P3a site-specific and cassette mutagenesis for seamless protein and plasmid engineering

Xiang-Jiao Yang[1,2,3,4,*]

[1]Rosalind and Morris Goodman Cancer Institute, McGill University, Montreal, Quebec H3A 1A3, Canada

[2]Department of Medicine, McGill University, Montreal, Quebec H3A 1A3, Canada

[3]Department of Biochemistry, McGill University, Montreal, Quebec H3A 1A3, Canada

[4]McGill University Health Center, Montreal, Quebec H3A 1A3, Canada

*Correspondence: xiang-jiao.yang@mcgill.ca

Keywords: Germline mutation, somatic mutation, neurodevelopmental disorder, histone H3, BRPF1, KAT6A, KAT7, KAT8, p300, CBP, synthetic biology, AI-assisted protein design

Abstract

Site-directed mutagenesis is a basic molecular tool required for protein and plasmid engineering. We have recently adopted an innovative strategy using primer pairs with 3′-overhangs and developed an optimized protocol, referred to as P3 site-directed mutagenesis, with an efficiency of ~50%. This raises the important question how to enhance the efficiency to the ideal level of 100%. Here we report P3a site-directed mutagenesis with an efficiency at or close to 100% and have systematically evaluated this improved method by engineering >100 point mutations and small deletions (or insertions) on >20 mammalian expression vectors encoding various epigenetic regulators and the spike protein of SARS-CoV-2. Moreover, all known mutagenesis methods are limited to point mutations and small deletions/insertions (typically up to a dozen nucleotides), raising the technical problem how to carry out cassette mutagenesis for replacement, deletion or insertion of large DNA fragments. Accordingly, the high efficiency of P3a mutagenesis and the 'handshaking' feature of primer pairs with 3′-overhangs have inspired us to adapt this method for efficient and seamless cassette mutagenesis, including high efficiency in epitope tagging and untagging, deletion of small and large DNA fragments and insertion of oligonucleotide duplexes (such as LoxP sites and sgRNA-coding sequences). Thus, this new site-specific and cassette mutagenesis method is highly efficient, likely resulting in its wide use for biomedical research.

Introduction

Site-directed mutagenesis is a basic method required for research in different branches of biomedical sciences [1]. Whole-genome and exome sequencing have identified numerous germline or somatic mutations in various diseases, as illustrated by genetic disease-linked mutations listed in the recently established ClinVar database [2] and cancer-associated mutations documented in the cBioPortal and COSMIC databases [3,4]. Because many of these mutations are missense and their functional impact is difficult to predict, an important question is to test whether these mutations are causal. For this, it is necessary to utilize site-directed mutagenesis and engineer the mutations for functional analysis *in vitro,* followed by generation and analysis of model organisms with equivalent mutations *in vivo*. Site-directed mutagenesis is also required for modeling evolution of viruses, such as SARS-CoV-2, to assess the impact of their newly acquired mutations [5,6]. This technique is also essential for protein engineering, such as testing mutant protein candidates from machine learning-based design [7] and generating artificial SH2 domains like phospho-tyrosine super-binders for deep proteomics and clinical profiling [8,9]. Other applications of protein engineering include generation of super-antibodies and artificial intelligence (AI)-assisted *de novo* design of novel proteins for biomedical research, clinical diagnosis and therapy [10-12].

Furthermore, instead of targeting one or a few residues, 'cassette' mutagenesis replaces, deletes or inserts DNA fragments, so it is valuable for introducing small or large deletions, engineering epitope tags, optimizing plasmids, inserting different types of DNA sequences (such as LoxP sites, related DNA-recombination elements, promoters and the coding sequences for nuclear import and export signals) and constructing sgRNA and shRNA expression vectors. So far, the only means to carry out cassette mutagenesis is through subcloning, which is restricted to

the availability of suitable restriction sites. Seamless cassette mutagenesis is an ideal choice and should have much wider applications.

Site-directed mutagenesis was initially developed for engineering mutations at specific sites of single-stranded phage or phagmid DNA [1,13]. By contrast, the QuickChange™ mutagenesis method utilizes double-stranded plasmids as the templates [14-16]. It is based on Pfu DNA polymerase-mediated PCR with a pair of completely complementary primers containing a given mutation (Fig. 1A), followed by DpnI digestion to destroy parental plasmids methylated at GATC sites. While it has been widely used in different laboratories, this method suffers from at least five limitations: 1) the complementary primers self-anneal and generate primer-primer dimers during PCR, thereby decreasing the amplification efficiency; 2) newly synthesized plasmid DNA is 'nicked' and thus unsuitable as a template for further amplification; 3) unwanted mutations at the primer sites resulting from impurity of primers; 4) unwanted mutations at the mutation and primer sites due to the processivity or strand-displacing activity of Pfu [17,18]; and 5) this polymerase synthesizes at a rather slow rate of ~1 kb/min, so it takes 4-5 hours to complete the initial PCR reaction.

Mitigating limitations 1) and 2), an alternative approach utilizes a pair of partially complementary primers with two 3′-protruding ends (Fig. 1B), which allows the use of newly synthesized strands as templates for subsequent PCR amplification (Fig. S1A) [19,20]. This improved method is referred to as P3 (primer pair with 3′-protruding ends) site-directed mutagenesis [17,18]. About limitation 3), we have shortened the length of primers to ~30 nucleotides (Fig. 1B) as shorter oligonucleotides tend to be synthesized easily at high quality. Related to limitations 4)-5), we utilized PfuUltra (an improved version of Pfu) [21]. We have evaluated this resulting method (Fig. 1B-C) with a dozen of mammalian expression vectors

ranging from 7.0-13.4 kb [17,18]. Compared to the QuickChange™ method, the success rate has increased significantly, reaching an average efficiency of close to or over 50%. To improve this method further, we sought to overcome limitations 4)-5) by replacing Pfu DNA polymerase with other high-fidelity thermostable DNA polymerases with better quality. As a result, we tested 4 candidates, including: 1) Q5 high-fidelity DNA polymerase, composed of a thermostable DNA polymerase fused to the processivity-enhancing Sso7d DNA-binding domain, thereby improving speed, fidelity and reliability of PCR amplification [22]; and 2) Platinum SuperFi II polymerase, whose fidelity was claimed to be even better than Q5 DNA polymerase.

Results from >100 mutations that we have engineered indicate that these two polymerases are superior to Pfu and its derivatives by elevating the mutagenesis efficiency very close to 100%, speeding up PCR amplification by 2-3 folds and minimizing unwanted mutations at or near the primer sites. Due to the much higher success rate, only 1-2 bacterial colonies need to be sequenced per mutation reaction, thereby reducing labor and reagent costs. To our knowledge, this is the first site-specific mutagenesis method to achieve the efficiency so consistently at or close to the ideal level of 100%. Importantly, in light of the unique 'handshaking' feature of the primers with 3′-protruding ends (Fig. 1B), we have led us to develop a highly efficient cassette mutagenesis method (Fig. 1D-E) for seamless epitope tagging or untagging, deletion of small or large fragments, optimization of plasmids and insertion of any oligonucleotide duplexes with 3-overhangs. The reagent cost per mutagenesis reaction is only about $0.25 and $0.5 with Q5 and SuperFi II master mixes, respectively. Thus, compared to those based on Pfu and its derivatives [17,19,20], this newly developed mutagenesis method is not only much faster but also more efficient, versatile and economical, likely making it a standard tool in various biomedical laboratories.

Results

Improving P3 site-directed mutagenesis

We have recently evaluated the P3 site-directed mutagenesis method that relies on PfuUltra (Fig. 1B-D) [17,18]. The method reached an average efficiency of nearly 50% but encountered problems with some vectors [17]. To improve the method further, we considered to replace PfuUltra with other high-fidelity thermostable DNA polymerases comparable with or superior to PfuUltra in terms of PCR performance. We tested 4 candidates: 1) Q5 high-fidelity DNA polymerase, 2) Platinum SuperFi II high-fidelity DNA polymerase, 3) i7® high-fidelity DNA polymerase and 4) KOD high-fidelity DNA polymerase. In terms of mis-incorporation rates, the former three were estimated to be ~15 times lower than PfuUltra, whereas KOD high-fidelity DNA polymerase is comparable to Pfu. Moreover, commercially available master mixes containing Q5 high-fidelity DNA polymerase, Platinum SuperFi II high-fidelity Polymerase or i7® high-fidelity DNA polymerase contain hot-start and other ingredients to enhance PCR performance, especially for templates with high GC-content and complex DNA sequences. As described below, Q5 and Platinum SuperFi II high-fidelity DNA polymerases outperformed PfuUltra, increasing the mutagenesis efficiency to ~100%, shortening the PCR amplification time by 2-3 folds and reducing unwanted mutations at or near the primer sites. Compared to Q5, Platinum SuperFi II DNA polymerase yielded slightly more colonies in some cases. KOD DNA polymerase was comparable to PfuUltra, whereas i7® high-fidelity DNA polymerase yielded very few or no colonies at all. Compared to PfuUltra [17], Q5 and Platinum SuperFi II high-fidelity DNA polymerases yielded 5-10 more colonies per mutagenesis reaction, which enhances the success rate substantially. Thus, we chose Q5 and Platinum SuperFi II high-fidelity DNA polymerases for the rest of this study. To distinguish it from Pfu-based P3 mutagenesis, this

newly developed method is referred to as P3a site-directed mutagenesis, with the letter "a" denoting an updated version [17,18].

