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

George T Lyozin1, Paul C Bressloff2,5, Amit Kumar3,5, Yasuhiro Kosaka1, Bradley L Demarest4, H Joseph Yost1,4, Michael R Kuehn3 & Luca Brunelli1

Current methods to isolate rare (1:10,000–1:100,000) bacterial artificial chromosome (BAC) recombinants require selectable markers. For seamless BAC mutagenesis, selectable markers need to be removed after isolation of recombinants through counterselection. Here we illustrate founder principle–driven enrichment (FPE), a simple method 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 and confirmed BAC functionality by generating fluorescent reporter mice. We also isolated small indel recombinants without using selection. Statistical analysis revealed that 1:100,000 recombinants can be isolated with <40 PCRs, and we developed a web-based calculator to optimize FPE.

Recombination-mediated genetic engineering (recombineering) is a technique to manipulate DNA in *Escherichia coli* using short homology sequences1–3. BAC recombineering allows the study of gene expression and function4, but it needs to be seamless to avoid any unnecessary sequence in modified BACs. Currently, seamless mutagenesis requires two steps: isolation of mutants using selectable markers followed by marker removal through counterselection5 (Fig. 1). This approach has substantial drawbacks because counterselection is relatively complex, prone to false positives, and resource and time consuming6.

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 (1:90–1:260, up to 70%) that recombinants can be isolated without selection7,8. However, these frequencies are not routinely achieved9, and some DNA inserts, such as those encoding fluorescent protein reporters10, are too large to be encompassed in commercially available oligonucleotides. Seamless fluorescent reporters are less likely to interfere with regulatory elements, including those located in introns as is the case for Nodal11, a critical developmental gene belonging to TGF-β superfamily12. The recombination frequency for large inserts can be between one recombinant BAC per 10,000–100,000 electroporated cells. Current laboratory methods do not allow isolation of these rare recombinants without using selection. Techniques such as clone pooling or cloning by limiting dilution13 can be used for isolating only high-frequency (1:90–1:260) recombinants8.

In population genetics, the founder principle is defined as “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 alleles15. 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 nonrecombinant 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 simple, low-cost and efficient method to isolate rare BAC recombinants without using selectable markers. 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 (Online Methods). To estimate the recombinant
frequency (f) in recombineering mixes and establish FPE parental cultures, we dispensed 10⁵ total cells per well in a single 96-well plate (Fig. 2 and Supplementary Fig. 2). In this plate, PCR detected recombinant-containing wells, identifying positive parental saturated cultures (SCs).

To start enrichment cycles, we diluted positive SCs to establish a diluted culture (DC) containing 10⁵ total cells and, presumably, at least one recombinant (Fig. 2 and Supplementary Figs. 2 and 3). 10⁵ cells is the FPE threshold for a recombinant frequency of 10⁻⁵. Then we equally divided the DC in a single 96-well plate (~10⁻¹) (the second cycle, the frequency of mutant BACs was increased to ~1/1. The first enrichment cycle could be due to differential growth of recombinants generated by single-stranded oligonucleotide–directed recombineering. We started with a mini-P1 phage vector derived from pCYPA3 (ref. 17) in which a fragment of plasmid PL253 (ref. 18) with a 20-nucleotide repeat replaced the mini-P1 phage vector sequence containing loxP and I-SceI. We deleted the 20-nucleotide repeat and reinserted loxP and I-SceI in different sites of the vector, using FPE to isolate seamless PAC recombinants (Supplementary Table 1).

To test FPE in BACs, we deleted vector backbone 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 the cloned DNA of three BACs modified in the previous step, which was 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 (1:30–1:17,000). In all cases, FPE successfully isolated PAC and BAC recombinants and their structure was confirmed by DNA sequencing.

In a 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 versus wild types, we reconstituted cell mixes with known frequencies of EYFP-mutant BACs (Supplementary Fig. 5). We reisolated recombinant BACs from these mixes without observing any deviation from the expectations of our FPE model. Additionally, there was overlap of the confidence intervals between the DDC and SC recombinant frequencies

Figure 1 | Seamless BAC recombineering via the conventional approach and FPE. Neomycin-resistance (neo; selectable) and rpsL+ (counterselectable) marker cassettes are shown.

