve the same goal in five days utilizing only the second of the two targeting vectors (**Fig. 1**).

Our statistical analysis suggests that FPE can substantially improve genome-wide recombineering pipelines. Using selection markers in recombineering pipelines for functional genomics leads to the persistence of some unnecessary DNA sequences which can disrupt normal gene function. Additionally, selection in liquid cultures produces cell mixes containing up to hundreds of independent recombinants[4, 26-31]. This leads to the need of isolating clones, a step which is currently extremely time consuming[29]. In contrast, FPE enriches cultures with seamless recombinants originating from the same, single recombinant event. Barcoded genetic variant pipelines have proven invaluable for multiplex reverse genetic studies[26, 30], but they remain time-consuming. For example, to barcode functional open reading frames (ORFs) in yeast, two unique PCR products are generated for each ORF, followed by joining them with a cloning vector *in vivo*[32]. Similar work in *E. coli* could be performed by directly introducing barcodes *in vivo* through one-step seamless recombineering, followed by rapid recombinant isolation through FPE. FPE could also improve multiplex automated genome engineering (MAGE), a pipeline used to improve bacterial phenotypes through accelerated evolution[33]. MAGE generates an enormous degree of genetic diversity, but it lacks a procedure for sifting out genetic variants that have no effect or worsen the phenotype. Alternating cycles of MAGE with FPE might accelerate the generation of improved bacterial phenotypes. Finally, FPE could be used to isolate any rare genetic variant identifiable in culture with a properly selected screening assay. For example, it could be applied to breed transgenic plants with desired traits without drug and pesticide resistance markers[34].

This work demonstrates that rare BAC recombinants can be isolated without using selectable markers, representing a valid alternative to the classical selection method. Because of its rigorous mathematical basis and statistical optimization, FPE can enable the efficient implementation of genome-wide seamless BAC recombineering pipelines.

ONLINE METHODS

Chemicals and enzymes

Chemicals were purchased from Sigma-Aldrich or Mallinckrodt Baker. All enzymes were purchased from New England Biolabs, except the following polymerases: ampliTaq (Life technologies) and KOD (Clontech Laboratories).

Plasmids

Wildtype *Nodal* BAC, 204 kb RP23-450D10 BAC clone (http://bacpac.chori.org/), was used. Plasmids containing the coding sequences of mStrawberry and EYFP were received from Dr. Vitaly Boyko (National Cancer Institute).

Plasmid DNA isolation and purification

Plasmid DNA was extracted as previously described[35], but the composition of the non-ionic detergent (NID) isolation buffer was improved to promote bacterial debris pelleting and reduce insoluble components in DNA solution. The new formulation is as follows: 0.75 M NH_4Cl, 0.25 M Tris HCl pH 9, 5 mM EDTA, and 0.15% IGEPAL CA-630. *E. coli* chromosomal DNA and endotoxins were removed as previously described[36]. Purified BAC DNA was linearized by PI-SceI homing endonuclease digestion of the pBACe3.6 vector backbone. Linearized DNA was purified using anion exchange plasmid mini kit columns (121123; Qiagen Inc.).

DNA sequencing

DNA sequencing was performed using the ABI BigDye Terminator Cycle Sequencing Kit v1.1 or v3.1 according to the manufacturer's instructions on a Gene Amp 9700 PCR machine. The sequence fragments were detected on an ABI 3130XL Genetic Analyzer. Samples were then analyzed and base-called using the Applied Biosystems DNA Sequencing Analysis Software V5.2. The following primers were used to confirm the structures of the break points: 5'BP:f_Nodal_10317: CGC CCT CTT CTG GAG TGT CTG A; 3'BP:r_Nodal_11870: GCA TCA CTC AGG TCG CTG GGT CAA ACA CA. For recombinants generated by single-stranded oligonucleotide-directed recombineering, the primers used for detecting recombinants were also used to confirm the expected breakpoint structure (see Oligonucteotides section below). Primers for the fluorescent protein coding sequences: f_Str_588: CGC CTA CAT CGT CGG CAT CAA GTT; r_Str_145: CGG TCT GGG TGC CCT CGT A; f_YFP/Str_1: ATG GTG AGC AAG GGC GAG GA; and f_YFP_411: CCT GGG GCA CAA GCT GGA GTA CAA CTA. There was no need to use additional primers for sequencing recombinants generated by single-stranded oligonucleotide-directed recombineering because the modifications encompassed short DNA sequences.