P3a site-specific mutagenesis for efficiently engineering missense mutants of histone H3

Histone H3 is almost invariant from yeasts to humans and plays an important role in organizing the genomes and their access for various DNA-based nuclear processes, such as DNA replication, repair and RNA transcription [23]. At its N-terminal tail, there are multiple lysine residues (i.e., K4, K9, K14, K18, K27 and K36) for acetylation, methylation and other modifications, thereby bringing about various regulatory mechanisms [23]. Moreover, substitutions of K27, G34 and K36 have been linked to oncogenesis, which has led to the onco-histone hypothesis [24]. To test P3a mutagenesis, we sought to engineer different histone H3 mutants as they are important for molecularly investigating the role of these residues by analyzing the impact of their substitutions. As shown in Fig. 1F-G, for the K9R, K14R, K18R and K23R mutants, we utilized 4 different primer pairs, such as K9R-F and K9R-R for engineering the mutation K9R (Fig. S1B). We sequenced plasmids from three of the resulting colonies per mutation and found that 11 of them were mutants, reaching an overall efficiency of 11/12 (91.6%, Fig. 1F). This is impressive as we have not seen any known site-directed mutagenesis methods with such an almost ideal success rate. One drawback of P3 or P3a site-directed mutagenesis is the need to use a pair of primers for engineering one mutation, which is different from classical methods that utilize only one primer per mutation. Thus, we next utilized the primer design strategies described in Fig. 1C to generate two mutants per primer pairs. For example, we employed one primer pair (K27R-F and K27M-R) with a mismatch for generating two mutations (K27R and K27M; Fig. S1B). Similarly, we designed the two primers G34R-F

and G34A-R with two mismatched bps for generating the two mutations G34R and G34A, respectively (Fig. S1C). For this, we analyzed 6 colonies per primer pair and found 2 to 5 colonies containing plasmids with the expected mutations (Fig. 1F), thereby reaching an overall efficiency of 12/18 (66.7%). Typically, for P3a mutagenesis, it is often adequate to sequence 2 colonies per mutation, thereby reducing labor and reagent costs. Thus, this newly developed method is efficient and economical.

P3a mutagenesis for engineering missense mutants of two epigenetic regulators

To assess the versatility of the method, it is necessary to test it with a range of vectors, including those much larger than the one described above for histone H3 (only ~5.7 kb). In addition to histones, there are hundreds of epigenetic regulators that modify or bind to histones for regulating genome organization during different nuclear processes. To study molecular functions of these epigenetic regulators, it is necessary to engineer various mutants. We thus sought to apply P3a mutagenesis for rapidly generating such mutants. Another practical reason is that the expression vector for histone H3 is ~5.7 kb, which is small compared to many expression vectors for epigenetic regulators. Typically, it becomes more difficult for PCR-based mutagenesis as the vector becomes larger. One epigenetic regulator is BRPF1 (bromodomain- and PHD finger-containing protein 1) [25]. It was originally identified as a 140 kDa protein with a bromodomain [25], so its expression vector reaches ~9.2 kb and is significantly larger than that for histone H3.

In addition to the bromodomain, BRPF1 possesses multiple other domains for epigenetic regulation, including the N-terminal part for interacting with lysine acetyltransferase 6A (KAT6A) and its paralog KAT6B [26]. The interaction is required for activation of these acetyltransferases to acetylate and propionylate nucleosomal histone H3 [27,28]. In addition,

BRPF1 possesses two EPC (Enhancer of Polycomb)-like motifs, the second of which interacts with ING4 (or its paralog ING5) and MEAF6 [29]. Moreover, BRPF1 has two PHD fingers flanking a zinc knuckle, with the first PHD finger for recognition of the unmodified N-terminus of nucleosomal histone H3 [30]. Recent studies have linked *BRPF1* mutations to a neurodevelopmental disorder [27,31-34]. According to the ClinVar database [2], additional *BRPF1* mutations have been identified in patients. While some of the identified mutations have been molecularly characterized [27,31,32], many more remain to be analyzed for establishing their pathogenicity firmly. For molecular characterization of different BRPF1 domains, we also needed to engineer more mutants (Fig. 2A). We sought to employ P3a mutagenesis to engineer new BRPF1 mutants. As shown in Fig. 2A, among 31 bacterial colonies analyzed, 27 contained the mutant plasmids, so the efficiency was 87.1%.

At the sequence and domain-organization levels, BRPF1 is highly homologous to BRPF2, but KAT7 is a preferred partner of BRPF2 for acetylating nucleosomal histone H3 [29,35-37]. By contrast, BRPF1 activates KAT6A and KAT6B [27,31]. Thus, BRPF1 is functionally different from BRPF2. Crystal structural analysis has identified a hydrophobic interface for interaction between BRPF2 and KAT7 [38]. This involves residues I53, L57, I59, I60 and L61 of BRPF2 [38]. No functional validation has been reported on the functional importance of this physical interaction. One approach is to engineer BRPF2 mutations at this interface. We thus utilized P3a mutagenesis to generate such mutants, including I53A, L57A, I59A, I60A and L61A (Fig. 2B). Moreover, we engineered G38R, H275P and P275A. G38R is to investigate the importance of a polymorphic substitution, whereas H275P is to correct an unexpected substitution identified in an expression plasmid that we have. P275A was engineered on the repaired plasmid to investigate the functional importance of P275. Amazingly, 23 out of

25 colonies that we analyzed were the correct mutants, reaching the efficiency of 92% (Fig. 2B). Representative sequence analysis results are shown in Fig. 2C. In another experiment, we engineered 11 point mutants of BRPF2 using 10 pairs of primers. We sequenced plasmids from 2-3 colonies per mutagenesis reaction. 22 out of 23 of the resulting colonies were correct, so the mutagenesis efficiency was 95.7%. The BRPF2 expression vector is ~9.2 kb and similar to that for BRPF1 in size, so these results indicate that P3a mutagenesis is highly efficient for vectors up to 9.2 kb. If an empty vector itself is 5-6 kb, the insert can encode a protein up to 150 kDa. This is important because a majority of proteins encoded by various genomes range from 30-100 kDa, so the method is expected to be applicable for many proteins.

P3a mutagenesis for engineering mutants of small and large histone acetyltransferases

To evaluate the versatility of the method, we tested it with additional expression vectors. For this, we targeted histone acetyltransferases as they are our main research subjects [39]. The human genome encodes over a dozen of such enzymes [40]. As mentioned above, one such enzyme, KAT7, interacts with (and is activated by) BRPF2. Crystal structural analysis has identified a surface for interaction with a hydrophobic core of BRPF2 [38]. This involves five residues (K537, I539, I541, V574 and K576) for interacting with I53, L57, I59, I60 and L61 within the hydrophobic core of BRPF2, [38]. With P3a mutagenesis, we carried out alanine substitutions of these five residues of KAT7 [38]. Moreover, we sought to substitute two zinc-chelating residues within an uncharacterized zinc finger of KAT7 and replace its K432, a potential autoacetylation site that has never been characterized [41]. As shown in Fig. 2D, 23 of 24 colonies that we analyzed were mutants, thus reaching a high efficiency of 95.8%, further attesting to effectiveness of the method.

To assess the general applicability of the method further, we next tested it with six different mammalian expression vectors encoding five different lysine acetyltransferases, including KAT6A, KAT6B, KAT8, p300 and CBP. The expression vectors range from 6.6-13.4 kb in size: human KAT6A (full-length), 11.9 kb; its HAT domain, 6.6 kb; the HAT domain of KAT6B, 6.8 kb; KAT8, 7.0 kb; p300, 12.8 kb and mouse CBP, 13.4 kb. Among them, the expression vectors for full-length KAT6A, p300 and CBP are quite large, ranging from 11.9-13,4 kb, so they are good candidates for testing the versatility of P3a mutagenesis with large vectors. This is important as PCR-based mutagenesis methods are more efficient with small plasmids and tend to encounter problems with large plasmids, especially those close to or over 12 kb.