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 inserts16. 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 pCYPA3 (ref. 17) in which a fragment of plasmid PL253 (ref. 18) with a 20-nucleotide repeat replaced the mini-P1 phage vector sequence containing loxP and I-SceI. We deleted the 20-nucleotide repeat and reinserted loxP and I-SceI in different sites of the vector, using FPE to isolate seamless PAC recombinants (Supplementary Table 1).

To test FPE in BACs, we deleted vector backbone 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 the cloned DNA of three BACs modified in the previous step, which was 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 (1:30–1:17,000). In all cases, FPE successfully isolated PAC and BAC recombinants and their structure was confirmed by DNA sequencing.

In a 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 versus wild types, we reconstituted cell mixes with known frequencies of EYFP-mutant BACs (Supplementary Fig. 5). We reisolated recombinant BACs from these mixes without observing any deviation from the expectations of our FPE model. Additionally, there was overlap of the confidence intervals between the DDC and SC recombinant frequencies

Figure 2 | Conceptual basis of FPE. Red circles, recombinant cells (n): gray circles, prevailing wild-type cells in N total cells. Recombinant frequency f = n/N. (i) The saturated culture (SC) state has a high bacterial titer. The recombinant frequency is illustrated as fSC = 1/6. (ii) Diluted culture (DC) state brings DC close to the FPE threshold (six cells, f = 1/6). At this threshold, cells become sensitive to founder-principle effects upon further size reduction. (iii) Founder state, i.e., diluted and divided culture (DDC), where culture size (three cells) is below the FPE threshold (six cells). Cultures are either enriched or contain no recombinants. After DDCs reach saturation, positive cultures are identifiable by an appropriate screening procedure.
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 nonrecombinant BACs (Supplementary Fig. 7). Thus, we speculate that the occasional 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. 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 (Online Methods). However, because recombination can generate repeats, deletions and plasmid concatenates, 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 6.5 (Fig. 3). 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 (Online Methods and Supplementary Note), with any two of them being independent. As shown in equation (7) (Supplementary Note), any recombinant frequency could be obtained in one enrichment cycle by properly selecting the product $a \times b$. However, because $b$ 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 for calculating the risk of failure (Supplementary Note, equation (16)), which can be managed by changing $b/r$, that is, the average number of recombinants sampled in DCs (Online Methods). We derived formulas for the expected number of cycles, and hence number of PCRs to complete FPE, for any given risk of failure (Supplementary Note, equation (18)).

Optimizing FPE for high-throughput applications
To make FPE suitable for high-throughput applications, we optimized it by building an FPE cost function. Optimal FPE parameters minimize the relative expense of resources (number of PCRs), $\alpha$, and time (number of cycles), $1 - \alpha$. From equation (21) (Supplementary Note), the cost function equaled the mean number of cycles to complete FPE when $\alpha = 0$ (time-saving strategy) (Supplementary Note, equation (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 $\alpha$, 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 the 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 investigator can perform 100–400 PCRs per day, the optimal scenario is sampling two or 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 $r_{\text{min}} = \exp(1) + e = 2.72$ (see Supplementary Note, equation (22)) regardless of $b$, implying that the enrichment rate was constant ($\sim 2.72$) 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 $14.62 (9.21/(1 - 0.37))$ and was $9.69 (9.21/(1 - 0.05))$ when sampling three mutants (Supplementary Note, equation (18), in which 9.21 is calculated using equation (10)). The expected minimal number of PCRs was $2.72 \times 1 \times 14.62 = 39.77$ and $2.72 \times 3 \times 9.69 = 79.07$, respectively. Thus, the number of PCRs could be as low as ~40 for isolating 1:100,000 mutant BACs.

We optimized the cost function for any given $\alpha$ 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 ($b_{\text{min}}(\alpha)$, $r_{\text{min}}(\alpha)$) that minimized the cost function (Fig. 4b,c). As expected, with $\alpha$ closer to 0, i.e., the time-saving strategy, FPE optimization led to increasing the division number $b$; with $\alpha$ closer to 1, i.e., the resource-saving strategy, it led to an enrichment rate $r$ approaching $e$ ($\sim 2.72$) (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#FPE_Calculator; also available online).

