

# 1 Deep-mutational scanning libraries using tiled-region exchange 2 mutagenesis

3  
4 **Kortni Kindree<sup>1,\*</sup>, Claire A. Chochinov<sup>1,\*</sup>, Keerath Bhachu<sup>1,\*</sup>, Yunyi Cheng<sup>2</sup>, Amelia  
5 Caron<sup>1</sup>, Molly McDonald<sup>3</sup>, Zaynab Mamai<sup>3</sup>, Alex N. Nguyen Ba<sup>1,3,&</sup>**

6 <sup>1</sup>*Department of Cell and Systems Biology, University of Toronto, Toronto ON Canada, M5S*

7 *3G5* <sup>2</sup>*Department of Computer Science, University of Toronto, Toronto ON Canada, M5S*

8 *2E4* <sup>3</sup>*Department of Biology, University of Toronto at Mississauga, Mississauga ON Canada,*  
9 *L5L 1C6*

10 *\*These authors have contributed equally to this work*

11 *&Corresponding author*

12 Email address of corresponding author: *alex.nguyenba@utoronto.ca*

13 Running head: Tiled-region exchange mutagenesis

14 Keywords: Deep-mutational scanning, cloning

## 15 **ABSTRACT**

16 The analysis of gene function frequently requires the generation of mutants. Deep-  
17 mutational scanning (DMS) has emerged as a powerful tool to decipher important functional  
18 residues within genes and proteins. However, methods for performing DMS tend to be  
19 complex or laborious. Here, we introduce Tiled-Region Exchange (T-REx) Mutagenesis, which  
20 is a multiplexed modification of the EMPIRIC mutagenesis approach. Self-encoded removal  
21 fragments are cloned in parallel in non-overlapping gene locations and pooled. In a one-pot  
22 reaction, oligonucleotides are then swapped with their corresponding self-encoded removal  
23 fragments in bulk using a single Golden Gate reaction. To aid in downstream phenotyping,  
24 the library is then fused with unique DNA barcodes using the Bxb1 recombinase. We  
25 demonstrate this approach and its optimizations, to show that it is both easy to perform and  
26 efficient. This method offers simple and expedient means to create comprehensive  
27 mutagenesis libraries.

## 28 **INTRODUCTION**

29 In the field of genomics and proteomics, assessing the functional impacts of mutations has  
30 been a crucial area of research. Efficient strategies for site-directed mutagenesis have  
31 allowed improved understanding of catalytic enzymes(Plapp 1995), binding  
32 interfaces(Casipit et al. 1998), and regulatory elements(Pattanaik et al. 2010). Standard  
33 methods for site-directed mutagenesis frequently assess mutations at specific sites to  
34 specific residues and therefore only offer a small glimpse of possible mutations on a gene of  
35 interest(Huttanus et al. 2023). In proteins, a less biased method is alanine scanning(Weiss  
36 et al. 2000), which aims to decipher the functional contribution of each native amino acid in

© The Author(s) 2026. Published by Oxford University Press on behalf of Genetics Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

1 a protein. More recently, alanine scanning has been supplanted by deep-mutational  
2 scanning (DMS) where the activity of each possible amino acid is compared at each position  
3 in a protein(Fowler and Fields 2014). These screens offer an extremely rich view of the  
4 functional and evolutionary constraints on proteins as amino acids of similar biochemical  
5 properties can be leveraged to obtain deeper insight on functional residues(Fowler and  
6 Fields 2014; Fowler et al. 2023). However, one major challenge with obtaining these  
7 improved mutagenesis screens is that deep-mutational scanning is far less accessible to  
8 current molecular biology labs due to cost and difficulty of production(Wei and Li 2023). This  
9 limitation has precluded the genome-scale analysis of variants.

10 To capture a complete set of comprehensive variants for a gene of interest, the DMS  
11 approach must be unbiased and provide a saturated library pool(Rubin et al. 2017; Coyote-  
12 Maestas et al. 2020a). Ideally, every possible amino acid change across the target gene or  
13 region should be present, encompassing all variants identified through evolutionary  
14 comparisons or variant screening and large-scale sequencing efforts. Moreover, each  
15 variant created in the pool should be equally represented before having undergone any  
16 phenotypic assays. This approach must also have the capability of being performed in bulk,  
17 using high-throughput screening techniques, while being scalable to any desired gene size.

18 Several methods have been previously developed for generating variant libraries to assess  
19 mutations in bulk including parallel site-directed mutagenesis (SDM)(Watanabe et al. 2021),  
20 error-prone PCR (ep-PCR), Saturated Programmable Integration of Elements  
21 (SPINE)(Coyote-Maestas et al. 2020b), and various oligonucleotide-based approaches  
22 (Kunkel(Kunkel 1985), PFunkel(Firnberg and Ostermeier 2012), PALS(Kitzman et al. 2015),  
23 POPcode(Weile et al. 2017), and EMPIRIC(Hietpas et al. 2011)). These oligonucleotide-  
24 based approaches have become more popular in recent years due to the rapidly decreasing  
25 cost of oligo pool synthesis. While these methods all have their advantages, to our  
26 knowledge no existing method is both easy to perform and generates all possible single  
27 amino acid variants or missense mutations in a gene, while simultaneously ensuring that  
28 each plasmid copy contains only one single amino acid variant, and that the wildtype  
29 sequences are minimized in this library.

30 To address these current limitations, we have developed a simple, novel DMS approach that  
31 multiplexes the EMPIRIC methodology, called Tiled-Region Exchange (T-REx) Mutagenesis  
32 (Figure 1). T-REx aims to achieve a high mutational efficiency while multiplexing the workflow  
33 within a streamlined one-pot reaction, for the purpose of creating a library of evenly  
34 represented single amino acid mutations along a gene of interest. This approach also aims  
35 to increase throughput, while decreasing the off-target mutation rate, to improve scalability  
36 for variant effect mapping and functional interpretation. Here, we outline our novel  
37 approach, its advantages, and implications as an integral DMS approach that generates a  
38 variant library enriched in mutations which may inform biomedical research and be relevant  
39 to disease.

40

## 41 MATERIALS AND METHODS

## 1 **In silico data-assisted design**

2 To multiplex the EMPIRIC mutagenesis method, self-encoded removal fragments (SERFs)  
3 are cloned in non-overlapping locations (or tiles) in a gene in parallel, and then subsequently  
4 pooled followed by exchange with desired fragments using simultaneous Golden Gate  
5 assembly in a one pot reaction, where T4 ligase and BsaI restriction endonuclease work in  
6 tandem for scarless ligation reactions over non-palindromic overhangs. Reaction fidelity is  
7 ensured by the specific overhangs used in this Golden Gate reaction, but minimizing off-  
8 target ligations is still desirable by choosing tile locations that produce overhangs with  
9 minimal crosstalk.

10 To select these tile locations, we took inspiration from the Data-optimized Assembly Design  
11 process that was developed for simultaneous gene assembly of 52 fragments (Potapov et al.  
12 2018; Pryor et al. 2020). We begin by randomly choosing tile boundary locations and  
13 calculating an objective function, which scores the probability of on-target assembly, the  
14 probability of off-target assembly, the presence of palindromes, the presence of vector-only  
15 relegation, and the tile-size variance. Tile positions are then randomly moved in a local region  
16 and positions that maximize this objective function are maintained. After a few iterations, we  
17 obtain sets of tile locations with compatible overhangs that can be assembled in one-pot. A  
18 script implementing this algorithm can be obtained at: <https://github.com/annb-lab/TRex>.

## 19 **Library generation and cloning**

### 20 *ccdB* entry vector cloning

21 Tiled *ccdB*-containing plasmids were generated with parallel PCR reactions, where the  
22 *Escherichia coli* plasmid containing a gene of interest was first amplified to exclude the tile  
23 region. Each of these PCR reactions were performed separately, so that there was an  
24 amplified fragment for each tile. The primers used for these PCRs contained a  
25 complementary overhang to the *ccdB* PCR fragment, which were then cloned with Gibson  
26 Assembly (Gibson et al. 2009) in a 10 µl reaction by combining equal amounts of the *ccdB*  
27 fragment and backbone PCR fragment. Reactions were then incubated at 50 °C for 1 hour on  
28 the thermocycler. 1 µl of this assembled product was transformed into *ccdB* Survival 2  
29 competent cells (Invitrogen), which allow for the otherwise toxic *ccdB* fragment to be  
30 propagated.

### 31 Oligo extension

32 Oligos were ordered from IDT (Integrated DNA Technologies) as standard desalted oligos, as  
33 Ultramers, or as oPools (oligo pools from IDT, which use the Ultramer synthesis platform).  
34 They were then resuspended in 10 mM Tris, 0.1 mM EDTA pH 8.0 to 10 µM. Conversion of the  
35 single-stranded oligonucleotide to double-stranded DNA and preparation of the *E. coli*  
36 library was performed as in previously with small modifications (which will be described in  
37 their respective sections) (Nguyen Ba et al. 2019). In some trials, gBlocks would substitute  
38 for these Ultramers, with specific introduced mutations, for a site-directed mutagenesis  
39 approach (three to four mutations in different tiles can be ordered as a single 500 bp gBlock)  
40 in a more cost effective and simple manner.