Recombineering

Mini-λ was provided by Dr. Shyam K. Sharan (National Cancer Institute). Electro and recombineering competent DH10B cells harboring wild type *Nodal* BAC were prepared as described previously[37]. 0.1-0.3 μg targeting vector DNA was electroporated at 16,000-18,000 V/cm at time constants of 7.5 -10 msec using BioRad Gene Pulser II or BioRad GenePulser Excell into 20-40 μl competent cells.

Oligonucleotides

All oligonucleotides were desalted and ordered from Invitrogen (Life Technologies), except the primers with ID #11 and #12 which were PAAG purified and ordered from IDT. Except for oligonucleotides used in DNA sequencing, all sequences are reported (**Supplementary**

Table 3). mStrawberry and EYFP targeting vectors replaced 168 bp *Nodal* exon1 sequence: chr10: 61418300-61418467 (NC_000076.6) leaving the last 28 bp of the exon.

PCR

Mixing bacterial cultures by pipetting up and down and sampling bacteria from 96-well plates were done using 12-channel pipettes (manufactured by Capp or Eppendorf). PCR of bacterial cultures: 0.7-1μl of every bacterial culture was transferred into 96-well PCR plates containing 10 μl PCR mix of 1x PCR buffer; 0.2-0.3 μM primers; 0.2 mM dNTP; 0.2 mg BSA/1ml (optional); and 20 U/ml ampliTaq. 5x PCR buffer composition: 100 mM TrisHCl pH 8.8, 50 mM $(NH_4)_2SO_4$, 15 mM $MgSO_4$, 0.5% Triton X-100, 0.25% Tween 20, and 50% glycerol. Addition of tracking dyes to the PCR buffer, such as 0.001% Xylene Cyanol, allows direct gel loading of PCR mixtures. PCR program: a 4 min first hold at 94°C to lyse bacteria and denature DNA, 10-15 sec for annealing primers at annealing temperatures between 55-62°C, and times for primer extension calculated according to a rate of 500 bp/min. The program consisted of 30-45 cycles. The PCR assay was able to detect up to one recombinant per 1,000,000 total cells (**Supplementary Fig. 5**). Gels were loaded using multichannel Impact 2 Matrix pipettes (Thermo Scientific). Targeting vector DNA PCR was performed using KOD or Phusion polymerases and linearized plasmids containing the coding sequences of the fluorescent proteins according to the manufactures' instructions. PCR products were consecutively treated with DpnI and Mung Bean Nuclease, and then purified using agarose gel electrophoresis. The following supplies were used: EU PCR Plate, 96x0.2ml, Non-Skirted (Bioplactics, from MidSci, Cat. B70504-CL); reusable Corning® Thermowell® Sealing Mats (Cat. 6555); 10mM dNTPs from GenScript (C01582-10); PerfectTaq DNA Polymerase from 5 PRIME (Cat. 2200070). The expected cost of major consumables for about 1,000 PCR reactions (10 μl) was ten 96-well PCR plates ($11) + dNTPs ($3.1) + Taq polymerase ($11) ≈ $25.

Establishing parental *E. coli* cultures

The recombineering cell mix contained ~10^9 cells. A fraction of the recombineering mix, containing ~10^7 cells, was mixed with an appropriate volume of growth media so that ~10^5 total DC cells were dispensed into each well of a 96-well cell culture or PCR plates in 12-100 μl aliquots. The LB/Super Broth (1:1) media contained 0.4% glucose, 5mM $MgSO_4$ and 12-15 μg/ml chloramphenicol. The plates were incubated in a PCR machine (for bacteria dispensed in PCR plates) or in an incubator (for bacteria dispensed in 96-well cell culture plates and incubated in moisture-tight containers) at 30-37°C overnight. Both repetitive (Distriman, Gilson Inc.) and multichannel pipettes were used for liquid sampling/dispensing. It was not necessary to outgrow cells before plating due to lack of drug selection in recombinants, but chloramphenicol was added to retain wild type and recombinant BACs in cells. In general, the number of parental cultures in one plate was more than enough to isolate at least one recombinant because the majority of saturated parental cultures were positive (**Supplementary Fig. 8**). If no positive parental cultures are detected, cultures can be reestablished taking greater amounts of the recombineering mix. The saturated parental cultures were also maintained as an independent source of recombinants in case of problems in subsequent steps. As trial and error experiments might be needed to estimate the frequency of recombinants in the recombineering mix, a 10% glycerol stock was prepared

from the rest of the recombineering mix for later use. A 30 min outgrowth of the recombineering mix improved the survival of the frozen cells after thawing.