Figure 3 | Functional integrity of fluorescent protein–modified Nodal BACs isolated by FPE. (a) Confocal microscopy of the expression pattern of mStrawberry in embryonic day 6.5 (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 extraembryonic ectoderm (EExE), recapitulating the embryonic Nodal expression. Scale bar, 50 µm.
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. Second, screening is done on replicas so that the culture never contacts the reagents or factors used for screening. This avoids any disturbance on cell metabolism, and isolated recombinants retain the genuine structures generated by recombineering. Third, FPE does not result in false positives because PCR can be designed to detect the expected features generated by recombineering. Third, FPE does not result in metabolism, and isolated recombinants retain the genuine structural features of recombinants progressively increases in a controllable and predictable manner.

The isolation of unselected recombinants is a classical problem in recombinant DNA technology. The classical solution has been to link a recombining DNA sequence of interest to selection markers in recombineering pipelines for functional genomics. However, this approach lacks a procedure for sifting the degree of genetic diversity, but it lacks a procedure for sifting genetic variants that have no effect on or that 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, for barcoding functional open reading frames in yeast, two unique PCR products are generated for each open reading frame, after which these products are joined with a cloning vector in vivo. 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.

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, but they remain time consuming. For example, for barcoding functional open reading frames in yeast, two unique PCR products are generated for each open reading frame, after which these products are joined with a cloning vector in vivo. 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.

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 be performed by directly introducing barcodes in vivo through one-step seamless recombineering followed by rapid recombinant isolation through FPE.

Articles
enable the efficient implementation of genome-wide seamless BAC recombineering pipelines.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We acknowledge S.K. Sharan (US National Cancer Institute) for providing mini-L and for 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 Y. Boyko (US National Cancer Institute) for providing plasmids containing the coding sequences of fluorescent proteins. This work was supported by funds from the University of Utah School of Medicine’s Department of Pediatrics (Neonatology) and an American Heart Association grant to L.B. (08BGIA2251076); the US National Heart, Lung, and Blood Institute Bench-to-Bassinet Consortium grant to H.J.Y. (U01HL098111); and the Intramural Research Program of the US National Institutes of Health and 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; nor does mention of trade names, commercial products or organizations imply endorsement by the US government.

AUTHOR CONTRIBUTIONS

G.T.L. and L.B. developed the project; G.T.L. and Y.K. performed experiments; B.L.D. developed the web calculator; G.T.L. and L.B. developed the project; G.T.L. and L.B. provided technical help in figure preparation; and V. Boyko (US National Cancer Institute) for providing plasmids containing the coding sequences of fluorescent proteins. This work was supported by funds from the University of Utah School of Medicine’s Department of Pediatrics (Neonatology) and an American Heart Association grant to L.B. (08BGIA2251076); the US National Heart, Lung, and Blood Institute Bench-to-Bassinet Consortium grant to H.J.Y. (U01HL098111); and the Intramural Research Program of the US National Institutes of Health and 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; nor does mention of trade names, commercial products or organizations imply endorsement by the US government.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Zhang, Y., Buchholz, F., Muysers, J.P. & Stewart, A.F. A new logic for DNA engineering using recombination in Escherichia coli. Nat. Genet. 20, 123–128 (1998).

2. Murphy, H.C. Use of bacteriophage λ recombination functions to promote gene replacement in Escherichia coli. J. Bacteriol. 180, 2063–2071 (1998).

3. Yu, D. et al. An efficient recombination system for chromosome engineering in Escherichia coli. Proc. Natl. Acad. Sci. USA 97, 5978–5983 (2000).

4. Skarnes, W.C. et al. A conditional knockout resource for the genome-wide study of mouse gene function. Nature 474, 337–342 (2011).

5. Sharan, S.K., Thomason, L.C., Kuznetsov, S.G. & Court, D.L. Recombineering: a homologous recombination-based method of genetic engineering. Nat. Protoc. 4, 206–223 (2009).