## 1 NEBridge and transformation

2 Golden Gate assembly(Engler et al. 2009) was performed using a cycling method with a  
3 modified reaction buffer that shows higher cloning efficiency. We usually used about 500 ng  
4 of *ccdB*-entry vector, and 1  $\mu$ l of purified dsDNA oligonucleotide in a 20  $\mu$ l reaction. We used  
5 the following cycling protocol: 1) 37°C for 1 minute, 2) 16°C for 1 minute, 3) GOTO 1, 60 times,  
6 4) 50°C for 5 minutes, 5) 12°C for infinite time. We used a 5x assembly buffer: 200  $\mu$ l of 10x  
7 Cutsmart buffer, 20  $\mu$ l of 100 mM ATP, 20  $\mu$ l of 1 M DTT, 60  $\mu$ l of propylene diol, 100 mg of PEG  
8 8000 and water to 400  $\mu$ l. Our 20x enzyme mix was: 10  $\mu$ l T4 ligase (2000 U/ $\mu$ l), 30  $\mu$ l Bsal-  
9 HFv2 (20 U/ $\mu$ l). The product of the cycling reaction can be directly transformed into  
10 homemade chemical competent cells(Inoue et al. 1990), and we routinely obtain  $10^5$ - $10^6$   
11 cfu. from this transformation.

## 12 Bxb1-integrase fusion

13 To barcode the deep-mutational scanning library, barcodes were cloned in a plasmid  
14 containing kanamycin resistance and an R6K origin of replication, which requires a strain  
15 containing the Pir element for propagation(Metcalf et al. 1994). Barcodes were cloned as in  
16 previously, except using strains BAN004 (DH10B, *uidA::pir116*) and BAN005 (*ccdB* Survival  
17 2, *uidA::pir116*) for plasmid propagation. The plasmid also contained a Bxb1 attP site, while  
18 the variant library, containing ampicillin resistance and a pUC origin of replication,  
19 contained a Bxb1 attB site(Singh et al. 2013). A Bxb1 fusion reaction was prepared as  
20 follows: 400 ng of the plasmid containing the mutagenized library, 200 ng of the barcode  
21 library, 2  $\mu$ l of 5x Bxb1 buffer (250 mM Tris pH 8, 250 mM KCl, 375 mM NaCl, 5 mM EDTA, 500  
22  $\mu$ g/mL BSA, 25% PEG 8000), 1  $\mu$ l of Bxb1 integrase enzyme (at approximately 0.5 mg/mL) and  
23 topped with water to 10  $\mu$ l. The reaction was then incubated at 37 °C for 2 hours in a  
24 thermocycler. Gel electrophoresis verification of the fusion reaction indicated that the  
25 reaction is essentially complete within 2 hours at 37 °C. The reaction was then transformed  
26 directly into chemical competent *E. coli* cells and recovered in LB+Amp+Kan to select for  
27 cells that received the fused plasmid only.

## 28 Purification of Bxb1 integrase

29 Purification of the Bxb1 integrase was performed using a Bxb1 integrase construct that was  
30 C-terminally tagged with a 6xHis. This construct was expressed in BL21(DE3) *E. coli* cells  
31 that were grown in ZYM505 medium(Studier 2005). The cultures were induced with IPTG at  
32 0.25 mM at mid-log phase ( $OD_{600} \sim 0.4$ – $0.8$ ) and incubated overnight at 18°C with shaking.  
33 Cells were then harvested by centrifugation and resuspended in lysis buffer (20 mM Tris (pH  
34 8), 1 M NaCl and 5% (v/v) glycerol). After cell lysis by sonication and centrifugation to remove  
35 cell debris, the clarified lysate was loaded into a nickel-nitrilotriacetic acid (Ni-NTA) resin  
36 column, which was pre-equilibrated with lysis buffer + 10mM Imidazole. The column was  
37 then washed with the same lysis buffer + 10mM Imidazole, to remove excess non-specific  
38 binders and outcompeting weakly bound proteins, before eluting the 6xHis-tagged Bxb1  
39 integrase with lysis buffer + 250 mM Imidazole. The eluted protein then underwent a buffer  
40 exchange to completely remove the imidazole and concentrate the protein to 1 mg/ml. The  
41 purified protein was then diluted with equal volumes of glycerol (final 50% v/v glycerol),

1 reaching a final concentration of ~8.5  $\mu$ M and stored at  $-20^{\circ}\text{C}$ . We did not remove the 6xHis-  
2 tag from the Bxb1 integrase protein since the tag is small (~0.8 kDa) and has not been seen  
3 to interfere with the robust, high-yielding Bxb1 reaction efficiency.

#### 4 Yeast transformations

5 After barcoding the library, the purified plasmid pools can be transfected into a model  
6 organism of choice. We used a standard Lithium Acetate (LiOAc) and 50% PEG protocol to  
7 transform our libraries into yeast (Gietz 2014). An overnight culture of the desired yeast strain  
8 was grown to saturation in YPD (2% peptone, 1% yeast extract, 2% (w/v) glucose), and 150-  
9 300  $\mu$ l was transferred into 5 ml of fresh YPD and grown for 4.5-5 hours at  $30^{\circ}\text{C}$  until an  
10 optimal OD600 of 0.4-0.6 was reached. The cells were then pelleted, washed, and the  
11 following were added on top of the cells: 240  $\mu$ l of 50% PEG 3350, 36  $\mu$ l of 1 M Lithium  
12 Acetate, 50  $\mu$ l denatured Salmon Sperm DNA (2 mg/mL, boiled 5 min, snap cooled on ice),  
13 10-20  $\mu$ l of PmeI digested mutagenized barcoded library. The PmeI digestion was done to  
14 release the mutagenized fragment with regions of homology for integration in the yeast  
15 genome. It is specific to our experiment and other means of transfection are possible. The  
16 reaction was then vortexed until homogenized and left to incubate in a  $42^{\circ}\text{C}$  heat bath for 1  
17 hour. The cells were then pelleted, the supernatant was removed, and the pellet was  
18 resuspended in 1 ml of ddH<sub>2</sub>O before being plated on standard dropout plates and incubated  
19 for 2-3 days at  $30^{\circ}\text{C}$ .

#### 20 **Sequencing library preparation**

##### 21 gDNA extraction

22 After growing the transformed yeast pools to saturation, the genomic DNA was extracted  
23 using the following gDNA extraction protocol. 1-3 mL was spun down and the supernatant  
24 was removed. The cell pellet was resuspended in 100  $\mu$ l of yeast lysis extraction buffer (5  
25 mg/mL Zymolyase 20T (100 U/mL), 100 mM Sodium Phosphate buffer pH 7.4 (0.5 M  
26 Na<sub>2</sub>HPO<sub>4</sub> and 0.5 M NaH<sub>2</sub>PO<sub>4</sub>), 10 mM EDTA, 0.5% SB3-14, 200  $\mu$ g/mL Rnase A, 1 M  
27 Sorbitol, 20 mM DTT, stored at  $-20^{\circ}\text{C}$ ) and placed at  $37^{\circ}\text{C}$  for 30 minutes or more until  
28 complete lysis. After lysis, 400  $\mu$ l of lysis/binding buffer (100 mM MES pH 5, 4.125 M  
29 Guanidine thiocyanate, 25% isopropanol, 10 mM EDTA) was added to the tube, and vortexed  
30 until all precipitates were dissolved. The tube was spun down for 30 seconds if unlysed cells  
31 remained. The supernatant of the tube was passed onto a standard miniprep silica column  
32 and spun for 30 seconds to pass the supernatant through the column. 1x wash with 400  $\mu$ l  
33 Wash buffer 1 (10% GuSCN, 25% Isopropanol, 10 mM EDTA) followed by 1x wash with 600  $\mu$ l  
34 10 mM Tris/80% Ethanol was performed and spun for 30 seconds after each wash. The  
35 column can then be dried by spinning for 3 minutes at maximum speed. The gDNA was then  
36 eluted with 50  $\mu$ l of elution buffer (10 mM Tris-HCl, pH 8.5), where the expected DNA  
37 concentration was around 20-30 ng/ $\mu$ l for a 1 mL culture. Success of the extraction was  
38 verified by agarose gel electrophoresis.

##### 39 PCR and sequencing

1 Unique barcodes linked with mutated codons can be sequenced using long-read  
2 sequencing or with Illumina in the case of short libraries. This can be done with PCR  
3 (necessary for libraries integrated inside model organisms), or by restriction digest followed  
4 by adapter ligations according to the recommendations of the sequencing platforms. Our  
5 FKBP1a library was sequenced on a MiSeq v2 500-cycle kit using PCR protocols as  
6 previously described in a previous study (Nguyen Ba et al. 2019).

## 7 **Barcode association bioinformatics**

### 8 Read extraction of genes and barcode sequences

9 Reads were de-multiplexed from inline indexes and barcodes were analyzed as in a previous  
10 study (Nguyen Ba et al. 2019). To extract the gene from the reads, 20 bp anchors  
11 corresponding to the promoter and terminators were used, allowing 2 mismatches per  
12 anchor. The barcode was similarly extracted, removing fixed bases included in the barcode  
13 to prevent BsaI restriction enzyme sites.