KAT6A and KAT6B interact with BRPF1 for acylating histone H3 [27,28]. For full-length KAT6A, we engineered two missense mutants and three nonsense mutants (Fig. S2A). In addition, we generated five missense mutants of its HAT domain (Fig. S2A). While K604R possesses an arginine substitution of K604, the sole autoacetylation site required for the acyltransferase activity of KAT6A [27], the other mutants of KAT6A and its HAT domain are clinical variants identified in patients with a neurodevelopmental disorder [2]. For example, L1061* is derived from a patient with intellectual disability [27]. These clinical variants have not been adequately analyzed for their pathogenicity, so it is necessary to engineer them for functional analysis *in vitro*. We thus applied P3a mutagenesis to engineering these mutants. For the five mutations on the backbone for full-length KAT6A, we analyzed plasmids from 23 bacterial colonies by Sanger sequencing and identified 19 as correct mutants, reaching efficiency of 82.6% (Fig. S2A). This result is significant as the vector is quite large, 11.9 kb. For the five mutations engineered on the vector for the HAT domain of KAT6A, 9 out of 12 colonies

analyzed possessed the mutant plasmids (Fig. S2A), yielding an efficiency of 75%. The overall mutagenesis efficiency for these two KAT6A expression vectors was 80.0% (Fig. S2A).

KAT6B is paralogous to KAT6A. For KAT6B, we engineered 7 missense mutants on an expression vector for the HAT domain (Fig. S2B). These missense mutants are based on clinical variants identified in patients with a neurodevelopmental disorder [2]. These clinical variants have not been analyzed for their pathogenicity, so it is important to engineer them for functional analysis *in vitro*. Moreover, to define the boundary of the HAT domain, it is necessary to generate two deletion mutants (Fig. S2B). We utilized P3a mutagenesis to engineer these 9 missense or deletion mutants. As shown in Fig. S2B, plasmids from 26 out of 27 colonies analyzed by Sanger sequencing were the mutants. Thus, the efficiency was 26/27 (96.3%), very close to the ideal level of 100%.

We next tested KAT8, which shares its catalytic domain with KAT6A, KAT6B and KAT7 [39]. Importantly, *KAT8* mutations have recently been linked to a new neurodevelopmental disorder [42]. Moreover, additional missense mutants of KAT8 have been identified as clinical variants in patients with this disorder [2]. These clinical variants have not been analyzed for their pathogenicity, so it is important to engineer them for functional analysis *in vitro*. We thus applied P3a mutagenesis to engineer 13 missense mutants on a mammalian expression vector encoding FLAG-tagged KAT8. The size of this vector is 6.9 kb. As shown in Fig. S2C, plasmids from 24 out of 31 colonies analyzed by Sanger sequencing were the mutants. Thus, the mutagenesis efficiency was 24/31 (77.4%).

In the superfamily of lysine acetyltransferases encoded by the human genome, p300 and CBP are the largest, each possessing ~2,400 residues. Their mammalian expression vectors reach 12.8 and 13.4 kb in size. They are among the largest plasmids that have been constructed. Thus,

they are good candidates for testing the versatility of the P3a mutagenesis method. For p300, we sought to engineer 11 missense mutants and 3 C-terminal truncation mutants (Fig. 2E). Some of these mutants are derived from somatic mutations identified in cancer and deposited in the COSMIC and cBioPortal databases [3,4]. For example, D1399N, D1399Y, Y1414C, W1466C and Y1467N are five hotspot mutations [3,4]. The others (such as C1204S/R, D1384N/Y and the three truncation mutants) are to understand functions of different domains of p300. As shown in Fig. 2E, plasmids from 32 out of 38 colonies analyzed by Sanger sequencing carried the expected mutations, resulting in the mutagenesis efficiency of 84.2%.

For CBP, we sought to engineer 7 missense mutants and 2 C-terminal truncation mutants. The first four mutants (i.e., R1446C, R1466H, Y1503D and Y1503H; Fig. S2D) are derived from somatic mutations identified in cancer and deposited in the COSMIC and cBioPortal databases [3,4]. They are four hotspot mutations [3,4]. The remaining three missense mutants (K1832E, R1869Q and R1968W) and the two truncation mutants are to understand functions of different domains of CBP. As shown in Fig. S2D, plasmids from 17 out of 20 colonies analyzed by Sanger sequencing carried the expected mutations. Thus, the efficiency was 85.0%. Together with the high efficiency to engineer full-length KAT6A and p300 mutants, these results support that P3a mutagenesis is also highly efficient for engineering mutants encoded by large plasmids.

P3a mutagenesis for engineering missense spike variants of SARS-CoV-2

The coronavirus disease 2019 (COVID-19) pandemic resulted in tragic loss of life, affected health care and crippled the economy around the world. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the culprit [5]. Since its initial identification four years ago, it has evolved rapidly, gained various mutations and yielded many variants and subvariants [6].

Derived from Omicron variant, JN.1 and its descendants are currently contributing to active infections around the world. One major gene of variation encodes the spike protein, which is also the target of vaccines that have been or are being developed. Compared to Omicron variant, JN.1 and its descendants have gained additional substitutions in the spike protein, including R346T, L455S, F456L, Q493E and V1104L [43]. To elucidate the underlying pathological mechanisms, it is necessary to engineer such substitutions in the wild-type spike protein for analysis of their functional impact. We thus investigated whether P3a mutagenesis can be employed for rapidly engineering these mutations in the backbone of spike proteins of the D614G and Omicron variants. The D614G variant is a SARS-CoV-2 derivative that appeared in the early part of 2020 and led to development of many major variants that drove the pandemic. We also engineered L981F in the backbone of the spike protein of the D614G variant as this substitution is present in Omicron variant [6]. For L455S and F456L, we only utilized two primers, with one encoding F456L and another for both L455S and F456L, using the strategies illustrated in Fig. 1C. As shown in Fig. 3, among 32 colonies that we analyzed, 29 were the correct mutants and one was a mixed clone of the wild-type and a mutant, so the overall mutagenesis efficiency was 92.2%. Once the primers were available, the initial mutagenesis experiment was carried out on day 1 and the colonies were inoculated on day 2. On day 3, the plasmids were prepared and sent for Sanger sequencing. Typically, 1-3 days later, the sequencing results were available for analysis. Thus, the mutant plasmids could be obtained easily within a week. Thus, this method is applicable for rapidly generating and molecularly modeling new mutations that JN.1 and its descendants will gain during evolution.

P3a cassette mutagenesis to introduce seamless large deletion

Completely complementary primers (Fig. 1A) can be used to introduce small deletions (such as removal of a few base pairs), but it is quite challenging to introduce large deletions (such as those from a few hundred base pairs to several kb). This is a common problem for all known mutagenesis methods. One approach is to amplify by PCR and subclone a DNA fragment with deletion, but this is time-consuming. This fragment functions as a cassette, so such an approach is also known as 'cassette' mutagenesis. The 'handshaking' feature of the primer pair with 3′-overhangs inspired us to postulate that such a primer pair can be leveraged for introducing large deletions (Fig. 1D). Related to this, the ideal efficiency of 100% in generating the two mutants del364-394 and del364-427 of the HAT domain of KAT6B (Fig. S2B) made us reason that P3a mutagenesis can indeed be used to delete large DNA fragments. To explore this concept further, we tested it with additional vectors, some of which are much larger than that for the HAT domain of KAT6B (Fig. S2B). We first assessed the concept for generating two N-terminal truncation mutants of BRPF1 (Fig. 4A). Shown in Fig. 4B-C is the primer design strategy to delete 393-bp fragment encoding the N-terminal 131 residues. A similar strategy was used to design the primer pair for engineering dN204, the truncation mutant lack of the N-terminal 204 residues (Fig. 4A). All six colonies analyzed by Sanger sequencing carry the expected deletions (Fig. 4A), thereby reaching the ideal efficiency of 100%. We also generated a mutant with an internal deletion of 11 amino acids (d594-604) and obtained two corrected colonies out of three, yielding an efficiency of 66.7%(Fig. 4A).