6. Bird, A.W. et al. High-efficiency counterselection recombineering for site-directed mutagenesis in bacterial artificial chromosomes. Nat. Methods 9, 103–109 (2012).

7. Sawitzke, J.A. et al. Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. J. Mol. Biol. 407, 45–59 (2011).

8. Swaminathan, S. et al. Rapid engineering of bacterial artificial chromosomes using oligonucleotides. Genesis 29, 14–21 (2001).

9. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31, 233–239 (2013).

10. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. & Prasher, D.C. Green fluorescent protein as a marker for gene expression. Science 263, 802–806 (1994).

11. Norris, D.P. & Robertson, E.J. Asymmetric and node-specific nodal expression patterns are controlled by two distinct cis-acting regulatory elements. Genes Dev. 13, 1575–1588 (1999).

12. Zhou, X., Sasaki, H., Lowe, L., Hogan, B.L. & Kuehn, M.R. Nodal is a novel Tgf-β-like gene expressed in the mouse node during gastrulation. Nature 361, 543–547 (1993).

13. Taswell, C., MacDonald, H.R. & Cerottini, J.C. 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. 151, 1372–1385 (1980).

14. Mayr, E. Animal Species and Evolution 2nd edn. 182–214 (Belknap, 1965).

15. Provine, W.B. in Genetics, Speciation, and the Founder Principle (eds. Giddings, L.V., Kaneshiro, K.Y. & Anderson, W.W.) 43–76 (Oxford Univ. Press, 1989).

16. Branda, C.S. & Dynecki, S.M. Talking about a revolution: the impact of site-specific recombinases on genetic analyses in mice. Dev. Cell 6, 7–28 (2004).

17. Ioannou, P.A. et al. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nat. Genet. 6, 84–89 (1994).

18. Liu, P., Jenkins, N.A. & Copeland, N.G. A highly efficient recombineering-based method for generating conditional knockout mutations. Genome Res. 13, 476–484 (2003).

19. Poteete, A.R., Fenton, A.C. & Nadkarni, A. Chromosomal duplications and cointegrates generated by the bacteriophage lambda-red system in Escherichia coli K-12. BMC Mol. Biol. 5, 22 (2004).

20. Thomason, L.C., Costantino, N., Shaw, D.V. & Court, D.L. Multicopy plasmid modification with phage lambda Red recombineering. Plasmid 58, 148–158 (2007).

21. Lederberg, J. Replica plating and indirect selection of bacterial mutants: isolation of preadaptive mutants in bacteria by sbi selection. Genetics 121, 395–399 (1989).

22. Rothstein, R.J. One-step gene disruption in yeast. Methods Enzymol. 101, 202–211 (1983).

23. Swingle, B. et al. Oligonucleotide recombination in Gram-negative bacteria. Mol. Microbiol. 75, 138–148 (2010).

24. Cairns, J., Overbaugh, J. & Miller, S. The origin of mutants. Nature 335, 142–145 (1988).

25. Gizatullin, F.S. & Lyozin, G.T. The origin of His revertants of Salmonella typhimurium obtained on selective medium. Res. Microbiol. 143, 711–719 (1992).

26. Shoemaker, D.D., Lashkari, D.A., Morris, D., Mittmann, M. & Davis, R.W. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nat. Genet. 14, 450–456 (1996).

27. Valenzuela, D.M. et al. High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. Nat. Biotechnol. 21, 651–659 (2003).

28. Sarov, M. et al. A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. Nat. Methods 3, 839–844 (2006).

29. Poser, I. et al. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat. Methods 5, 409–415 (2008).

30. Warner, J.R., Reeder, P.J., Karimpour-Fard, A., Woodruff, L.B. & Gill, R.T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat. Biotechnol. 28, 856–862 (2010).

31. Pfander, C. et al. A scalable pipeline for highly effective genetic modification of a malaria parasite. Nat. Methods 8, 1078–1082 (2011).

32. Ho, C.H. et al. A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. Nat. Biotechnol. 27, 369–377 (2009).

33. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature 460, 894–898 (2009).