### 14 Barcode clustering

15 Sequenced barcodes contain a mixture of barcodes with and without sequencing errors.  
16 Assuming that most sequencing reads have no base calling mistakes on the barcodes, we  
17 can perform clustering and error correction to obtain a final set of barcodes. This was  
18 performed essentially as in (Nguyen Ba et al. 2019), by sorting barcodes by their counts, and  
19 error correcting barcodes starting from least common to most common, using a threshold  
20 of 2 Hamming distance. Finally, barcodes that were observed fewer than 10 times in the  
21 whole sequencing library were removed from further analyses.

### 22 Mutation association

23 To associate mutations within the gene to a barcode, we tabulated for each barcoded read  
24 the list of mutations detected. Each read can contain, in principle, four different types of  
25 mutations: 1) the correct codon that was mutated by Tiled-Region Exchange mutagenesis,  
26 2) an oligo synthesis mistake that is associated with the mutated codon, 3) a sequencing  
27 error due to basecalling, and 4) a sequencing error due to PCR chimeras. Because oligo  
28 synthesis errors will occur at the same frequency as the desired mutated codon (forming a  
29 mutation set), we take advantage of the Apriori algorithm (Odland 2025) to identify the most  
30 common set of mutations for each barcode at the nucleotide level. Given a total list of  
31 possible mutations associated with a barcode, the Apriori algorithm can return, for each  
32 possible set of mutations, the most common to least common sets. Sequencing errors  
33 occur usually as singletons and can be filtered out by a simple frequency threshold.  
34 However, PCR chimeras occur during the amplification reaction and can therefore appear in  
35 several reads. To investigate the effect of PCR chimeras in our analysis, we first cumulated  
36 the frequency of the most common mutation set and the second most common mutation  
37 set for each barcode (assuming that the most common mutation set is the real mutation  
38 associated with a specific barcode). We found that the most common mutation set was  
39 usually found at over 30% of the reads, while the second most common mutation set was  
40 usually below 10% of the reads. Thus, implementing a mutation set threshold of 30% and a

1 minimum read count of 10, both accelerates the Apriori algorithm and likely returns true  
2 barcode-mutation associations.

3 If two mutations were found at over 30% of reads, this would indicate barcodes that are  
4 shared between different mutations. These were discarded from our analysis, as well as  
5 barcodes that were associated with an indel.

## 6 **RESULTS AND DISCUSSIONS**

### 7 **Efficient and comprehensive deep-mutational scanning libraries using tiled Golden- 8 Gate assembly**

9 There is no shortage of approaches for systematically mutagenizing genes of interest on a  
10 plasmid. The conceptually simplest approach to do this is by performing standard site-  
11 directed mutagenesis (SDM) at every desired position, however this is extremely  
12 laborious (Watanabe et al. 2021). Higher throughput methods have used mutagenic  
13 oligonucleotides and a polymerase, or error-inducing enzymes (such as error-prone  
14 PCR (Ossa-Hernández et al. 2024) or fusion of T7 RNA polymerase with AID (Ali et al. 2025)).  
15 All these approaches, however, require fine-tuning the Poisson rate of introduced mutations  
16 and cannot guarantee that all constructs contain a single mutation (and not zero or more  
17 than one). In contrast, the EMPIRIC (Extremely Methodical and Parallel Investigation of  
18 Randomized Individual Codons) method is an oligonucleotide-based approach that clones  
19 the designed oligonucleotides directly with the use of a ligase (Hietpas et al. 2011) and thus  
20 falls into a different class of mutagenesis techniques that can guarantee the final product.  
21 In EMPIRIC, plasmids receiving the oligonucleotide cassettes have a Self-Encoded Removal  
22 Fragment (SERF) flanked by inverted BsaI restriction sites (Hietpas et al. 2011). As such,  
23 plasmids can be treated enzymatically with BsaI to ligate oligo fragments flanked by sticky  
24 ends that have been designed to ligate at the desired position in the target vector. Where the  
25 EMPIRIC method falls short is that it only mutates a single region of a gene. However, it is  
26 trivial to imagine how the EMPIRIC approach can be parallelized to mutagenize a complete  
27 gene.

28 In light of this, the theoretical ideal approach would generate one and only one mutation,  
29 with high mutational efficiency in a one-pot reaction. This would allow the production of  
30 comprehensive libraries of all possible missense variants in a gene of interest. To this end,  
31 we developed Tiled-Region Exchange (T-REx) mutagenesis, which was greatly inspired by the  
32 EMPIRIC approach. Briefly, the approach is a *multiplexed* version of EMPIRIC where all the  
33 SERFs (referred to as ‘tiles’) are replaced with their corresponding oligonucleotides in a  
34 single reaction. The following four objectives were prioritized prior to the development of this  
35 method: ease of use, generation of all possible single-mutant variants, production of one  
36 and only one variant per clone, and minimization of unmutated sequences in the library.

37 To multiplex EMPIRIC and decrease unmutated sequences in the library, we made two major  
38 modifications to the protocol: 1) we include the toxic *ccdB* gene in the SERF, and 2) we use  
39 tiles that are positioned such that they contain unique and optimized overhangs that can be  
40 used to clone synthetic mutagenizing oligos in a single reaction. The addition of *ccdB* in the

1 SERF virtually guarantees that only cells that have undergone a successful tile-exchange will  
2 be viable (Bernard 1996), thus decreasing the frequency of clones that maintain the WT,  
3 unmutated sequence. The unique optimized overhangs enable the entire gene to be mutated  
4 in a single reaction vessel as these minimize off-target ligations of oligonucleotides. Finally,  
5 to aid in downstream phenotyping of mutants, we introduce an approach to add unique  
6 barcodes to the mutagenesis libraries using the Bxb1 recombinase.

7 A brief overview of the workflow of T-REx is as follows: parallel cloning of SERFs in a gene of  
8 interest on a plasmid, converting designed oligonucleotide pools to dsDNA, one-pot cloning  
9 of all the oligonucleotides with a plasmid pool of SERF-containing genes (Figure 1a),  
10 followed by one-pot fusion with a unique barcode (Figure 1b). The final library can be  
11 introduced in a model system of choice for further phenotyping (Figure 1c).

## 12 **Automated and optimized design of tile locations**

13 One critical component of T-REx is the initial insertion of SERFs. Simultaneous cloning of  
14 oligonucleotides that potentially encode different tile sequences requires that their ligation  
15 and exchange with corresponding SERF is highly specific. Here, a trade-off between the  
16 number of tiles and the length of the cloning oligonucleotide must be considered, as longer  
17 oligonucleotides are more costly and may have a higher rate of synthesis errors (Ultramers  
18 from IDT are expected to yield full length oligos at 50% of the unpurified products at 140 bp  
19 (oPools Oligo Pools | IDT), while having more tiles increases the complexity of the reaction  
20 (notwithstanding the increased number of SERFs that must be cloned in parallel).

21 To explore these constraints, we first verified whether the length of mutagenic  
22 oligonucleotides is a limiting factor in T-REx mutagenesis. We cloned SERFs inside regions  
23 of different lengths into the eforRed pink-producing chromoprotein (Liljeruhm et al. 2018),  
24 and performed a single oligonucleotide exchange, counting resulting colonies for cloning  
25 efficiency and yield. A successful reaction yields pink colonies, while oligonucleotide  
26 synthesis errors (which are frequently indels) would yield a white colony. Across a variety of  
27 overhangs and a large range of oligonucleotide sizes (from <60 bp to 175 bp), using standard  
28 desalted oligos or using Ultramers from IDT (for longer oligos), we found no prohibiting  
29 differences in correct assembly or colony counts after *E. coli* transformation  
30 (Supplementary Figure 1). Others have also found similar rates of full-length synthesis with  
31 Ultramers from IDT (Filges et al. 2021) and as will be described later, even tiles of 120 bp (40  
32 amino acids) that require oligonucleotides of about 180 bp can be mutated successfully with  
33 minimal errors (~5% indel rate, and ~5% wrong base synthesized). Therefore, contrary to our  
34 original assumption, the length of the oligonucleotides or the synthesis platform is not a  
35 major limiting factor for EMPIRIC and the requirement for semi-accurate synthesis on both  
36 5' and 3' ends of the purchased oligo is sufficient to ensure high-efficiency cloning. While  
37 shorter oligos may be more cost-effective and accessible, they do not strongly influence the  
38 reaction efficiency, and thus the trade-off for a comprehensive mutagenesis library is simply  
39 the cost of oligonucleotides and the total number of SERFs. In a practical sense, the span of  
40 a tile can therefore be about 40 amino acids.

1 To reduce cost, we further explored the number of bases needed for cleavage close to the  
2 end of DNA fragments for the BsaI restriction endonuclease. According to NEB, BsaI can  
3 cleave fragments when the recognition sequence is 1 bp from the end of a double-stranded  
4 fragment, but they recommend 6 base pair for Golden Gate assembly (Cleavage Close to the  
5 End of DNA Fragments | NEB). We thus varied the number of bases from 0-20 bp past one of  
6 the two BsaI site in the oligo (one site is necessarily longer due to the need for a primer during  
7 conversion of ssDNA oligos to dsDNA, see Methods) and tested the efficiency of the reaction  
8 using the same colorimetric assay as previously discussed except that we mutagenized a  
9 plasmid containing the amilCP blue-producing chromoprotein. In our hands, all extensions,  
10 including 0, 1, 2, 3, 4, 5, 6, 7, 8, and 17 bp, supported a robust assembly with high colony  
11 counts (Figure 2a and Supplementary Figure 2 for similar experiment using the amilOrange  
12 chromoprotein).