Before starting screening parental cultures for recombinants, we electroporated control targeting vector DNA to ensure that recombineering has taken place. For selection based controls (**Supplementary Table 2**), the DNA produces colony forming units (CFUs) only in recombineering competent cells; for counterselection based controls (**Supplementary Table 1**), the DNA produces higher CFUs in recombineering competent cells compared to the background CFUs. For example for PCR with primers ID #3 and #4, pUC series plasmid DNA generated about 1 Kb DNA targeting vector with 30 nucleotides homology arms to the BAC vector backbone at its flanks. The targeting vector incorporated the Amp^r marker from the plasmid and upon recombination with the BAC vector backbone the cells lose resistance to chloramphenicol but acquire resistance to ampicillin. The DpnI-treated targeting vector DNA is free of replicative plasmid DNA template contamination and should not produce Amp^r cells when electroporated in DH10B cells with an electroporation efficiency of about 5×10^9 CFU/1μg pUC18. The number of drug resistant colonies in the positive control could also be used for estimating the number of positive parental cultures when plating the recombineering mix.

Computation of the confidence interval

Binomial confidence intervals were calculated using the Clopper-Pearson exact method as implemented in R[1] by the binom package[2]. [1] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. [2] Sundar Dorai-Raj (2014). binom: Binomial Confidence Intervals For Several Parameterizations. R package version 1.1-1. http://CRAN.R-project.org/package=binom.

The basics of FPE

There are two FPE prerequisites. The first one is to retain at least one recombinant in a culture with reduced size. The second prerequisite is that the size of this culture should be below the FPE threshold. These prerequisites allow the recombinant frequency to increase in each successive enrichment cycle (**Fig. 2**).

With f_i being the recombinant frequency in the current cycle and f_{i+1} the one in the next cycle, a specific enrichment rate, r, defined as f_{i+1}/f_i, can be reached by diluting (a) and dividing (b) the culture in the following way. To sample any (n) number of recombinants on average per ml (n/ml) into a DC from a PCR-positive SC with a given number of total cells per ml (N/ml) and a recombinant frequency (f), SC should be diluted $a = N\times f/n$. Dividing 1 ml DC containing n recombinants $n\times r$ times will reduce the total cell number of cells in DDCs to $N/(N\times f/n)/n\times r = 1/f\times r$, which is r times less than the FPE threshold. The positive DDCs most likely have one recombinant among $1/f\times r$ totals and the recombinant frequency in the next cycle is $1/1/f\times r = f\times r$. Thus, to establish positive DDCs enriched r times for recombinants, 1 ml DC containing n recombinants should be equally divided $b = n\times r$. It follows that $b/r = n$, where n is the number of recombinants in DC.

FPE has the interesting feature of being more efficient at low recombinant frequencies. Since $f_{i+1} = 1/N_{i+1}$, the enrichment rate r is equal to $1/f_i \times N_{i+1}$, showing that in order to enrich cultures the denominator $f_i \times N_{i+1}$ should be less than one. This implies that when f_i is low the total number of cells N_{i+1} in DDCs can be high, i.e. requiring less effort to divide and dilute SC. Thus, FPE is more efficient at lower recombinant frequencies.

For the implementation of FPE, the following numbers need to be known: 1) *f*, 2) *N*/ml in SC, and 3) *r*. In positive SCs, *f* is estimated as the reciprocal of the number of colony forming units (CFUs) per DDC. *N*/ml, i.e. SC total cell titer, is determined as CFU number per ml in SCs. This can be determined experimentally before starting the next enrichment cycle (spending one more day), or the currently known total cell titer can be used. We found that with the culture media used in our laboratory *N* has consistently been ~1-3×10^9 CFUs/ml, and the experimental determination of exact *N* does not enrich cultures more efficiently than using an estimated number of 1-2×10^9 for all enrichment cycles. Selecting *r* is connected to selecting *b* and depends on whether FPE is designed to be time and/or reagent-saving (see Computational modeling of FPE in Results), but *b* cannot be greater than the practical limit of the number of screening PCRs, *B*, that an operator can analyze in one day. In our experience, up to 400 PCRs per day are reasonable for a single operator using an electronic multichannel pipette.