34. Tester, M. & Langridge, P. Breeding technologies to increase crop production in a changing world. Science 327, 818–822 (2010).
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), PerfectTaq (5 PRIME) and KOD (Clontech Laboratories).

Plasmids. Wild-type Nodal BAC, 204-kb RP23-450D10 BAC clone (http://bacpac.chori.org/) and BACs reported in Supplementary Table 2 were used. Plasmids containing the coding sequences of mStrawberry and EYPF were received from V. Boyko (National Cancer Institute).

Plasmid DNA isolation and purification. Plasmid DNA was extracted as previously described35, 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 NH4Cl, 0.25 M Tris-HCl, pH 9, 5 mM EDTA and 0.15% (v/v) IGEPAL CA-630. E. coli chromosomal DNA and endotoxins were removed as previously described36. 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 kits (121123; Qiagen).

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:r_Nodal_11870: GCA TCA CTC AGG and f_Nodal_10317: CGC CCT CTT CTG; 3′BP:r_Nodal_11870: GCA TCA CTC AGG and f_YFP/Str_1: ATG 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 primers used for detecting recombinants were also used to confirm the expected breakpoint structure (see “Oligonucleotides below”). Primers for the fluorescent protein–coding sequences were 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 GGT 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 S.K. Sharan (National Cancer Institute). Electro- and recombineering-competent DH10B cells harboring wild-type Nodal BAC were prepared as described previously37. 0.1–0.3 µg targeting vector DNA was electroporated at 16,000–18,000 V/cm at time constants of 7.5–10 ms using Bio-Rad Gene Pulser II or Bio-Rad GenePulser Excvell into 20–40 µl competent cells.

Oligonucleotides. All oligonucleotides were desalted and ordered from Invitrogen (Life Technologies), except the primers with IDs #11 and #12, which were PAAG purified and ordered from IDT. Except for oligonucleotides used in DNA sequencing (see above), all sequences are reported (Supplementary Table 3; mStrawberry and EYPF targeting vectors replaced 168 bp Nodal exon1 sequence chr10: 61418300-61418467 (NC_000076.6), with the fluorescent protein coding sequences 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: 0.7–1 µl of every bacterial culture was transferred into 96-well PCR plates containing 10 µl PCR mix of 1× PCR buffer; 0.2–0.3 µM primers; 0.2 mM dNTP; 0.2 mg/ml BSA (optional); and 20 U/µl AmpliTaq or PerfectTaq. 5× PCR buffer composition was 100 mM Tris-HCl, pH 8.8, 50 mM (NH4)2SO4, 15 mM MgSO4, 0.5% (v/v) Triton X-100, 0.25% (v/v) Tween 20, and 50% (v/v) glycerol. Addition of tracking dyes to the PCR buffer, such as 0.001% (w/v) xylene cyanol, allows direct gel loading of PCR mixtures. The PCR program had a 4-min first hold at 94 °C to lyse bacteria and denature DNA, 10–15 s for annealing primers at temperatures between 55 and 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 manufacturers’ instructions. PCR products were treated consecutively with DpnI and mung bean nuclease and then purified using agarose gel electrophoresis. The following supplies were used: EU PCR Plate, 96 × 0.2 ml, non-skirted (Bioplastics, from MidSci, cat. B70504-CL); reusable Corning Thermowell sealing mats (cat. 6555); 10 mM dNTPs from GenScript (C01582-10); and PerfectTaq DNA Polymerase from 5 PRIME (cat. 2200070). The expected cost of major consumables for about 1,000 PCRs (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 ~10⁹ cells. A fraction of the recombineering mix containing ~10⁷ cells was mixed with an appropriate volume of growth medium so that ~10⁶ total DC cells were dispensed into each well of a 96-well cell culture or PCR plates in 12- to 100-µl aliquots. The LB/Super Broth (1:1) medium contained 0.4% glucose, 5 mM MgSO4 and 12–15 µg/ml chloramphenicol. The plates were incubated in a PCR machine (for bacteria dispersed in PCR plates) or in an incubator (for bacteria dispersed in 96-well cell culture plates and incubated in moisture-tight containers) at 30–37 °C overnight. Both repetitive (Distriman, Gilson) and multichannel pipettes were used for liquid sampling/dispensing. It was not necessary to outgrow cultures before plating owing 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 there were some positive parental cultures (Supplementary Figs. 8 and 9). If no positive parental cultures are detected, cultures can be reestablished with greater amounts of the recombineering mix. The saturated parental cultures were also kept as an independent
source of recombinants in case of problems during 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 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 2–3 orders more CFUs in recombineering competent cells compared to the background CFUs. For example, for PCR with primer IDs #3 and #4, pUC series plasmid DNA generated about 1 kb of DNA targeting vector with 30-nucleotide homology arms to the BAC vector backbone at its flanks. The targeting vector incorporated the Amp<sup>r</sup> 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<sup>r</sup> cells when electroporated in DH10B cells with an electroporation efficiency of about 5 × 10<sup>9</sup> CFU per microgram 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 binomial confidence intervals