13 Upon testing many different overhangs, we encountered one case where an increase in white  
14 colonies was observed and traced this to a case where overhangs differing by one base could  
15 promote re-circularization of the plasmid without the oligonucleotide. While this is a  
16 common occurrence in some restriction-endonuclease cloning workflows, it was not  
17 expected to occur based on the measured fidelity of T4 DNA ligase (5-GGAA ligated to 3-  
18 CCCT was seen approximately 0.1% of the time in NEB's screen (Potapov et al. 2018)).  
19 Despite this, we observed about 25% white colonies in this reaction (Figure 2b). Thus, care  
20 must be taken when designing tile locations to minimize these off-target ligations.

21 Our previous results suggest that T-REx could be used as a comprehensive mutagenesis  
22 technique that can mutagenize whole genes with minimal constraints. Thus, to aid in  
23 developing this methodology, we developed a script that can automate the cloning process  
24 and choose optimal tile locations, inspiring ourselves from data-optimized assembly design.  
25 The efficiency and fidelity of T4 DNA ligation on all possible 4 bp overhangs was previously  
26 measured by NEB (Potapov et al. 2018), and serves as a guide for optimized tile locations.  
27 The number of tiles is chosen by the user (taking into account the total cost of assembly),  
28 and all oligonucleotides, including mutagenic oligo pools, are returned in a simple output.

29 While it has been routine in synthetic biology labs to remove BsaI sites from genes for cloning  
30 purposes (Marillonnet and Grütznér 2020), it is also necessary for deep-mutagenesis to  
31 consider that sets of random nucleotides can generate a *de novo* BsaI restriction sequence  
32 within the mutagenized region. In our lab, we usually perform degenerate NNK mutagenesis  
33 (though NNN or NNS is also possible) to randomly generate all 20 amino acids at a single  
34 site, while reducing redundancy and the number of stop codons. In certain instances (for  
35 example NNK CTC), a spurious BsaI site can be produced. In these cases, the script  
36 automatically attempts to further mutagenize surrounding bases to preserve the coding  
37 amino acid (for example NNK CTC to NNK CTT) or will resort to NNS mutagenesis. One  
38 advantage of data-optimized assembly design is that it can easily incorporate further  
39 constraints in the design, and so other restriction endonuclease sites can also be  
40 automatically screened and removed in a similar manner.

41 **Tiled-Region Exchange mutagenesis can be performed as a one-pot reaction**

1 One strength of T-REx is that, by choosing unique overhangs where off-target ligations have  
2 been minimized, the reaction might be performed in one-pot. Though not necessarily  
3 required for successful and comprehensive mutagenesis, this may improve throughput and  
4 cost as long as the reaction remains specific. For example, the list price for IDT oligo pools  
5 in small scale is 5 cents per base, while it is 1 cent per base at very high scale (oPools Oligo  
6 Pools | IDT). To showcase the specificity of this one-pot reaction, we cloned SERFs into 5  
7 different chromoproteins. Single-stranded oligos that ‘repair’ these SERFs were ordered and  
8 pooled prior to conversion to dsDNA, and a one-pot Tiled-Region Exchange mutagenesis  
9 reaction was performed. The tile locations were chosen such that incorrect assembly (an  
10 oligo swapped with the wrong SERF) would yield an unpigmented colony. Under ideal  
11 conditions, we would expect approximately equal cloning efficiencies for all five  
12 chromoproteins and thus observe an equal proportion of coloured colonies. The results of  
13 this experiment are shown in Figure 2c-d, suggesting that the reaction is highly efficient and  
14 specific, depicting <1% of colonies having been incorrectly ligated and appearing white, with  
15 each colour appearing at a similar frequency.

16 To further highlight the specificity of this reaction, we performed the same experiment but  
17 this time omitting either one plasmid entirely (Figure 2c) (and thus having too many ‘repair  
18 oligos’) or omitting one swapping oligonucleotide (Figure 2d) (and thus having one plasmid  
19 with a SERF that cannot be successfully cloned). In neither of these cases did we observe a  
20 detrimental effect. As shown in Supplementary Figure 3, all 4 of the other chromoproteins  
21 were properly ligated with their exchange oligos and there was no excess of white colonies.  
22 As expected, no colonies produced the omitted intact chromoprotein.

### 23 **Barcodes can be attached to the mutagenic library using recombinase-mediated fusion**

24 In the EMPIRIC approach, the mutagenized fragment can be sequenced directly for  
25 phenotyping purposes using standard short-read sequencing. However, when mutagenizing  
26 whole-genes in one-pot reactions, this approach cannot be guaranteed to sequence the  
27 intended tile and long-read sequencing typically do not offer the throughput required for  
28 phenotyping thousands of variants effectively. Previously, it has been shown that short,  
29 unique DNA barcodes can be linked to variations of interest and used to phenotype libraries  
30 effectively with short-read sequencing (Chochinov and Nguyen Ba 2022).

31 In many deep-mutational scanning protocols, DNA barcodes are incorporated into the  
32 mutagenized library using PCR, by incorporating random nucleotides in the primers and  
33 cloning of the amplified product (Frank et al. 2022), or by direct ligation (Fowler et al. 2014).  
34 Due to the large number of random bases in these designs, this approach virtually  
35 guarantees that barcodes do not associate with more than one gene fragment, enabling high  
36 fidelity in the following analyses. However, if the PCR product that is amplified with the  
37 barcodes is the mutagenized library, then the final libraries may contain PCR chimeras and  
38 contain more than one variant. To remedy this, mutagenized libraries can be isolated by  
39 restriction enzymes and ligated on a barcoded PCR product of the desired plasmid  
40 backbone.

1 Here, we chose a different strategy where barcode libraries can be sequenced and  
2 characterized beforehand, effectively generating subsamples of the random nucleotide  
3 space. These barcodes can be sequenced at high depth, aiding in future long-read mutation  
4 association sequencing. This library, however, must be fused, or linked, to the mutagenized  
5 genes. To do this, we designed our mutagenesis to be performed on a plasmid containing a  
6 Bxb1 attB(Singh et al. 2013) (bacterial attachment) integration site and ampicillin resistance.  
7 On another plasmid, a barcode library is cloned on an attP-containing (phage attachment)  
8 plasmid(Singh et al. 2013) with an R6K origin of replication with kanamycin  
9 resistance(Metcalf et al. 1994). Bxb1 serine-integrase can thus fuse both plasmids *in vitro*  
10 on the conserved “GA” dinucleotide sequence within the attachment sites, yielding a  
11 barcoded deep-mutation scan library after transformation into standard DH5a cells and  
12 selecting on both kanamycin and ampicillin. The linkage between the mutagenized variant  
13 and the unique barcode can be determined by using any of the long-read sequencing  
14 technologies available and the use of Bxb1 integrase enables flexibility for downstream  
15 users that may wish to use the Gateway(Reece-Hoyes and Walhout 2018) system to clone  
16 their library onto different backbones while preserving this linkage. Thus, this approach  
17 allows the unique tagging, quantification, and identification of each variant in the  
18 sequencing analysis from complex multiplexed mixtures.

19 In our workflow, we have optimized this Bxb1 fusion reaction through several parameters and  
20 assessed these reactions on an agarose gel and by transformation. We first tested a range of  
21 temperatures, 20 °C, 25 °C, 30 °C, and 37 °C, where we noticed that the most complete fusion  
22 reaction occurred at 37 °C (Figure 3a). We also tested various incubation times, at 37 °C, by  
23 periodically plating the reaction every 15-30 minutes and found that the number of colonies  
24 obtained was generally sufficient after two hours, which roughly corresponded to results  
25 from the agarose gel electrophoresis (Supplementary Figure 4). We also tested various salt  
26 concentration and additives (Spermidine, Bovine Serum Albumin (BSA), propylene glycol,  
27 polyethylene glycol (PEG), etc.) to the fusion buffer, where good fusion results occurred in  
28 the presence of at least 75 mM NaCl (Figure 3b), 50 mM KCl and incorporation of 5% PEG  
29 8000 (Figure 3c and Supplementary Figure 5). Finally, to confirm fusion of both plasmids,  
30 restriction enzyme digestion was used to show a successful reaction (Figure 3d).