Success with these three BRPF1 deletion mutants led us to apply a similar strategy to generate larger deletion mutants of a paralog, BRPF2, and a related protein, JADE2, as we needed such mutants to investigate their interaction with KAT7. As shown in Fig. 4A, for two BRPF2 deletion mutants and one JADE2 deletion mutant, we analyzed 11 colonies and found

that 8 were mutants, so the efficiency was 8/11 (72.3%). We next tested the strategy with the expression vectors for p300 and CBP, which are 12.8 and 13.4 kb in size, respectively. We aimed to engineer two mutants to delete the coding sequences for the N-terminal 1301 residues of p300 and the N-terminal 1068 residues of CBP. As shown in Fig. 4A, all six colonies analyzed by Sanger sequencing have the expected mutations, thereby reaching the ideal efficiency of 100%. These results attest to the notion that P3a cassette mutagenesis is highly efficient for introducing small or large deletion mutants.

Seamless epitope tagging and untagging by P3a cassette mutagenesis

The high efficiency in engineering deletions via P3a mutagenesis led us to investigate the possibility to use this strategy to introduce insertions (Fig. 1E). We first tested a mammalian expression vector for human p300, which does not possess a FLAG epitope for affinity purification on M2 agarose, so we sought to insert such a tag at the N-terminus of p300 (Fig. 5A). In the design, we also sought to remove 92 bp between the CMV promoter and the p300 coding sequence (Fig. 5B-C). As shown in Fig. 5A, all three colonies analyzed by Sanger sequencing are the expected mutants with the deletion of the 92-bp fragment and the insertion of the FLAG-coding sequence, thereby reaching the ideal efficiency of 100%. This supports that P3a cassette mutagenesis is highly efficient for epitope tagging.

We also tested the strategy to modify additional vectors. One such expression vector for CBP encodes two epitope tags, FLAG and HA, initially designed in another lab for tandem affinity purification. The HA tag interferes with detection of CBP partners in co-immunoprecipitation experiments, as they are also expressed as HA-tagged fusion proteins. We thus sought to remove the HA tag from the CBP expression vector using the strategy illustrated

in Fig. S3A-B. As shown in Fig. 5A, one of three colonies analyzed by Sanger sequencing carries the expected deletion. Similarly, the p300 expression vector has a 6xHis tag, which may cause non-specific interaction. We employed a strategy analogous to that depicted in Fig. S3A-B to remove this tag. One of three colonies analyzed by sequencing carries the expected deletion. These results support that P3a cassette mutagenesis is efficient for epitope untagging.

We next investigated epitope tag conversion. Related to this, we had purchased expression vectors for multiple epigenetic regulators, including SIN3A, YEATS2, ZZZ3, MIER1, PHF6 and PH12 (Fig. 5A). These vectors were initially designed for expression of FLAG-tagged proteins, but for co-immunoprecipitation with partners that are also expressed as FLAG-tagged proteins, it is necessary to replace the FLAG tag with another tag, such as the HA tag. For the SIN3A vector, we utilized a strategy depicted in Fig. S4A-B to achieve tag conversion. As shown in Figs. 5A and S4C, all three colonies analyzed by Sanger sequencing were the expected mutants, thereby reaching the ideal efficiency of 100%. Similar strategies were employed to replace the FLAG tag on the expression vectors for YEATS2, ZZZ3, MIER1, PHF6 and PH12. As shown in Figs. 5A and S4D, 12 of 15 colonies analyzed by Sanger sequencing were the expected deletion mutants, thereby reaching the efficiency of 75%. Using this strategy (Fig. S3C), we converted a FLAG expression vector to a 3xFLAG vector. Notably, the 'handshaking' region is only 18 bp (Fig. S3C), allowing 60-bp insertion with a pair of primers only 58 nucleotides long. This is based on the general strategy illustrated in Fig. 1E. As shown in Fig. 5A, 5 of 6 colonies analyzed by Sanger sequencing harbor the expected plasmids, thereby reaching an efficiency of 83.3%. The overall efficiency for manipulating epitope tags on the 10 vectors is 75.6% (Fig. 5A), attesting to the high efficiency of P3a cassette mutagenesis for epitope tagging and untagging.

As shown above for introducing epitope tags (Figs 5 & S3-S4), this method should also allow seamless insertion of oligonucleotide duplexes for engineering gRNA and shRNA vectors. Related to this, we tested a pair of primers to mutate an sgRNA-coding sequence on an sgRNA expression vector for marker-free genome editing (Addgene, Cat 86611, 8.4 kb [44]) and analyzed 8 of the resulting colonies. Among them, 6 contained the mutated plasmid, so the efficiency was 6/8 (75%), supporting that P3a cassette mutagenesis allows seamless insertion or removal of sgRNA-coding sequences.

Seamless restriction site engineering by P3a site-specific mutagenesis

High efficiency in deletion of large DNA fragments encoding different epigenetic regulators by P3a mutagenesis (Fig. 4) led us to consider the use of this method for plasmid engineering. For example, the mammalian expression vector for p300 is almost 12.8 kb (Fig. S5A) and its preparation yield from bacterial cultures is much lower than other plasmids of similar size. To enhance the yield, we considered to delete some unnecessary sequences. Related to this, the Fi origin that was designed decades ago for preparation of phagmid forms is obsolete and no longer needed (Fig. S5A). Moreover, the Neo gene marker and its promoters are not necessary either. Thus, P3a mutagenesis was carried out to delete 1,928-bp fragment encompassing the F1 origin and the Neo elements. The resulting plasmid was expected to be 10.8 kb (Fig. S5B). The two primers were designed as illustrated in Fig. S5C. As shown in Fig. S5D, 2 of 4 colonies analyzed by Sanger sequencing are the expected deletion mutants. The preparation yield of the mutant plasmids was still similar to the parental plasmid, indicating that sequences other than the deleted region are responsible for the low yield. Among the remaining two colonies, one was the original plasmid and the other contained a much larger deletion. Thus, the mutagenesis efficiency was

50%. Due to its reduced size, the resulting plasmid with the 1,928-bp deletion should allow easy insertion of other sequences, such as those of KAT6A and KAT6B for engineering related leukemic fusion proteins [39].

We next sought to engineer simultaneous double deletions for plasmid optimization. Like the p300 expression plasmid, many plasmids in use now contain unnecessary sequences that can be deleted for optimization. This is because molecular cloning and cell biology have evolved rapidly, thereby rendering some old sequence elements designed decades ago obsolete. One such example is the F1 origin, which was initially designed for phagmid isolation, but this technique is hardly used anymore. Thus, we sought to optimize a commonly used 5.4-kb vector in our laboratory. This vector, known as pAW51 (Fig. S6A), was initially derived from a pcDNA3.1 vector and encodes an HA tag. pAW51 contains two regions that are no longer needed. One of them is a 430-bp fragment containing the F1 origin, and the other is a 390-bp fragment that encompasses a CAP-binding site and a Lac operator. To optimize this plasmid, we sought to delete these two fragments. The resulting plasmid is smaller (4.6 kb, Fig. S6B). Two pairs of primers were designed (Fig. S6C-D). These four primers were used in the same mutagenesis reaction to introduce both deletions simultaneously. In the first trial, we used 15 ng of the parental plasmid as the PCR template (condition A, Fig. S6E). We analyzed plasmids from 9 colonies by restriction digestion (Fig. S6E). Two plasmids possess the 390-bp deletion and 4 contain the 430-bp deletion. Among these 6, one carries both deletions. As the efficiency to obtain the mutant plasmid with both deletions is low, 1/9 (11.1%), we tested two additional conditions by using 5 ng of the parental plasmid as the PCR template (conditions B-C, Fig. S6E). For each of these two conditions, we analyzed plasmids from 8 colonies by restriction digestion (Fig. S6E). For both conditions, 7 colonies contain one of the two deletions. For condition B,

three colonies harbor the mutant plasmid with both deletions (Fig. S6E). For condition C, none of the 8 colonies possess the mutant plasmid with both deletions (Fig. S6E). For these three trials, the overall efficiency to obtain the mutant plasmid with both deletions is only 4/25 (16%). These results suggest that while P3a mutagenesis is efficient in engineering single-site deletions,it is feasible for P3a mutagenesis to introduce two mutations simultaneously; however, alternative approaches are needed to enhance the efficiency. One possibility is to engineer two deletions with sequential mutagenesis reactions.