To determine whether recombinant cells differ in growth rate from nonrecombinant cells, we calculated the confidence intervals of the frequency of recombinants for 15 different constructs both before (in DDC) and after overnight growth (in SC) (Supplementary Tables 1 and 2 and Supplementary Fig. 6). In all 15 cases, confidence intervals overlap broadly, which is consistent with our hypothesis that recombinant genotype does not affect growth rate. Because our predicted effect size was 0, we did not perform a power analysis. We chose instead to rely upon overlapping confidence intervals to illustrate the negative results. Binomial confidence intervals were calculated using the Clopper-Pearson exact method as implemented in R (ref. 38; http://www.r-project.org/) by the “binom” package (http://CRAN.R-project.org/package=binom). This method assumes only that the data are binomially distributed, an assumption that is well justified given that we have counts of bacterial colonies where each is assessed for a binary outcome (recombinant vs. nonrecombinant). Equal variance is not an assumption of the Clopper-Pearson method. Variance for each construct is not explicitly estimated but is well illustrated by the confidence intervals.

### 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 \( N/(N \times f/n) \times n \times r = 1/f \times r \), 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 \( 1/(1/f \times r) = f/r \). 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 1. This implies that when \( f_i \) is low, the total number of cells \( N_{i+1} \) in DDCs can be high, requiring less effort to divide and dilute SC. Thus, FPE is more efficient at lower recombinant frequencies.

For the implementation of FPE, the following values need to be known: (i) \( f \), (ii) \( N/ml \) in SC and (iii) \( r \). In positive SCs, \( f \) is estimated as the reciprocal of the number of 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 medium used in our laboratory, \( N \) has consistently been \(~1-3 \times 10^9\) CFUs/ml, and the experimental determination of exact \( N \) does not enrich cultures more efficiently than using an estimated number of \( 1 \times 10^9 \)– \( 2 \times 10^9\) 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 1 d. 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 cloning by limiting dilution (LDC), i.e., establishing DDCs by dividing few differentially diluted DCs (instead of just one). See the detailed mathematical analysis of FPE in the 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 breakpoints 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 (pH 8.0). Embryos were transferred in MatTek 35-mm glass-bottom dish no. 1.0 (#P35G-1.0-14-c) in 250 µl of the above medium. Images of the embryos were acquired with a Zeiss LSM510 confocal microscope using a Plan-Neofluor 40×/1.3–numeral aperture oil DIC objective lens.

35. Lezin, G., Kosaka, Y., Yost, H.J., Kuehn, M.R. & Brunelli, L. A one-step miniprep for the isolation of plasmid DNA and lambda phage particles. *PloS ONE* **6**, e23457 (2011).
36. Lezin, G., Kuehn, M.R. & Brunelli, L. Hofmeister series salts enhance purification of plasmid DNA by non-ionic detergents. *Biotechnol. Bioeng.* **108**, 1872–1882 (2011).
37. Court, D.L. *et al.* Mini-λ: a tractable system for chromosome and BAC engineering. *Gene* **315**, 63–69 (2003).
38. R Development Core Team. *R: A Language And Environment For Statistical Computing* (R Foundation for Statistical Computing, 2013).