At the last enrichment cycle, when *N* is the lowest, an alternative approach to meet the first FPE prerequisite is to use LDC, i.e. establishing DDCs by dividing few differentially diluted DCs (instead of just one). See detailed mathematical analysis of FPE in **Supplementary Note**.

Generation of transgenic mice

All animals (*Mus musculus*) were cared for and used humanely according to the following policies: The US Public Health Service Policy on Humane Care and Use of Animals (1996); the Guide for the Care and Use of Laboratory Animals (1996); and the US Government Principles for Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training (1985). All Frederick National Laboratory for Cancer Research animal facilities and the animal program are accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International. All animal handling and procedures were approved by the Animal Care and Use Committee, NCI-Frederick. Purified modified BACs were diluted in the injection buffer (10 mM Tris, 0.025 mM EDTA) and 1-1.5 ng/μl DNA was microinjected into pronuclei of C57BL/6NCr zygotes, which were then transferred into pseudopregnant B6D2F1 females. In embryos, we confirmed the expected modified BAC structures of the break-points as well as the coding sequences of the fluorescent proteins by PCR and DNA sequencing to ensure that no gross structure rearrangement had occurred upon integration in the mouse genome. For assessing endogenous *Nodal* expression, a minimum of 60 embryos from 3 different injections were analyzed.

Confocal microscopy

The day of vaginal plug was considered as e0.5, and embryos were collected at e6.5 in PBS containing 50% heat inactivated FBS and 25 mM Hepes buffer (pH8.0). Embryos were

transferred in the MatTek 35 mm glass bottom dish no.1.0 (Part #P35G-1.0-14-c) in 250 μl of the above media. Images of the embryos were acquired with a Zeiss LSM510 confocal microscope using a Plan-Neofluar 40x/1.3 oil DIC objective lens.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We acknowledge S.K. Sharan (National Cancer Institute) for providing mini-λ and helpful discussions; M.L. Condic, F.S. Gizatullin, J.T. Hill, and R.A. McKnight for comments on the manuscript; V. Lyozin and B. Stephan for technical help in figure preparation; and V. Boyko (National Cancer Institute) for providing plasmids containing the coding sequences of fluorescent proteins. This work was supported by funds from the Department of Pediatrics (Neonatology) and an American Heart Association grant to L.B. (09BGIA2251076); NHLBI Bench-to-Bassinet Consortium grant to H.J.Y. (U01HL0981); as well as funds from the Intramural Research Program of the US National Institutes of Health, National Cancer Institute to M.R.K. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services and nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

REFERENCES

1. Zhang Y, Buchholz F, Muyrers JP, Stewart AF. A new logic for DNA engineering using recombination in Escherichia coli. Nat. Genet. 1998; 20:123–128. [PubMed: 9771703]
2. Murphy KC. Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. J. Bacteriol. 1998; 180:2063–2071. [PubMed: 9555887]
3. Yu D, et al. An efficient recombination system for chromosome engineering in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 2000; 97:5978–5983. [PubMed: 10811905]
4. Skarnes WC, et al. A conditional knockout resource for the genome-wide study of mouse gene function. Nature. 2011; 474:337–342. [PubMed: 21677750]
5. Sharan SK, Thomason LC, Kuznetsov SG, Court DL. Recombineering: a homologous recombination-based method of genetic engineering. Nat. Protoc. 2009; 4:206–223. [PubMed: 19180090]
6. Bird AW, et al. High-efficiency counterselection recombineering for site-directed mutagenesis in bacterial artificial chromosomes. Nat Methods. 2011; 9:103–109. [PubMed: 22138824]
7. Sawitzke JA, et al. Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. J. Mol. Biol. 2011; 407:45–59. [PubMed: 21256136]
8. Swaminathan S, et al. Rapid engineering of bacterial artificial chromosomes using oligonucleotides. Genesis. 2001; 29:14–21. [PubMed: 11135458]
9. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 2013; 31:233–239. [PubMed: 23360965]
10. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. Science. 1994; 263:802–805. [PubMed: 8303295]
11. Norris DP, Robertson EJ. Asymmetric and node-specific nodal expression patterns are controlled by two distinct cis-acting regulatory elements. Genes Dev. 1999; 13:1575–1588. [PubMed: 10385626]
12. Zhou X, Sasaki H, Lowe L, Hogan BL, Kuehn MR. Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation. Nature. 1993; 361:543–547. [PubMed: 8429908]
13. Taswell C, MacDonald HR, Cerottini JC. Clonal analysis of cytolytic T lymphocyte specificity. I. Phenotypically distinct sets of clones as the cellular basis of cross-reactivity to alloantigens. J. Exp. Med. 1980; 151:1372–1385. [PubMed: 6445930]
14. Mayr, E. Animal Species and Evolution. 2nd edn.. Vol. 211. The Belknap Press of Harvard University Press; 1965.