31 One drawback of our approach is that the barcode diversity is necessarily much lower than  
32 barcodes added by PCR or by ligation. Indeed, this diversity is limited exactly by the number  
33 of barcodes found in the barcoded R6K plasmid pools. However, several lines of evidence  
34 suggest that this disadvantage can be circumvented by simply having a modestly large  
35 number of barcodes. First, a barcoded R6K plasmid pool from our barcoding construction  
36 protocol usually contains about 250,000 barcodes as established from colony counts, far  
37 exceeding the number of variants found in most deep-mutational scan studies. For example,  
38 a gene of 1000 amino acids will have 32 000 possible single-NNK mutants, allowing about  
39 ~8 barcodes per variant as biological replicates. We believe having fewer biological

1 replicates (such as ~8) is preferential to enable high-quality phenotyping as 1000 reads per  
2 barcode may be required to estimate barcode frequencies. Second, even if barcodes get  
3 associated with two different mutations, this should occur relatively infrequently as we will  
4 discuss in the next section. To assess this more generally, we reanalyzed the dataset from  
5 (Nguyen Ba et al. 2019), which used the same barcode library to insert random barcodes into  
6 several yeast strains that contained a fixed known DNA barcode. Analyzing the total set of  
7 barcodes from the libraries yielded 188478 barcodes (from the expected ~250000 from  
8 colony counts). This library was inserted into two starting yeast strains, with the first having  
9 41207 barcodes and the second yeast strain having 28013 barcodes. Between the two yeast  
10 populations, the number of shared barcodes was 5227, which was close to the expectation  
11 of 6119. Thus, the number of barcodes that will be associated with the same mutations is  
12 directly controlled by the barcode library size, which can be built to be sufficiently large for  
13 most deep-mutational scanning studies. Finally, if this overlap is found to be too high, it is  
14 trivial to obtain more barcodes by constructing more barcoded R6K plasmid pools.

### 15 **Deep-mutational scanning libraries produced by tiled-region exchange mutagenesis** 16 **are comprehensive**

17 To confirm that Tiled-Region Exchange mutagenesis can be used for the production of  
18 comprehensive single-variant libraries, we produced a mutant library for the human FKBP1a  
19 gene, whose protein product binds to TOR in the presence of rapamycin to inhibit cell  
20 growth (Sabers et al. 1995). The library was produced in two technical replicates and was  
21 integrated into the yeast genome at the benign *ho* locus, which contains loss-of-function  
22 mutations in the laboratory yeast strain (SGD Project 2011 Jan 18). We then sequenced the  
23 barcoded libraries after yeast integration and sought to obtain several quality metrics that  
24 could confirm successful mutagenesis. The biological insights gained from performing  
25 comprehensive variant scanning on FKBP1a after selection will be described elsewhere.

26 We first verified whether NNK mutagenesis ordered as oPools had a nucleotide bias, which  
27 would result in a skewed amino acid representation. In standard desalted oligos, IDT offers  
28 ‘hand-mixing’ to ensure equal base representation, but this cannot be done at the scale of  
29 oPools. We found a modest synthesis bias with 30%G: 29%T: 24%A: 17%C. With NNK  
30 mutagenesis, this translated to an overrepresentation of glycine amino acids, and a  
31 decrease in histidine and glutamines. While proline was also reduced compared to  
32 expectations, there are more codons that code for proline than histidine, and as such the  
33 absence of prolines in the final mutagenesis library was not very pronounced. To rectify this  
34 bias, we suggest that oligos can be synthesized as both Cricks and Watson strand (so as to  
35 have MNN as well), however this will double the cost of making such mutagenesis libraries  
36 and it may be more cost-effective to transform the libraries at higher multiplicity if possible.

37 Finally, to show that the library can produce variants at every position, we cumulated  
38 sequencing reads with barcodes and their consensus mutations. In total, 14,476 barcode-

1 mutation association pairs were obtained (7104 from replicate 1, and 7372 from replicate 2).  
2 As mentioned previously, only a small number of these barcodes were associated with oligo  
3 synthesis errors despite pooled oligo lengths of 140 bp to 179 bp: 588 contained a frameshift  
4 (4%) and 425 (3%) barcodes were associated with two mutations, either by oligo synthesis  
5 error or through fusing of the same barcode to two variants. To verify whether our barcode  
6 library was diverse enough, we found 269 barcodes present in both libraries that were  
7 associated with different mutations (close to the expectation of 259). Thus, the diversity of  
8 our R6K barcode library of ~250,000 barcodes was relatively high enough to ensure only a  
9 small number of unusable barcodes.

10 We found on average 4.35 barcodes per NNK codon, with the median being 3. However, we  
11 observed a slight mutational bias in the first tile (Figure 4a), presumably due to uneven  
12 plasmid mixing or exchange efficiency, our mutational coverage was 94.66% and a heatmap  
13 indicating positions with the number of barcodes representing the mutation is shown in  
14 Figure 4b. Obtaining about 100 colony-forming units per mutational position appears  
15 sufficient to ensure high coverage of all variants. These results suggest that T-REx  
16 mutagenesis can be used for the production of comprehensive deep-mutational scanning  
17 libraries.

18

## 19 **CONCLUSION**

20 Tiled-Region Exchange (T-REx) mutagenesis combines the simplicity of the EMPIRIC  
21 approach and the throughput of other one-pot mutagenesis techniques. One caveat,  
22 however, is that the initial *ccdB* cloning stage must be done in parallel and can be laborious  
23 for large genes or for more systematic studies. Further, while T-REx can essentially guarantee  
24 one mutation per plasmid within the pool, it is often desirable to introduce multiple  
25 mutations per fragment. With T-REx, multiple variants can be introduced by designing  
26 oligonucleotides containing several mutations, however the user is constrained to a small  
27 region within a tile. In principle, this can be rectified by assembling several oligonucleotides  
28 simultaneously (in a three- or four-piece Golden Gate assembly), however this can only  
29 produce combinatorial libraries. In cases where single variants must be observed under  
30 different gene backgrounds, it may be preferable to combine T-REx with error-prone PCR or  
31 with other mutagenic techniques.

32 During the writing of this manuscript, another study showcased a similar assembly  
33 technique for deep-mutational scanning (Jann et al. 2025). Our approach differs in a few  
34 ways, namely by using *ccdB* during one of the reaction steps, and by allowing pooling of the  
35 complete reaction during assembly. In contrast, one particular strength of that assembly  
36 method is that barcodes are programmed during mutagenesis, which avoids the need for  
37 long-read sequencing and can even bypass sequencing to associate barcodes with  
38 mutations. Nevertheless, as deep-mutational scanning becomes more accessible, there  
39 will be a rise of new methodologies that increase efficiency, cost, and ease of use. New

1 developments in this area will enable combining the strength of different established  
2 methodologies to ultimately give flexibility to researchers in this field.

3 Despite the limitations of Tiled-Region Exchange mutagenesis, our lab has leveraged the  
4 relatively straightforward methodology to train undergraduate students with minimal  
5 molecular biology experience, and we anticipate that its ease of use will be useful for labs  
6 seeking to produce routine comprehensive variant libraries without much trial and error.  
7 Finally, Our FKBP1A library cost about \$600 CAD for the oligos required to produce it, while  
8 a pre-made variant libraries from Twist Bioscience was quoted at \$5400 USD. Thus, the cost  
9 of making libraries in-house is also relatively competitive compared to commercial  
10 products.

11

## 12 **DATA AVAILABILITY**

13 Strains and plasmids are available upon request. Raw sequencing data for FKBP1A deep-  
14 mutational scanning map has been submitted to Sequence Read Archive. BioProject ID:  
15 PRJNA1316630. Sample SAMN51115240 corresponds to paired variant and barcode reads  
16 for replicate 1 (SRA: SRR35259694). Sample SAMN51115242 corresponds to paired variant  
17 and barcode reads for replicate 2 (SRA: SRR35259692). Code is available on  
18 <https://github.com/annb-lab/TRex>.

19

## 20 **ACKNOWLEDGMENTS**

21 We also acknowledge Fritz Roth for input during the development of this method. Finally, we  
22 acknowledge The Centre for Applied Genomics sequencing center for technical help during  
23 the sequencing of the libraries.

## 24 **FUNDING AND COMPETING INTERESTS**

25 KK, CC, KB, and ANNB acknowledge funding from the Canadian Institute of Health Research  
26 (CIHR: Canada Graduate Scholarship-Master's, OGB-185738). ANNB also acknowledges  
27 funding from NSERC (RGPIN-2021-02716), Connaught, and the OVPRI at UTM. The authors  
28 declare that they have no conflict of interest.