For plasmid engineering, it is often necessary to introduce or remove restriction sites for subcloning. Related to this, we needed to remove one of the two NheI site that pAW51 harbors to make this remaining site unique. Thee site to be deleted is located upstream from the coding sequence for the HA tag (Fig. S6A). This was achieved easily with the primers NheI-F and NheI-R, at the mutagenesis efficiency of 100% (Fig. S7A). Into the resulting vector (pAw51a), we then introduced an XhoI site and an NheI site with another pair of primers, XhoI-F and XhoI-R (Fig. S7B). Again, this was easily carried out, at the mutagenesis efficiency of 100% (Fig. S7A). Similarly, we introduced an HindIII site into pCL36, a vector almost the same as pAW51, but with a different poly-linker downstream from the coding sequence for the HA tag (Fig. S7A). As shown at the top of Fig. S8, we designed one pair of primers to introduce the HindIII site at two different reading frames. We analyzed 4 colonies by Sanger sequencing. All of them were the expected mutants, with 3 for one reading frame and the 4th for the other reading frame (Fig. S8), also reaching the ideal efficiency of 100%. For these cases, restriction digestion could be used as the initial screening for the resulting colonies, so the results could be obtained within 3 days after primers were synthesized. The candidates from the initial screens could then be validated by Sanger sequencing.

To produce mammalian expression vectors for human SIN3B, USP10 and histone H4, we obtained their coding sequences from the ORFeome Collaboration Collection [45]. We initially tried PCR to amplify the coding sequences for subcloning, but due to unknown reasons, the amplification failed. To overcome this technical problem, we sought to utilize P3a mutagenesis to introduce necessary restriction sites for subcloning. Into the vectors for SIN3B and USP10, we introduced an HindIII site upstream from their coding sequences. The resulting plasmids were analyzed by restriction digestion, and Sanger sequencing was performed for verification. The analyses revealed that the efficiency was 44.4% and 75% for introducing the HindIII sites into the SIN3B and USP10 expression vectors, respectively. The large size of the SING3B insert (~3.6 kb) might have contributed to the lower efficiency. For the histone H4 expression vector, we needed to introduce XhoI and BamHI sites upstream and downstream from the coding sequence, respectively. We designed two pairs of primers and used both pairs in one mutagenesis reaction. We analyzed 4 colonies by digestion. Two had the correct digestion pattern and were verified by Sanger sequencing, so the efficiency to obtain both restriction sites on the same vector is 2/4 (50%), supporting the feasibility of P3a mutagenesis to introduce two distant mutations simultaneously. Together with the results on deletion of large fragments from the p300 expression vector (Fig. S5) and pAW51 (Fig. S6), these experiments on introduction and deletion of restriction sites (Figs S7-S8) indicate that P3a mutagenesis is efficient and valuable for seamless plasmid engineering.

Discussion

Site-directed mutagenesis is a fundamental technique for biomedical research, used to analyze disease-associated germline or somatic mutations [2-4], assess the impact of viral genetic

alterations during evolution [5,6], generate specific mutants of natural proteins [7] and engineer artificial proteins from natural backbones [8,9] or AI-assisted *de novo* design [10-12]. An innovative strategy involves the use of partially complementary primers with 3′-protruding ends (Fig. 1B) [17,19,20]. These primers anneal to nicked regions of newly synthesized DNA strands, enabling their use as new templates for PCR amplification (Fig. S1A). In the previous study [17], we evaluated this method with a dozen mammalian expression plasmids. While it is generally reliable for many cases, the efficiency was close to 50% and can be improved further. Moreover, other issues, such as 1) unwanted mutations at the primer sites, 2) relatively low fidelity of Pfu polymerase and 3) slow synthesis rate of this enzyme, require further improvement of the method [17]. In the current study, we have developed a new approach by replacing Pfu and its derivatives with Q5 and Platinum SuperFi II high-fidelity DNA polymerases, both of which elevate the efficiency close to the ideal level of 100%, speed up PCR amplification and eliminate unwanted mutations at or close to the primer sites (Figs 1-3). Typically, the new method requires fewer PCR cycles than what was described previously [17]. We have evaluated the new method with over ten mammalian expression vectors encoding various epigenetic regulators (Figs 1-2, 4 & S1-S2) and two vectors for the spike protein of SARS-CoV-2 (Fig. 3).

In light of the unique handshaking feature of the primers with 3′-overhangs, we have extended the method for seamless cassette mutagenesis to engineer deletion (Fig. 1D) and insertion (Fig. 1E) as exemplified by introduction of large deletions (Fig. 4), epitope tagging/untagging (Figs. 5 & S3-S4) and plasmid optimization (Figs S5-S8). The largest plasmid that we tested is the 13.4-kb CBP expression vector (Figs. 4A & S2D). For seamless plasmid minimization, we tested the method with two plasmids (Figs. S5-S7). But it should be applicable

to many other plasmids, especially for those (such as viral vectors for preclinical research and gene therapy), where the size limit becomes a critical bottleneck.

For seamless epitope tagging and untagging, we tested the method with four tags, including FLAG, 3xFLAG, HA and 6xHis tags (Fig. 5A), but the method should be applicable to many other tags, such as a modified tandem affinity purification tag [46]. It needs to be noted, however, that the size of insertion is limited to the two primers that can be synthesized chemically. The largest insertion that we have made was the conversion of a FLAG tag to a 3xFLAG tag (Fig. 5A), where the inserted sequence is ~60 bp. As shown for introducing epitope tags (Figs S3-S4), this method should also allow efficient and seamless insertion of oligonucleotide duplexes for engineering gRNA and shRNA vectors. Similarly, this method shall also be useful for seamless insertion of LoxP and other recombination sites for construction of targeting vectors for homologous recombination and conditional knockouts. Theoretically, the method can also be used for seamless insertion or deletion of the coding sequences for nuclear localization and export signals.

One common challenge with site-directed mutagenesis is occurrence of unwanted mutations due to three sources: 1) impurity in primers, 2) unwanted recombination at the mutation or primer sites, and 3) misincorporation during synthesis mediated by DNA polymerase. Related to source 3), Q5 and SuperFi II DNA polymerases are high-fidelity enzymes with misincorporation rates 10-15-fold lower than PfuUltra that we used in the previous study [17] and 30-35-fold lower than regular Pfu. Related to this, results from over 100 Sanger sequencing reactions performed for the current study have not yielded any such mutations away from the primer sites. Related to source 2), the previous study frequently uncovered unwanted mutations at the region at or close to the primer sites when PfuUltra was used, perhaps due to its

strand-replacement activity or unwanted synthesis when reaching the 5′ end of an annealed primer [17]. By contrast, we have found only two such cases in the current study (Figs. S4D & S5D). One of them (Fig. S4D) is right at a primer site, so it could also be due to source 1). Instead of unwanted recombination at the primer sites or misincorporation during PCR, impurity in primers is a major source of unwanted mutations, so it is important to verify the sequence at the primer sites. The results from the current study and those from the previous study [17] also indicate that Q5 and SuperFi II polymerases replicate much more faithfully than Pfu and its derivatives when DNA synthesis reaches the primer sites.

The current study focuses on engineering single-site mutations. For pAW51 (Fig. S6) and the histone H4 vector (Fig. S7), we tested the feasibility to engineer double mutations simultaneously. The results show that the dual-site mutagenesis efficiency was 50% for the histone H4 vector (Fig. S7A) and only 16% for pAW51(Fig. S6). Theoretically, PCR amplification between the forward primer for site 1 and reverse primer for site 2, as well as PCR amplification between the reverse primer for site 1 and forward primer for site 2, becomes dominant and may overshadow the expected two PCR amplifications from two pairs of primers at sites 1 and 2 for P3a mutagenesis. This problem is expected to get worse for multi-site mutagenesis. This is an intrinsic flaw of all mutagenesis methods based on primer pairs, including the QuickChange™ and P3 site-directed mutagenesis methods (Fig. 1A-B). One potential solution is to carry out sequential mutagenesis, but this is time-consuming, especially for multi-site mutagenesis where the site number is 5 or more. Therefore, alternative approaches are needed to generate multi-site mutations fast and efficiently.

In summary, we have developed a fast, economical and highly efficient site-specific mutagenesis method based on Q5 and SuperFi II high-fidelity DNA polymerases, along with

primer pairs with 3′-overhangs (Fig. 1). We have evaluated this method systematically by engineering >100 mutations on >20 mammalian expression vectors, up to 13.4 kb in size. The results indicate that the method is superior to those based on Pfu and its derivatives [17,19,20], by elevating the efficiency to or close to 100%, speeding up PCR amplification and reducing unwanted mutations at the primer sites. Even more importantly, the high efficiency and the 'handshaking' feature of the primer pairs with 3′-protruding ends have led us to extend the method for seamless cassette mutagenesis, which allows highly efficient epitope tagging or untagging, deletion of small or large DNA fragments, optimization of plasmid constructs and insertion of oligonucleotide duplexes with 3′-overhangs. Given its versatility, P3a site-specific and cassette mutagenesis is expected to be of wide use in biomedical research.