15. Provine, WB. Genetics, Speciation, and the Founder Principle. Giddings, LV.; Kaneshiro, KY.; Anderson, WW., editors. Oxford University Press; Oxford: 1989. p. 43-76.
16. Branda CS, Dymecki SM. Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. Dev. Cell. 2004; 6:7–28. [PubMed: 14723844]
17. Ioannou PA, et al. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nat. Genet. 1994; 6:84–89. [PubMed: 8136839]
18. Liu P, Jenkins NA, Copeland NG. A highly efficient recombineering-based method for generating conditional knockout mutations. Genome Res. 2003; 13:476–484. [PubMed: 12618378]
19. Poteete AR, Fenton AC, Nadkarni A. Chromosomal duplications and cointegrates generated by the bacteriophage lamdba Red system in Escherichia coli K-12. BMC Mol. Biol. 2004; 5:22. [PubMed: 15596011]
20. Thomason LC, Costantino N, Shaw DV, Court DL. Multicopy plasmid modification with phage lambda Red recombineering. Plasmid. 2007; 58:148–158. [PubMed: 17434584]
21. Lederberg J. Replica plating and indirect selection of bacterial mutants: isolation of preadaptive mutants in bacteria by sib selection. Genetics. 1989; 121:395–399. [PubMed: 2653959]
22. Rothstein RJ. One-step gene disruption in yeast. Methods Enzymol. 1983; 101:202–211. [PubMed: 6310324]
23. Swingle B, et al. Oligonucleotide recombination in Gram-negative bacteria. Mol. Microbiol. 2010; 75:138–148. [PubMed: 19943907]
24. Cairns J, Overbaugh J, Miller S. The origin of mutants. Nature. 1988; 335:142–145. [PubMed: 3045565]
25. Gizatullin FS, Lyozin GT. The origin of His+ revertants of Salmonella typhimurium obtained on selective medium. Res. Microbiol. 1992; 143:711–719. [PubMed: 1488555]
26. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nat. Genet. 1996; 14:450–456. [PubMed: 8944025]
27. Valenzuela DM, et al. High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. Nat. Biotechnol. 2003; 21:652–659. [PubMed: 12730667]
28. Sarov M, et al. A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. Nat Methods. 2006; 3:839–844. [PubMed: 16990816]
29. Poser I, et al. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat Methods. 2008; 5:409–415. [PubMed: 18391959]
30. Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LB, Gill RT. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat. Biotechnol. 2010; 28:856–862. [PubMed: 20639866]
31. Pfander C, et al. A scalable pipeline for highly effective genetic modification of a malaria parasite. Nat Methods. 2011; 8:1078–1082. [PubMed: 22020067]
32. Ho CH, et al. A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. Nat. Biotechnol. 2009; 27:369–377. [PubMed: 19349972]
33. Wang HH, et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature. 2009; 460:894–898. [PubMed: 19633652]
34. Tester M, Langridge P. Breeding technologies to increase crop production in a changing world. Science. 2010; 327:818–822. [PubMed: 20150489]
35. Lezin G, Kosaka Y, Yost HJ, Kuehn MR, Brunelli L. A one-step miniprep for the isolation of plasmid DNA and lambda phage particles. PLoS One. 2011; 6:e23457. [PubMed: 21858126]
36. Lezin G, Kuehn MR, Brunelli L. Hofmeister series salts enhance purification of plasmid DNA by non-ionic detergents. Biotechnol. Bioeng. 2011; 108:1872–1882. [PubMed: 21351074]
37. Court DL, et al. Mini-lambda: a tractable system for chromosome and BAC engineering. Gene. 2003; 315:63–69. [PubMed: 14557065]