## 29 **REFERENCES**

- 30 Ali M, Khan A, Nursimulu TV, Shin JA. 2025. Unlocking genetic potential: harnessing phage for  
31 targeted mutagenesis in phage-assisted evolution. *Nucleic Acids Res.* 53(14):gkaf746.  
32 <https://doi.org/10.1093/nar/gkaf746>
- 33 Bernard P. 1996. Positive selection of recombinant DNA by CcdB. *Biotechniques.* 21(2):320–323.  
34 <https://doi.org/10.2144/96212pf01>
- 35 Casipit CL et al. 1998. Improving the binding affinity of an antibody using molecular modeling and  
36 site-directed mutagenesis. *Protein Science.* 7(8):1671–1680.  
37 <https://doi.org/10.1002/pro.5560070802>

- 1 Chochinov CA, Nguyen Ba AN. 2022. Bulk-Fitness Measurements Using Barcode Sequencing  
2 Analysis in Yeast. *Methods Mol Biol.* 2477:399–415. [https://doi.org/10.1007/978-1-0716-2257-5\\_22](https://doi.org/10.1007/978-1-0716-2257-5_22)
- 3 Cleavage Close to the End of DNA Fragments | NEB. [accessed 2025 Aug 27].  
4 [https://www.neb.com/en-ca/tools-and-resources/usage-guidelines/cleavage-close-to-the-end-of-](https://www.neb.com/en-ca/tools-and-resources/usage-guidelines/cleavage-close-to-the-end-of-dna-fragments?srsId=AfmBOooATgYFi0SUivckaCZzW92rpiT2wte5UIHmkm7I9V5K20wHsRY6)  
5 [dna-fragments?srsId=AfmBOooATgYFi0SUivckaCZzW92rpiT2wte5UIHmkm7I9V5K20wHsRY6](https://www.neb.com/en-ca/tools-and-resources/usage-guidelines/cleavage-close-to-the-end-of-dna-fragments?srsId=AfmBOooATgYFi0SUivckaCZzW92rpiT2wte5UIHmkm7I9V5K20wHsRY6)
- 6 Coyote-Maestas W et al. 2020a. Targeted insertional mutagenesis libraries for deep domain  
7 insertion profiling. *Nucleic Acids Res.* 48(2):e11. <https://doi.org/10.1093/nar/gkz1110>
- 8 Coyote-Maestas W et al. 2020b. Targeted insertional mutagenesis libraries for deep domain  
9 insertion profiling. *Nucleic Acids Res.* 48(2):e11. <https://doi.org/10.1093/nar/gkz1110>
- 10 Engler C, Gruetzner R, Kandzia R, Marillonnet S. 2009. Golden gate shuffling: a one-pot DNA  
11 shuffling method based on type IIs restriction enzymes. *PLoS ONE.* 4(5):e5553.  
12 <https://doi.org/10.1371/journal.pone.0005553>
- 13 Filges S, Mouhanna P, Ståhlberg A. 2021. Digital Quantification of Chemical Oligonucleotide  
14 Synthesis Errors. *Clin Chem.* 67(10):1384–1394. <https://doi.org/10.1093/clinchem/hvab136>
- 15 Firnberg E, Ostermeier M. 2012. PFunkel: efficient, expansive, user-defined mutagenesis. *PLoS*  
16 *One.* 7(12):e52031. <https://doi.org/10.1371/journal.pone.0052031>
- 17 Fowler DM et al. 2023. An Atlas of Variant Effects to understand the genome at nucleotide  
18 resolution. *Genome Biol.* 24(1):147. <https://doi.org/10.1186/s13059-023-02986-x>
- 19 Fowler DM, Fields S. 2014. Deep mutational scanning: a new style of protein science. *Nat Methods.*  
20 11(8):801–807. <https://doi.org/10.1038/nmeth.3027>
- 21 Fowler DM, Stephany JJ, Fields S. 2014. Measuring the activity of protein variants on a large scale  
22 using deep mutational scanning. *Nat Protoc.* 9(9):2267–2284.  
23 <https://doi.org/10.1038/nprot.2014.153>
- 24 Frank F et al. 2022. Deep mutational scanning identifies SARS-CoV-2 Nucleocapsid escape  
25 mutations of currently available rapid antigen tests. *Cell.* 185(19):3603–3616.e13.  
26 <https://doi.org/10.1016/j.cell.2022.08.010>
- 27 Gibson DG et al. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat*  
28 *Methods.* 6(5):343–345. <https://doi.org/10.1038/nmeth.1318>
- 29 Gietz RD. 2014. Yeast transformation by the LiAc/SS carrier DNA/PEG method. *Methods Mol Biol.*  
30 1205:1–12. [https://doi.org/10.1007/978-1-4939-1363-3\\_1](https://doi.org/10.1007/978-1-4939-1363-3_1)
- 31 Hietpas RT, Jensen JD, Bolon DNA. 2011. Experimental illumination of a fitness landscape. *Proc*  
32 *Natl Acad Sci USA.* 108(19):7896–7901. <https://doi.org/10.1073/pnas.1016024108>
- 33 Huttanus HM et al. 2023. Targeted mutagenesis and high-throughput screening of diversified gene  
34 and promoter libraries for isolating gain-of-function mutations. *Front Bioeng Biotechnol.*  
35 11:1202388. <https://doi.org/10.3389/fbioe.2023.1202388>

- 1 Inoue H, Nojima H, Okayama H. 1990. High efficiency transformation of *Escherichia coli* with  
2 plasmids. *Gene*. 96(1):23–28. [https://doi.org/10.1016/0378-1119\(90\)90336-p](https://doi.org/10.1016/0378-1119(90)90336-p)
- 3 Jann J et al. 2025. Making deep mutational scanning accessible: a cost-efficient approach to  
4 construct barcoded libraries for genes of any length. 2025.05.29.656836 [accessed 2025 Sept 5].  
5 <https://www.biorxiv.org/content/10.1101/2025.05.29.656836v1>.  
6 <https://doi.org/10.1101/2025.05.29.656836>
- 7 Kitzman JO et al. 2015. Massively parallel single-amino-acid mutagenesis. *Nat Methods*. 12(3):203–  
8 206, 4 p following 206. <https://doi.org/10.1038/nmeth.3223>
- 9 Kunkel TA. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc*  
10 *Natl Acad Sci U S A*. 82(2):488–492. <https://doi.org/10.1073/pnas.82.2.488>
- 11 Liljeruhm J et al. 2018. Engineering a palette of eukaryotic chromoproteins for bacterial synthetic  
12 biology. *J Biol Eng*. 12:8. <https://doi.org/10.1186/s13036-018-0100-0>
- 13 Marillonnet S, Grütznér R. 2020. Synthetic DNA Assembly Using Golden Gate Cloning and the  
14 Hierarchical Modular Cloning Pipeline. *Curr Protoc Mol Biol*. 130(1):e115.  
15 <https://doi.org/10.1002/cpmb.115>
- 16 Metcalf WW, Jiang W, Wanner BL. 1994. Use of the rep technique for allele replacement to  
17 construct new *Escherichia coli* hosts for maintenance of R6K gamma origin plasmids at different  
18 copy numbers. *Gene*. 138(1–2):1–7. [https://doi.org/10.1016/0378-1119\(94\)90776-5](https://doi.org/10.1016/0378-1119(94)90776-5)
- 19 Nguyen Ba AN et al. 2019. High-resolution lineage tracking reveals travelling wave of adaptation in  
20 laboratory yeast. *Nature*. 575(7783):494–499. <https://doi.org/10.1038/s41586-019-1749-3>
- 21 Odland T. 2025. tommyod/Efficient-Apriori. [accessed 2025 Nov 27].  
22 <https://github.com/tommyod/Efficient-Apriori>
- 23 oPools Oligo Pools | IDT. Integrated DNA Technologies; [accessed 2025 Sept 5].  
24 [https://www.idtdna.com/pages/products/custom-dna-rna/dna-oligos/custom-dna-oligos/opools-](https://www.idtdna.com/pages/products/custom-dna-rna/dna-oligos/custom-dna-oligos/opools-oligo-pools)  
25 [oligo-pools](https://www.idtdna.com/pages/products/custom-dna-rna/dna-oligos/custom-dna-oligos/opools-oligo-pools)
- 26 Ossa-Hernández N, Marins LF, Almeida DV. 2024. Combination of error-prone PCR (epPCR) and  
27 Circular Polymerase Extension Cloning (CPEC) for improving the coverage of random mutagenesis  
28 libraries. *Sci Rep*. 14(1):15874. <https://doi.org/10.1038/s41598-024-66584-y>
- 29 Pattanaik S, Werkman JR, Kong Q, Yuan L. 2010. Site-Directed Mutagenesis and Saturation  
30 Mutagenesis for the Functional Study of Transcription Factors Involved in Plant Secondary  
31 Metabolite Biosynthesis. In: Fett-Neto AG, editor. *Plant Secondary Metabolism Engineering:*  
32 *Methods and Applications*. Humana Press; p 47–57 [accessed 2025 Aug 28].  
33 [https://doi.org/10.1007/978-1-60761-723-5\\_4](https://doi.org/10.1007/978-1-60761-723-5_4). [https://doi.org/10.1007/978-1-60761-723-5\\_4](https://doi.org/10.1007/978-1-60761-723-5_4)
- 34 Plapp BV. 1995. [4] Site-directed mutagenesis: A tool for studying enzyme catalysis. In: *Methods in*  
35 *Enzymology*. Vol. 249. Academic Press; p 91–119 (Enzyme Kinetics and Mechanism Part D:  
36 *Developments in Enzyme Dynamics*). [accessed 2025 Aug 28].