CONFLICT OF INTEREST

We declare no conflict of interests.

ACKNOWLEDGEMENT

We thank Paulina Varela-Castillo and Arezousadat Razavi for critically reading the manuscript. This work was supported by funds from Canadian Institutes of Health Research (CIHR), Natural Sciences and Engineering Research Council of Canada (NSERC) and Compute Canada (to X.J.Y.).

Materials and Methods

Plasmid vectors

The mammalian expression vector for wild-type *Xenopus* histone H3 with its C-terminus fused a FLAG tag was modified from a pcDNA3.1 derivative containing the coding sequence for this epitope tag. A bacterial expression vector encoding *Xenopus* histone H3 was generously provided by Karolin Luger [47]. Mammalian expression vectors for HA-tagged BRPF1, BRPF2 (also known as BRD1) and BRPF3 were described earlier [27,37,48]. The mammalian expression vector for FLAG-KAT6A was kindly provided by Dr. Issay Kitabayashi [49]. The mammalian expression vector for the catalytic domain of KAT6A was described previously [27,48]. The mammalian expression vectors for FLAG-KAT6B and its catalytic domain have been reported [27,48,50]. The expression plasmids for untagged p300 and FLAG/HA-tagged mouse CBP were obtained from Addgene (Cat. 23252 and 32908, respectively). Mammalian expression vectors for untagged D614G and Omicron spike proteins of SARS-CoV-2 were purchased from SinoBiological (Cat. VG40589-UT and VG40835-UT, respectively). Vectors containing the coding sequences for human SIN3B (BC172412), histone H4 (EU446978) and USP10 (DQ892553), initially produced as a part of the ORFeome Collaboration Collection [45], were obtained from the Platform for Cellular Perturbation at McGill University, Montreal, Canada. The mammalian expression vectors for human SIN3A, YEATS2, ZZZ3, MIER1, PHF6 and PHF12 were purchased from Genescript. They were expressed as fusion proteins with a FLAG tag fused to the C-termini, upon transfection into mammalian cells.

Primers for site-directed mutagenesis

Primers were designed with the aid of SnapGene software package (version 7.2.1). They were synthesized at Integrated DNA Technologies (IDT) as small scale, standard and desalted DNA oligos, without further purification. Primers used to generate Xenopus histone H3 mutants for expression in mammalian cells as fusion proteins with a FLAG tag at the C-terminus were as listed below, where the mutated nucleotides are shown in lowercase: K9R-F, CCGCCCGGAgATCCACCGGAGGGAAGGCTCCCCGC; K9R, CCGGTGGATcTCCGGGCGGTCTGCTTAGTACGAGC; K14R-F, CCGGAGGGAgGGCTCCCCGCAAGCAGCTGGCCACC; K14R-R, CGGGGAGCCcTCCCTCCGGTGGATTTCCGGGCGGT; K18R-F CTCCCCGCAgGCAGCTGGCCACCAAGGCAGCCAGG, K18R-R, GCCAGCTGCcTGCGGGGAGCCTTCCCTCCGGTGGA; K23R-F, TGGCCACCAgGGCAGCCAGGAAGAGCGCTCCGGCC; K23R-R, CTGGCTGCCcTGGTGGCCAGCTGCTTGCGGGGAGC; K27M-F, CAGCCAGGAtGAGCGCTCCGGCCACAGGCGGAGTC; 27R-R, GGAGCGCTCcTCCTGGCTGCCTTGGTGGCCAGCTG; G34R-F, GCCACAGGCaGAGTCAAGAAACCTCACCGTTACCG; G34A-R, TTCTTGACTgCGCCTGTGGCCGGAGCGCTCTTCCT, K36M-F, GCGGAGTCAtGAAACCTCACCGTTACCGGCCCGGC; and K36R-R TGAGGTTTCcTGACTCCGCCTGTGGCCGGAGCGCT.

The primers for engineering plasmids expressing deletion mutants of human BRPF1 were dN131-F, CGCCGAATTCaATGGCAGCAACAAGGAGAA; dN131-R, TGCTGCCATtGAATTCGGCGTAGTCAGGCA; dN204-F, CGCCGAATTCtCTGCAGAGGAGCTGGACGA; dN204-R,

CCTCTGCAGaGAATTCGGCGTAGTCAGGCA; d594-604-F,

CCTGGCAGCGGGTGGAATTGATCCGCAAGCG; d594-604-R,

GATCAATTCCACCCGCTGCCAGGACTTGAGCT; Y215A-F,

GAAGTAGAGgcTGACATGGACGAGGAGGACT; Y215A-R

TCCATGTCAgcCTCTACTTCCTCGTCCAGCT; W224A-F

GACTACATCgcGCTGGATATCATGAATGAGC; W224A-R

ATATCCAGCgcGATGTAGTCCTCCTCGTCCA; F245A-F

CAGGAGATCgcTGAGTACCTAATGGACCGAC; F245A-R

AGGTACTCAgcGATCTCCTGCGGGATGGGAC; R381A-F

CCACCAGCTgcCTGGAAGCTCACCTGCTACA; R381A-R

AGCTTCCAGgcAGCTGGTGGGATGTGCTCAA; K390A-F

TACATTTGCgcACAACGGGGCTCAGGGGCCT; K390A-R

CCCCGTTGTgcGCAAATGTAGCAGGTGAGCT; R392A-F

TGCAAACAAgcGGGCTCAGGGGCCTGCATCC; and R392A-R

CCTGAGCCCgcTTGTTTGCAAATGTAGCAGG, where the mutation is denoted with lowercase letters.

The primers to engineerthe expression plasmids for human BRD1 mutant were as follows: G38R-F, CAAGCTCAAaGGATGGTAGAGATAGAAATT; G38R-R

CTACCATCCtTTGAGCTTGAGCGTAGGTCA; I53A-F,

ACAGGATCAGTgcTTTTGATCCCCTGGAGATCA; I53A-R,

GGGATCAAAAgcACTGATCCTGTGCAAGCGCC; L57A-F,

TTTTTGATCCCgctGAGATCATATTGGAAGATGA; L57A-R,

AATATGATCTCagcGGGATCAAAAATACTGATCC; I59A-F,

TCCCCTGGAGgcCATATTGGAAGATGACCTCA; I59A-R,

CTTCCAATATGgcCTCCAGGGGATCAAAAATA; I60A-F,

CCTGGAGATCgcATTGGAAGATGACCTCACTG; I60A-R,

ATCTTCCAATgcGATCTCCAGGGGATCAAAAA; L61A-F,

GGAGATCATAgcGGAAGATGACCTCACTGC; L61A-R,

GTCATCTTCCgcTATGATCTCCAGGGGATCAA; P275A-F,

GTGCTGTGCgCCAACAAGGGTGGTGCCTTC; P275A-R,

CCTTGTTGGcGCACAGCACACAGTCGGCGG; H275P-F,

TGCTGTGCCCCAACAAGGGTGGTGCCTTCA; H275P-R,

CCCTTGTTGGGGCACAGCACACAGTCGGCG; d31-80-F,

GCGAGAAACggAGCGGCCTCCTGTCTGCTT; d31-80-R1,

GAGGCCGCTccGTTTCTCGCGTAGGTGAGTGTTTA; dN147-F,

CGCCGAATTCGAGGAACTGGACAACGAGGT; and dN147-R,

CCAGTTCCTCGAATTCGGCGTAGTCAGGCA.

The primers for engineering mutant plasmids for FLAG-tagged human KAT6A were as follows: K604R-F, TTGACCACAgAACCCTCTATTACGATGTGG; K604R-R TAGAGGGTTcTGTGGTCAAGAAACAACTTT; K350Kfs-F, ACGGTATCAAAGGTCCCTTCAGCAAAGTTCGAACTGGCCCTGGAA; K350Kfs-R TGAAGGGACCTTTGATACCGTGTTTTGTTTCTTTAACCTGTTTTT; R469*-F**,** AATGAGGAGtGACTTTTTGGGAGCCAGGAA**;** L668F-F CAGTTATTTcTTATCAAAGCGTGAAGGCCAAGCAG; L668F-R, CTTTGATAAgAAATAACTGAAATCGATGAGAAACC; P682Q-F, CAGAGAAACaGTTATCTGATCTGGGTCGTCTTTCC; P682L-R,

TCAGATAACaGTTTCTCTGGAGACCCTGCTTGGCC; Q743R-F,

GTAGTGACCgATTTGTGATTATCCGCCGGGAAAAA; Q743R-R,

ATCACAAATcGGTCACTACGGAAGTCCAGCATTCG; R740H-F,

TGGACTTCCaTAGTGACCAATTTGTGATTATCCGC; R740L-R,

TGGTCACTAaGGAAGTCCAGCATTCGTAGGTGGTG;

R469*-R, CAAAAAGTCaCTCCTCATTTTCTTGTTTGC; G814*-F,

ATCAGTGTGtGAAAGTCTGTGTCTCATGAG; G814*-R,

CAGACTTTCaCACACTGATCTCTAATTCTC; L1061*-F,

ATGCCAAGATaAGAACCCACGTTTGAGATCGATGAAGAAGAGGAG; L1061*-R

,CGTGGGTTCTtATCTTGGCATTGGCCTCTCGGAGTCAGAATCTTC; E1393*-F,

GCTGGGTCTtAGGACGACCACGAAGAAGAC; and E1393*-R,

GGTCGTCCTaAGACCCAGCCATCTGCTCTG.