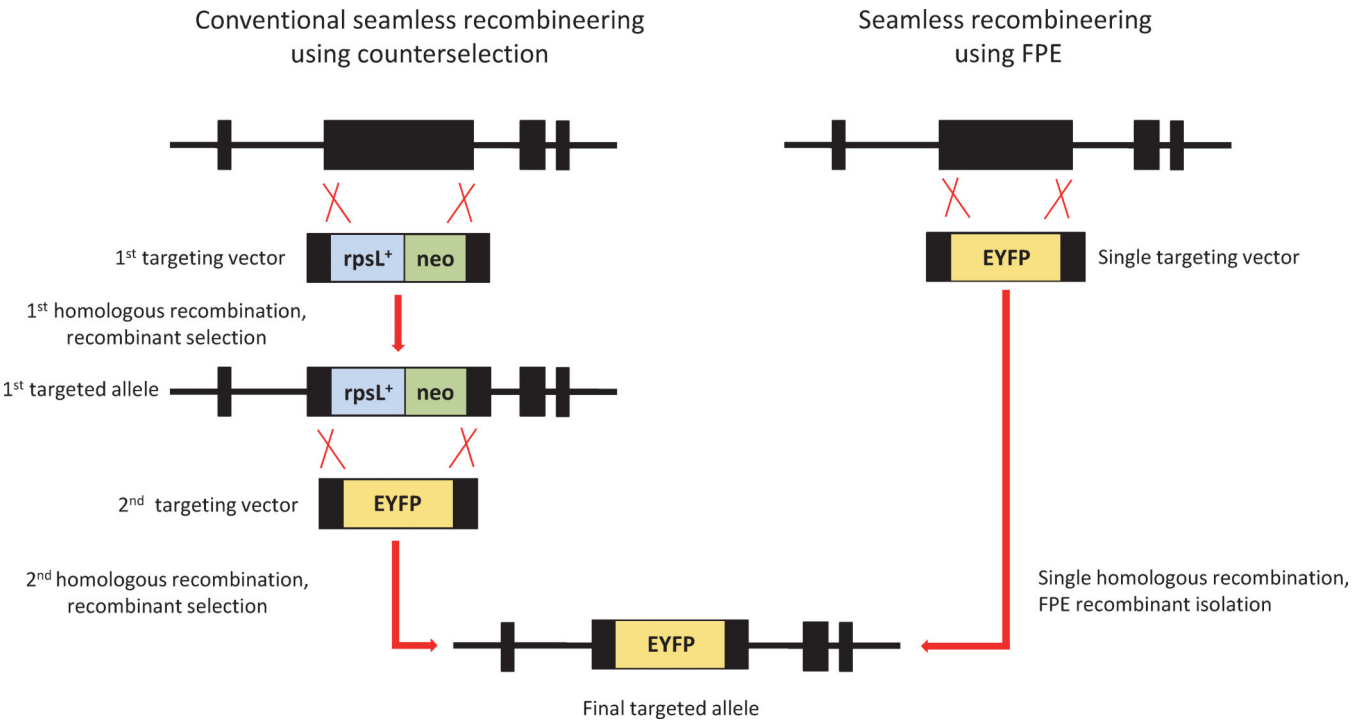

Figure 1.

Seamless BAC recombineering *via* conventional approach or FPE. The conventional method has two steps: 1) selectable (e.g. *neo*) and counterselectable marker (e.g. *rpsL*$^+$) cassettes are introduced in the locus of the DNA modification, followed by isolation of rare recombinants on kanamycin-containing plates; and 2) the whole cassette is replaced by EYFP through recombination without leaving any unnecessary sequences, followed by isolation of rare recombinants on streptomycin-containing plates. *neo*-gene (selection marker) encodes for a protein conferring resistance to aminoglycosides. *rpsL*$^+$ (counterselection marker) is the wild type allele of a gene encoding a ribosomal protein that makes cells sensitive to streptomycin because it is dominant over mutant *rpsL* which enables cells to grow on streptomycin. Cells in which *rpsL*$^+$ is deleted by recombination can grow on streptomycin due to *rpsL*, but some nonrecombinants can grow as well because *rpsL*$^+$ can be inactivated by spontaneous mutations (false positives). In contrast, the new strategy excludes the first step and generates a final targeting allele directly through recombination of the gene of interest with EYFP without introducing any unnecessary sequences. Isolation of rare recombinants containing only EYFP is achieved by FPE. EYFP: DNA sequence encoding the fluorescent protein EYFP.