- 1 <https://www.sciencedirect.com/science/article/pii/S0076687995490329>.  
2 [https://doi.org/10.1016/0076-6879\(95\)49032-9](https://doi.org/10.1016/0076-6879(95)49032-9)
- 3 Potapov V et al. 2018. Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA  
4 Ligase and Application to DNA Assembly. *ACS Synth Biol.* 7(11):2665–2674.  
5 <https://doi.org/10.1021/acssynbio.8b00333>
- 6 Pryor JM et al. 2020. Enabling one-pot Golden Gate assemblies of unprecedented complexity using  
7 data-optimized assembly design. *PLoS One.* 15(9):e0238592.  
8 <https://doi.org/10.1371/journal.pone.0238592>
- 9 Reece-Hoyes JS, Walhout AJM. 2018. Gateway Recombinational Cloning. *Cold Spring Harb Protoc.*  
10 2018(1). <https://doi.org/10.1101/pdb.top094912>
- 11 Rubin AF et al. 2017. A statistical framework for analyzing deep mutational scanning data. *Genome*  
12 *Biology.* 18(1):150. <https://doi.org/10.1186/s13059-017-1272-5>
- 13 Sabers CJ et al. 1995. Isolation of a protein target of the FKBP12-rapamycin complex in mammalian  
14 cells. *J Biol Chem.* 270(2):815–822. <https://doi.org/10.1074/jbc.270.2.815>
- 15 SGD Project. 2011. [accessed 2011 Jan 17]. <http://www.yeastgenome.org/>
- 16 Singh S, Ghosh P, Hatfull GF. 2013. Attachment site selection and identity in Bxb1 serine integrase-  
17 mediated site-specific recombination. *PLoS Genet.* 9(5):e1003490.  
18 <https://doi.org/10.1371/journal.pgen.1003490>
- 19 Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein*  
20 *Expr Purif.* 41(1):207–234
- 21 Watanabe S, Ito M, Kigawa T. 2021. DiRect: Site-directed mutagenesis method for protein  
22 engineering by rational design. *Biochemical and Biophysical Research Communications.* 551:107–  
23 113. <https://doi.org/10.1016/j.bbrc.2021.03.021>
- 24 Wei H, Li X. 2023. Deep mutational scanning: A versatile tool in systematically mapping genotypes  
25 to phenotypes. *Front Genet.* 14:1087267. <https://doi.org/10.3389/fgene.2023.1087267>
- 26 Weile J et al. 2017. A framework for exhaustively mapping functional missense variants. *Mol Syst*  
27 *Biol.* 13(12):957. <https://doi.org/10.15252/msb.20177908>
- 28 Weiss GA et al. 2000. Rapid mapping of protein functional epitopes by combinatorial alanine  
29 scanning. *Proc Natl Acad Sci U S A.* 97(16):8950–8954. <https://doi.org/10.1073/pnas.160252097>

30

31 **FIGURE CAPTIONS**

32 **Figure 1: (a)** The gene of interest (*goi*) is split into non-overlapping tiled regions with unique  
33 inward-facing *Bsa*I sites. The native sequence is temporarily replaced with a *ccdB* negative  
34 selection cassette, and then swapped for the desired dsDNA mutagenic oligo containing

1 NNK codons at each position in a Golden Gate reaction with 4-base overhangs. **(b)** The  
 2 variant library and barcode plasmid library are fused in a Bxb1 recombinase reaction and  
 3 can be sequenced to associate the mutation to their unique barcode. **(c)** To complete a  
 4 selection assay on the variant library, variants can compete for growth and their relative  
 5 fitness can be obtained by sequencing the abundance of each barcode before and after  
 6 selection.

7 **Figure 2: (a)** Barplot depicting colony formation vs the number of overhang bases past the  
 8 BsaI recognition sites, with three replicates per base length. **(b)** Stacked barplots indicating  
 9 the overhang specificity and its impact on ligation accuracy. **(c-d)** Stacked barplots  
 10 indicating colony counts grouped by chromoprotein identity (five chromoproteins each  
 11 represented by a different colour). **(c)** depicts the difference between all colours included in  
 12 the pool, versus omitting each colour plasmid individually from the pool, while **(d)** shows the  
 13 colony counts of the pool when omitting each oPool oligo individually from the pool of oligos.

14 **Figure 3: Optimization of Bxb1 integrase-mediated fusion between a linear R6K (ori<sub>y</sub>)**  
 15 **recipient vector and a donor fragment. A complete fusion reaction yields four diagnostic**  
 16 **bands corresponding to the linear R6K backbone, the donor fragment, and the two**  
 17 **recombination products and is consistent across each reaction. (a)** Agarose gel of  
 18 **temperature series. Lane 1: DNA size ladder. Lanes 2-5: fusion reactions performed at 20 °C,**  
 19 **25 °C, 30 °C, and 37 °C, respectively for 1 hour. (b)** Additive series. Lane 1: DNA ladder. Lanes  
 20 **2-5: reactions supplemented with 2 mM spermidine, 200 µg/mL BSA, 5% PEG, and 10%**  
 21 **propylene glycol, respectively. Reaction at 37 °C for 1 hour. (c)** Salt tolerance series. Lane 1:  
 22 **DNA ladder. Lanes 2-5: reactions supplemented with 25 mM, 50 mM, 75 mM, and 100 mM**  
 23 **NaCl, respectively. Reaction at 37 °C for 1 hour. (d)** Validation of plasmid fusions using PmeI  
 24 **restriction endonuclease which confirms correct linearization of fragments. Lane 1: 10kb**  
 25 **DNA ladder; Lane 2: PmeI digested barcode plasmid before Bxb1 fusion reaction; Lane 3:**  
 26 **PmeI digested mutagenized plasmid before Bxb1 fusion reaction; Lane 4: PmeI digested**  
 27 **fusion construct of barcode plasmid and mutagenized plasmid after Bxb1 fusion reaction.**  
 28 **Band sizes are consistent with PmeI restriction site locations.**

29  
 30 **Figure 4: (a)** Bar plot showing the number of barcodes per position along FKBP1A. The gene  
 31 was divided into three tiles: Tile 1 (positions 2-29), Tile 2 (positions 30-70), and Tile 3  
 32 (positions 71-108) and a total of 13,463 barcodes-mutation pair were detected, distributed  
 33 as follows: 6,279 in Tile 1, 3,743 in Tile 2, and 3,950 in Tile 3. Barcodes corresponding to  
 34 variants without a single identifiable mutation or an indel were excluded from this analysis  
 35 (1013 barcodes removed). **(b)** Heatmap of barcode frequency per mutation. Wild-type  
 36 positions are outlined in black boxes.

37

38

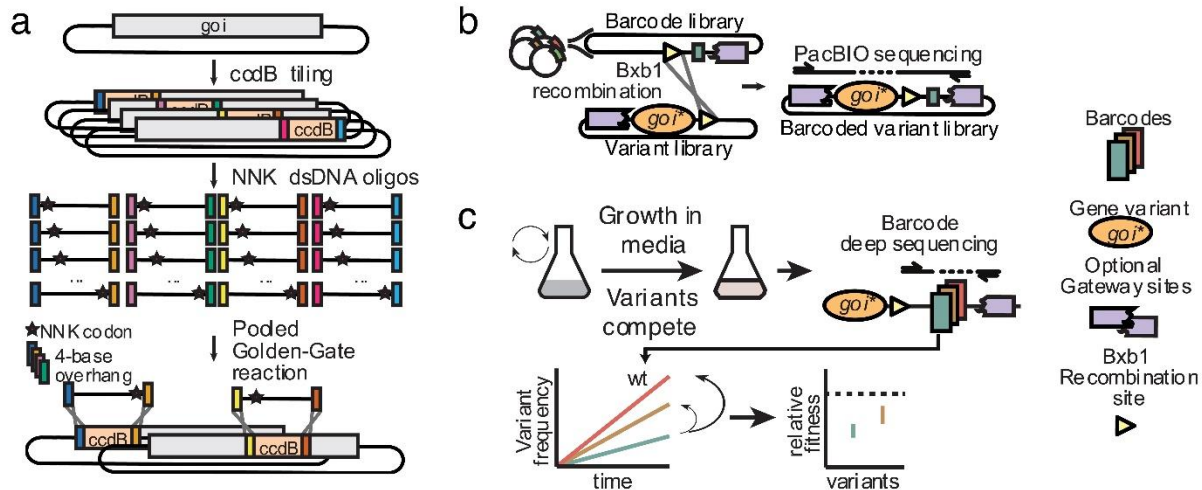


Figure 1  
165x68 mm (x DPI)