The primers for engineering mutant mammalian vectors to express human KAT6B as a FLAG-tagged fusion protein were as follows: F551L-F,

GTTGGATACCTCTCTAAGGAAAAGCTTTGC;

F551L-R, CCTTAGAGAGGTATCCAACCAGATGACAGC;

Q863H-F, CATGCCCCAcCACCAAAGGCAAGGATTTGGACGGT; Q863H-R,

CCTTTGGTGgTGGGGCATGATCATTATGCAGGAGA; E884G-F,

CTAGAAGAGgAGGCCAAGCAGGGTCTCCTGAAAAG; E884G-R,

GCTTGGCCTcCTCTTCTAGAAAGCAAATAGCTGAA; S889C-F,

AAGCAGGGTgTCCTGAAAAGCCTCTCTCCGATCTG; S889C-R,

TTTTCAGGAcACCCTGCTTGGCCTTCTCTTCTAGA; D896N-F,

CCTCTCTCCaATCTGGGCCGTCTCTCCTACCTGGC; D896N-R,

GGCCCAGATtGGAGAGAGGCTTTTCAGGAGACCCT; R899H-F, ATCTGGGCCaTCTCTCCTACCTGGCATATTGGAAG; R899H-R, TAGGAGAGAtGGCCCAGATCGGAGAGAGGCTTTTC; E912K-F, GTCATCTTGaAGTATCTCTACCACCACCATGAGAG; E912K-R, AGAGATACTtCAAGATGACGCTCTTCCAATATGCC; d364-394-F, ATTCGGATCCGTCGTTACTGAAGAGGATTT; d364-394-R, CAGTAACGACGGATCCGAATTCGGTACCTT; d364-427-F, ATTCGGATCCCCTTCTGTGATTGAATTTGG; and d364-427-R, TCACAGAAGGGGATCCGAATTCGGTACCTT.

The primers for engineering mutant plasmids for human KAT7 were as follows: C185S-F, ATATGAAGTcTCCTACACCAGGCTGTAACTCT; C185S-R, GGTGTAGGAgACTTCATATTGAAGTTGTAGCT; C209S-F, TCTCAGGATcCCCACTGTATCATAACCTCTCA; C209S-R, TACAGTGGGgATCCTGAGATGGAGAAATGTCT; K432R-F, CTGGACCACcgtACATTATATTATGATGTGGAGCCCT; K432R-R, AATATAATGTacgGTGGTCCAGAAAcAGTTTGGCCAA; I537A-F, TTTTCAAGGCgcAGAGATTTCTATCAAAGAAA; I537A-R, AGAAATCTCTgcGCCTTGAAAATTATGCAGGT; I539A-F, AGGCAAAGAGgcTTCTATCAAAGAAATCAGTC; I539A-R, TTTGATAGAAgcCTCTTTGCCTTGAAAATTAT; I541A-F, AAGAGATTTCTgcCAAAGAAATCAGTCAGGAGAC; I541A-R, GATTTCTTTGgcAGAAATCTCTTTGCCTTGAAA; V574A-F, AACACCTAGcTTTAAAGAGACAGGACCTGA; V574A-R,

CTCTTTAAAgCTAGGTGTTTTCCCTTCCAG; K576A-F,

ACCTAGTTTTAgccAGACAGGACCTGATTGATGA; and K576A-R,

AGGTCCTGTCTggcTAAAACTAGGTGTTTTCCCT.

The primers for engineering mutant plasmids for human KAT8 were as follows: E87D-F GAGGGCCGAGAtGAATTCTATGTACACTACGT; E87Q-R,

ATAGAATTCCTgTCGGCCCTCCTGGTCGTTCA; V91I-F,

GAATTCTATaTACACTACGTGGGCTTTAAC; V91I-R,

GGAAGTCCCaCAGGATCTCTAGCAGCACCC; R136C-F,

GCAGCCTGAGtGCAAGATCACTCGCAACCAA, R136H-R,

AGTGATCTTGtGCTCAGGCTGCTCTGCGAGC; R140H-F,

AAGATCACTCaCAACCAAAAGCGCAAGCATG; R140G-R,

TTTTGGTTGCcAGTGATCTTGCGCTCAGGCT; D190G-F,

ACGAAATTGgTGCCTGGTATTTCTCACCAT; D190G-R,

CATCGGGAACTACGAAATTGgTGCCTGGTA; R224H-F,

AAGAGCTACCaCTTCCACTTGGGTCAGTGCC; R224G-R,

AAGTGGAAGCcGTAGCTCTTCTCATATTTCA; Y241C-F,

AAGAGATCTgCCGCAAGAGCAACATCTCCG; Y241C-R,

TCTTGCGGcAGATCTCTTTCCCGGGtGGCT; P321H-F,

TGACCTTGCaCCCCTACCAACGCCGCGGCTA; P322L-R,

GCGTTGGTAGaGGGGCAAGGTCAGGATGCAG; R374W-F,

GAGATCCTGtGGGACTTCCGGGGCACACTG; and R374W-R,

AGGGCCGAGAGGAATTCTATaTACACTACG.

The primers for engineering mutant plasmids for human p300 were as follows:

p300-F, ACTACAAGGACGACGATGACAAAGCCGAGAATGTGGTGGAACC; p300-R, TCATCGTCGTCCTTGTAGTCCATGCTCGAGTCTAGAGGGCCCG; dN1031-F, ACTACAAGGACGACGATGACAAAAGTACTTCAGCTACCCAGTC; d6His-F, CTAGACATATAGTGATACTAAGCTTAAGT; d6His-R, GTATCACTATATGTCTAGTGTACTCTGTG; d2K-F, GGTATCCCCTCATGTCTGTATACCGTCGA; d2K-R, CAGACATGAGGGGATACCCCCTAGAGCCC; C1204S-F, TGTGAGAAGTcTTTCAATGAGATCCAAGGG ; C1204R-R, TCATTGAAACgCTTCTCACAGAAATGATACC; E1242K-F, ACTGGATCCTaAACTGTTTGTTGAATGTACA; E1242A-R; ACAAACAGTgCAGGATCCAGTGTGTCATTT; D1384N-F, TATGGCTCTaACTGCCCTCCACCCAACCAG; D1384H-R, GAGGGCAGTgAGAGCCATACTCTTGAACAT; D1399Y-F, ATCTTACCTCtATAGTGTTCATTTCTTCCGT; D1399N-F, GCCTGGGCTaCTGGGTGGCTGCAGCCTGCTY1414C-F, CTGCAGTCTgTCATGAAATCCTAATTGGAT; Y1414C-R, ATGAACACTATtGAGGTAAGATATGTATACT; W1466C-F, GCAGGAATGcTACAAgAAAATGCTTGACAAGGCTGT; Y1467N-R, ATACACAAGgCATAATCCTCACAGACAGTA; S1726*-F, GCCACCCAGtAGCCCAGGCGATTCTCGCCG; S1726*-R, ATTTCATGAcAGACTGCAGTCCTCAAGCAT; R1830*-F, ATGCTTCGCtAGGAGGATGGCCAGCATGCA; R1830*-R, CATCCTCCTaGCGAAGCATTTGGGCCTGCT; Q1845*-F

GGGCAGCAAtAGGGCCTCCCTTCCCCCACT; and Q1845*-R, GGAGGCCCTaTTGCTGCCCAACCACACCAG.