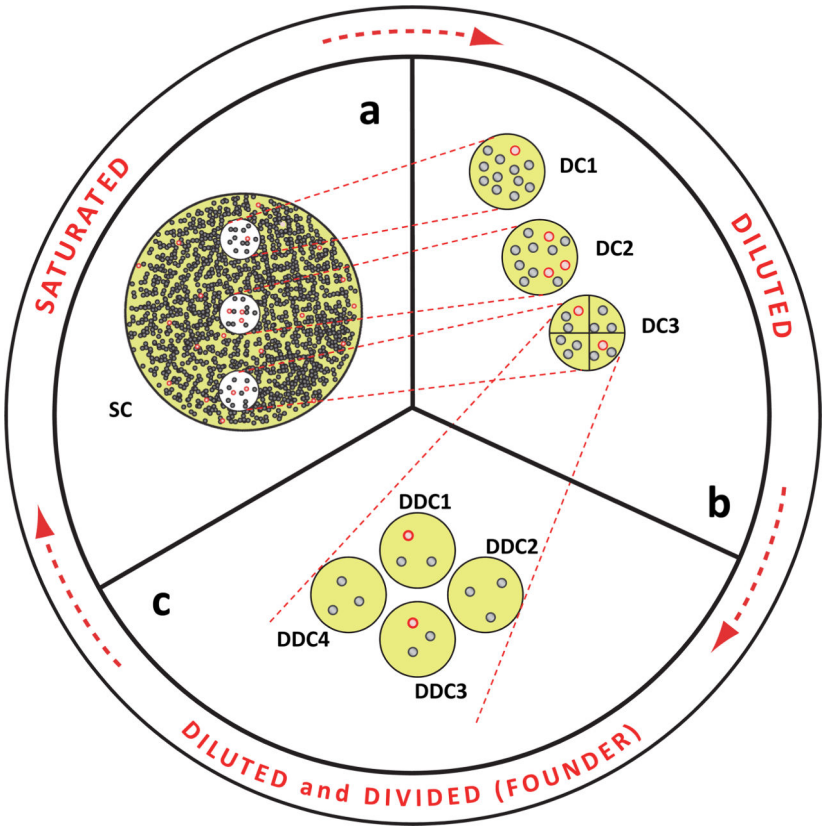

Figure 2.

Conceptual basis of FPE. Red circles: recombinant cells (*n*). Grey circles: prevailing wild type cells. These two sub-populations constitute the cell population with a number (*N*) of total cells. Recombinant frequency $f = n/N$. (**a**) The saturated culture (SC) state has a high bacterial titer. For illustrative purposes, $f_{SC} = 1/6$. The culture is not sensitive to founder principle effects and neither enrichment nor depletion is possible. (**b**) Diluted culture (DC) state. Dilution of SC reduces the number of total cells and brings DC close to the FPE threshold (six cells, $f = 1/6$) where cells become sensitive to founder principle effects upon further size reduction. Founder principle effects start taking place and both enrichment and depletion are possible. However, *f* does not change in a controllable way and there are only frequency fluctuations. (**c**) Founder state, i.e. diluted and divided culture (DDC), where culture size (three cells) is below the FPE threshold (six cells). Here, cultures are either enriched or contain no recombinants. With increasing number of divisions, *f* increases because *N* decreases but *n* in recombinant-containing cultures (positive cultures) is always ≥1. After DDCs reach SC (the culture state cycles back after a sufficient incubation period), positive cultures are identifiable by an appropriate screening procedure. Robust and controllable enrichment of recombinants takes place by cycling cultures through these three states.