1  
2  
3  
4

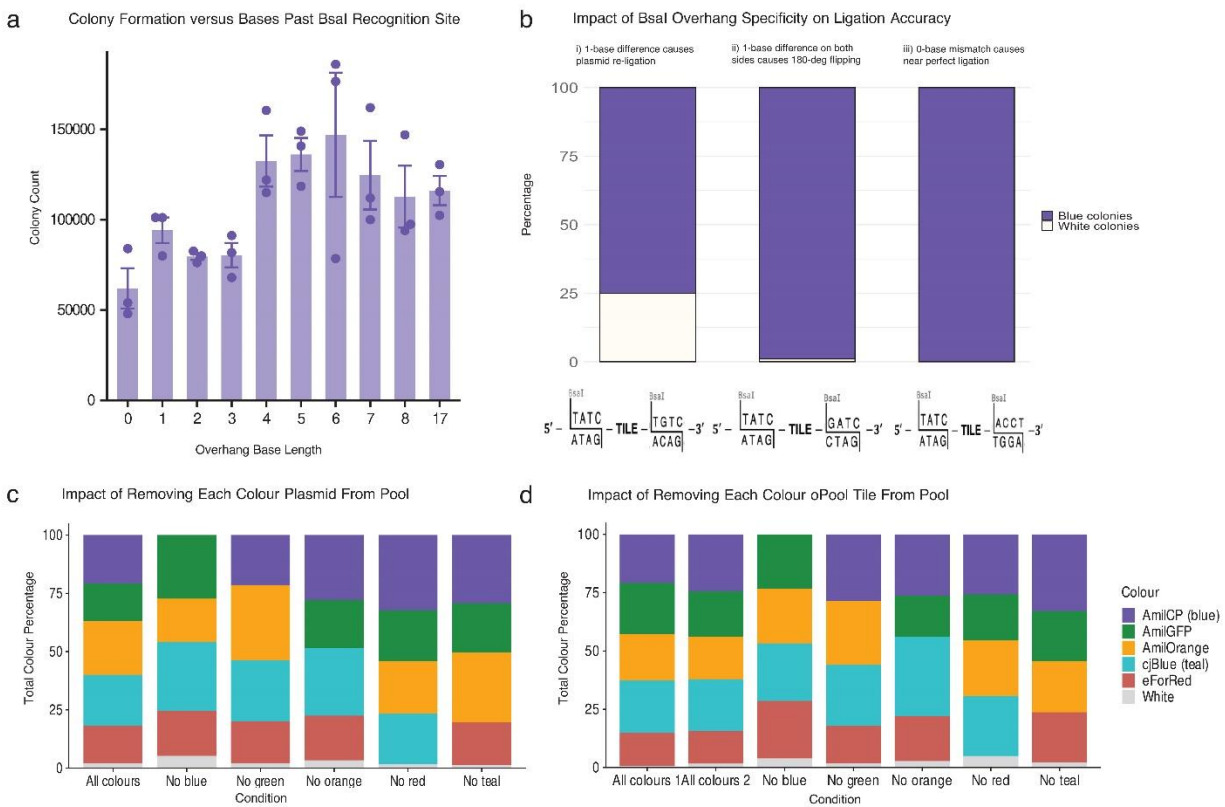


Figure 2  
 165x111 mm (x DPI)

1  
 2  
 3  
 4

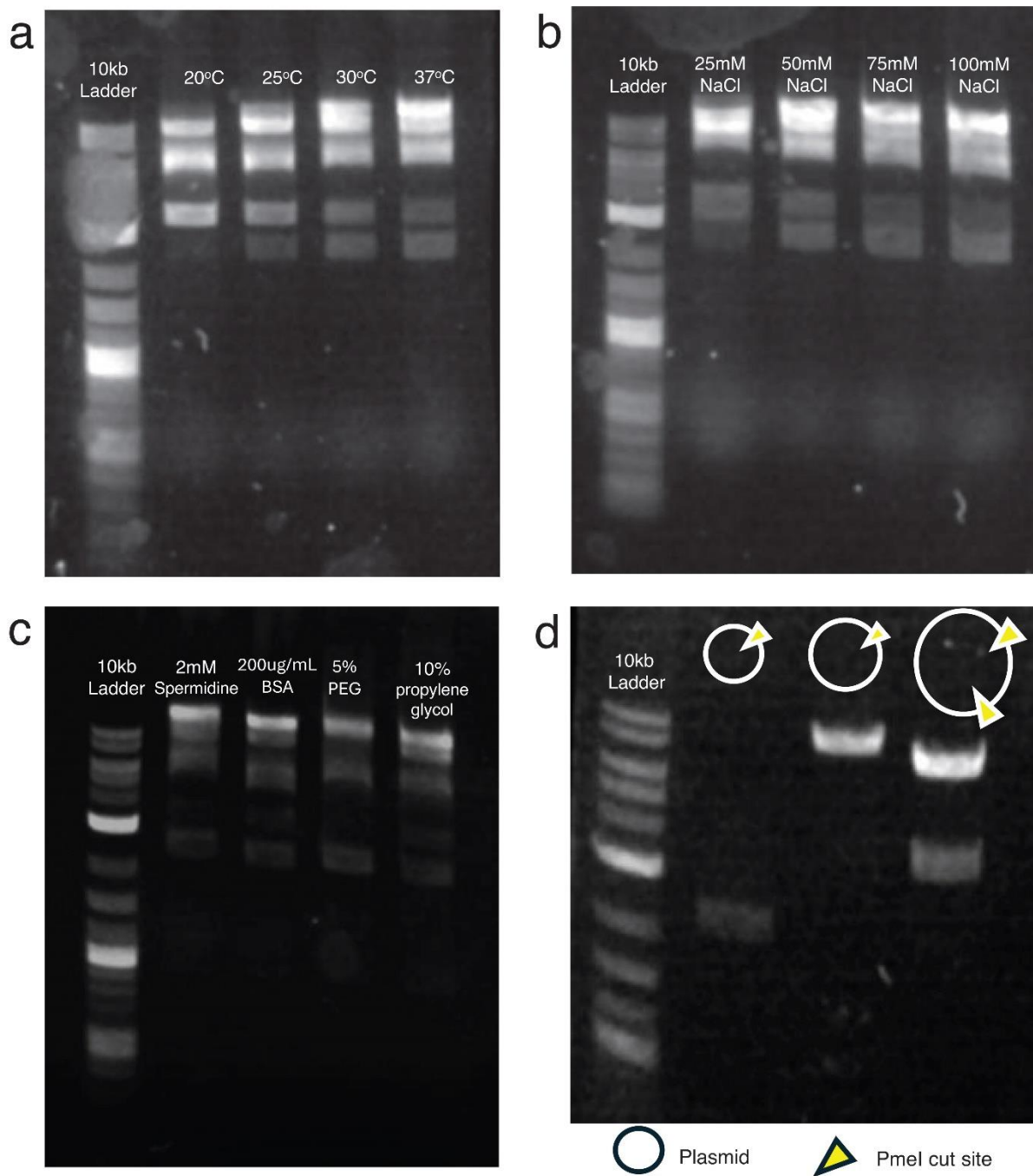
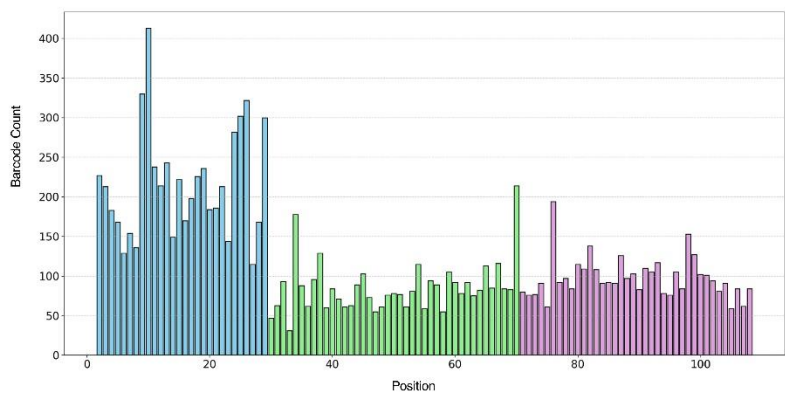


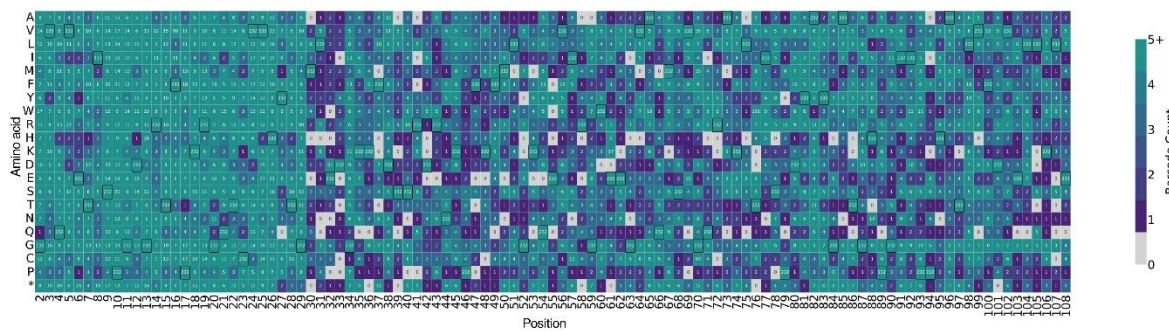
Figure 3  
165x185 mm (x DPI)

1  
2  
3  
4

a

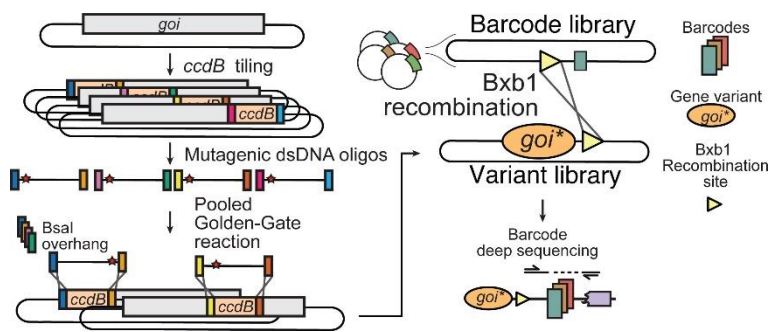


b



- 1
- 2
- 3
- 4

Figure 4  
165x106 mm (x DPI)



Graphical Abstract

1  
2