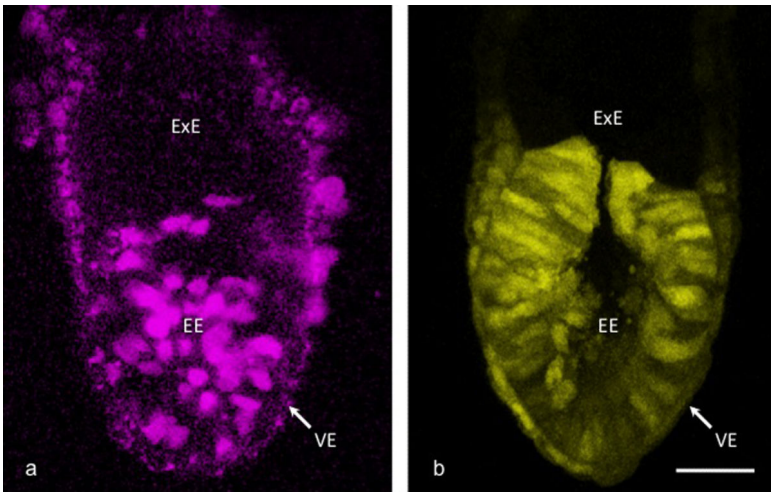

Figure 3.

Functional integrity of fluorescent protein-modified *Nodal* BACs isolated by FPE. (**a**) Analysis by confocal microscopy of the expression pattern of mStrawberry in e6.5 transgenic mouse embryos. (**b**) Similar analysis for EYFP in e6.5 transgenic mouse embryos. Expression of both fluorescent proteins was restricted to the embryonic ectoderm (EE) and visceral endoderm (VE) and absent from the extra-embryonic ectoderm (ExE), recapitulating the embryonic *Nodal* expression. Scale bar: 50 μm.

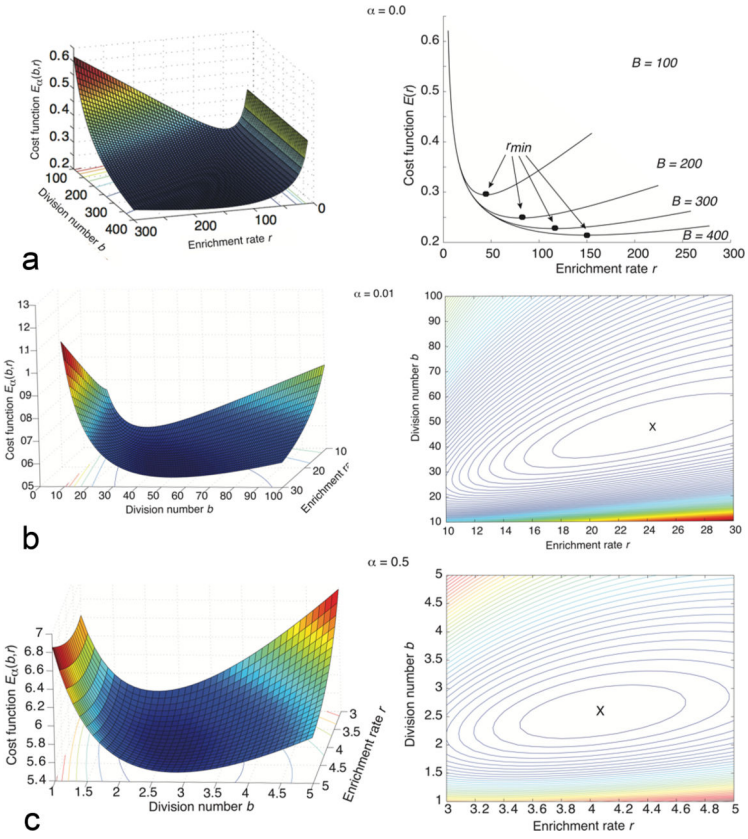

Figure 4.

Cost function analysis of FPE for high-throughput applications. (**a**) Plot of cost function for a 100% time saving strategy ($\alpha = 0$). Left panel: three dimensional (3D) image; right panel: frontal section of 3D image. There is no global minimum of the cost function for all possible b and r. The optimal strategy is to use the highest practical division number, $b = B$, and then minimize the cost function with respect to r. There is a unique enrichment rate for every value of B as shown in the right panel. (**b**) Plot of cost function for a 99% time saving strategy ($\alpha = 0.01$). Left panel: 3D image; right panel: transverse section of 3D image, contours. There is a global minimum as indicated by X ($b_{\min}$ ($\alpha = 0.01$), $r_{\min}$ ($\alpha = 0.01$)) in the contour plot. (**c**) Plot of cost function when time and resource savings are equally important ($\alpha = 0.5$). Left panel: 3D image; right panel: transverse section of 3D image, contours. There is a global minimum as indicated by X in the contour plot. As α increases the minimum occurs at smaller values of b and r.