The primers for engineering mutant plasmids for mouse CBP were as follows: dN1068-F, CGACGACAAAGACACAGCCTCACAATCAAC; dN1068-R, AGGCTGTGTCTTTGTCGTCGTCGTCCTTGT; R1446C-F CCGCTGCCTCtgtACAGCTGTTTACCATGAGAT; R1446H-R, AACAGCTGTatgGAGGCAGCGGGGCCGGAAGA; Y1503D-F, ACAGGAGTGGgACAAGAAGATGCTGGACAAG; Y1503H-R, GGATTCCTGtGGGCTCTaGGACTGTGGCTCACCCT; R1868*-F, GCTCATGCGCtaaCGAATGGCAACCATGAACAC; R1868*-R, TTGCCATTCGttaGCGCATGAGCTGAGCCTGCT; Q1881*-F, GTGCCTCAGtAGAGTTTGCCTTCTCCTACC; and Q1881*-R, GCAAACTCTaCTGAGGCACATTGCGGGTGT.

The primers for generating mutants of the D614G and Omicron spike proteins were previously described [21]. The backward primers for engineering an HA tag fused to the C-terminus of MIER1, SIN3A, PHF6, PHF12, ZZZ3 and YEATS2 were HA-MIER1 (TAGTCAGGCACGTCGTATGGGtagTCATCTGTGTTTTCAAGTT), HA-SIN3A (TAGTCAGGCACGTCGTATGGGtaaGGGGCTTTGAATACTGTGC), HA-PHF6 (TAGTCAGGCACGTCGTATGGGtaGTTTCCATTAAGTTGCTGCT), HA-PHF12 (TAGTCAGGCACGTCGTATGGGtaaGGAACAGAGTTGGAGCGCA), HA-ZZZ3 (TAGTCAGGCACGTCGTATGGGtaTCTGTTTGCTGGAAAGTAGT), HA-YEATS2 (TAGTCAGGCACGTCGTATGGGtaCTGGTCCTCATTCAATATTC),

respectively. The forward primer was the same: HA-F4, CCATACGACGTGCCTGACTACGCCTGATAAACCCGCTGATCAGC. The original plasmids, purchased from Genescript, were based on a pCDNA3.1 derivative encoding a FLAG tag fused to the C-terminus of a target protein. To convert the coding sequence for a FLAG tag on pAW48 (a pCDNA3.1 derivative) to that for a 3xFLAG tag, the two primers used were as follows: 3Flag-F, TGATAAGGGCAGCGATTATAAGGATGACGACGATAAGGAATTCCACCACACTGGACTA; and 3Flag-R, TAATCGCTGCCCTTATCATCATCGTCTTTATAATCTGCACCTTTGTCATCGTCGTCCT. Primers for modifying the sgRNA expression vector eSpCas9(1.1)_No_FLAG_ATP1A1_G3 (Addgene #86611) were OB352-F1, CACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGC; OB352-R, CTTCTCGAAGACCCGGTGTTTCGTCCTTTCCAC.

Primers for modifying pAW51, a pcDNA3.1 derivative encoding an HA tag, were as follows: NheI-F, AGCTGGCTAcgGTTTAAACGGGCCCTCTAGA; NheI-R, CGTTTAAACcgTAGCCAGCTTGGGTCTCCCT; XhoI-F, CTCGAGGATATCGCTAgcCACCACACTGGACTAGTGGA; and XhoI-R, GGCTAGCGATATCCTCGAGGAATTCGGCGTAGTCAGGCA. Primers for introducing an HindIII site into pCL36, another pcDNA3.1 derivative encoding an HA tag, were 36H-F GGAATTCTGaagcttTCCAGCACAGTGGCGGCCGC and 36H1-R, TGTGCTGGAaagctttCAGAATTCCGGCGTAGTCAG. These two primers led to two constructs with different reading frames downstream from the coding sequence for the HA tag.

Primers to introduce an HindIII site into the 5′ region of the encoding sequence for human Sin3B were SIN3B-HF, CCCACCATGaagcttGCGCACGCTGGCGGTGGCAGC and SIN3B-HR, AGCGTGCGCaagcttCATGGTGGGCCTTGACGGCC. Primers to introduce an XhoI site and a BamHI site flanking the encoding sequence of human histone H4 were HisH4-XF1, AGCAGGCTCgagCACCATGTCTGGTCGCGGCA; HisH4-XR1, GACATGGTGctcGAGCCTGCTTgTTTGTACAAAG; HisH4-BF1, CTTCGGCGGggatccTTGGACCCAGCTTTCTTGTA; and HisH4-BR1, TGGGTCCAAggatccCCGCCGAAGCCGTAAAGAGT.

P3 mutagenesis via Q5 and SuperFi II high-fidelity DNA polymerases

For a mutagenesis reaction, the template plasmid needed to be isolated from DH5α or any other Dam$^+$ *E. coli* hosts, as required for subsequent DpnI digestion. Primers were designed with the aid of the SnapGene software package and purchased from IDT as small-scale desalted oligos, without further purification. Their sequences were listed above. If primer sequences contain long stretches of A and/or G (typically more than 5), a silent mutation was introduced to break such stretches. This was also done with long stretches of T or C (typically more than 5). All oligonucleotides were used without polyacrylamide gel or HPLC purification. Upon receipt from IDT, autoclaved Nanopure water was used to dissolve lyophilized oligonucleotides to prepare 100 mM (i.e. 100 pmol/μl) stocks for further dilution to 10 mM (i.e. 10 pmol/μl) working solutions.

10 μl PCR reactions were set up in 0.2 ml 8-strip thin-wall PCR tubes (Diamed Cat. DIATEC420-1378), with each reaction containing 0.1-0.15 μl of plasmid DNA (0.1 μg/μl), 0.5 μl forward/reverse primer mixture (2.5 pmol/μl for each primer), 4.35-4.40 μl autoclaved

Nanopure water and 5 μl 2x Q5 hot-start master mix (New England Biolabs, Cat. M0494S), 2x Platinum SuperFi II PCR master mix (Thermo Fisher Scientific, Cat. 12368010), 2x i7® high-fidelity DNA polymerase master mix (Intact Genomics, Cat. 3257 or 3257S) or 2x KOD hot-start master mix (Sigma-Aldrich, Cat. 71842). The amplification was carried out in a Bio-Rad PCR T100 Thermal Cycler using the following parameters: 96°C for three minutes as the initial step to denature plasmid DNA, followed by 20-25 cycles to amplify the DNA. Each amplification cycle was composed of 93°C for 15 seconds as the denaturation step, 50-52°C for 20 seconds as the annealing step and 72°C for 3-5 minutes as the extension step, where the extension time was calculated according to the plasmid size (20-30 seconds per kb). The amplification cycle number varied slightly from one plasmid to another, with the initial number set at 20. After amplification, an additional extension step of 72°C for 7.5-10 minutes was added to the PCR program. After that, the reaction was pre-programmed for short-term storage at 4°C. For DpnI digestion, 0.25 μl (20 units/μl; New England Biolabs, Cat. R0176) was added to each reaction mixture, which was then pipetted into a clean PCR tube for incubation at 37°C for 90 min in a Bio-Rad T100 Thermal Cycler. The clean PCR tube at this step helped eliminate contamination of undigested plasmid from the inner wall of the old PCR tube used for PCR amplification. The reagent per mutagenesis reaction was $0.5-1. To reduce the cost further, the reaction volume could be reduced to 5 μl, with 1-2 μl used for transformation (see below).

Preparation and transformation of competent *E. coli* cells

DH5α competent cells were prepared and stored in liquid nitrogen tanks as described previously [17,18]. Briefly, 2-2.5 μl of the digested PCR mixture was gently mixed with 20-25 μl of DH5α competent cells, as specifically prepared below, in a pre-chilled 0.5 ml sterile Eppendorf tube

and incubated on ice for 30 min. Even transformation of 1.0 μl of the digested PCR mixture into 10 μl of DH5α yielded sufficient colonies for analysis, which could save materials further, The remaining DpnI-digested PCR mixture was frozen at -20°C as the backup. After the 30 min-incubation on ice, the tube containing the DH5α competent cells and DpnI-digested PCR mixture was heat-shocked at 42°C for 45 seconds and immediately transferred back onto ice. 60-75 μl of cold SOC or NYZ^+ medium [17,18] was then added to the tube and after gentle mixing, the tube was incubated at a 37°C water bath for ~30 min (without shaking, as the cells are fragile at this stage). Afterwards, the entire cell mixture was pipetted onto a LB-agar plate (containing the appropiate antibiotic as required for the plasmid to be mutated) as 4-5 drops, which were further spread out relatively evenly with the outside surface at the bottom half of an autoclaved 1.5 ml Eppendorf tube. After that, the plate was placed in a 37°C incubator for 22-24 h to allow large bacterial colonies to appear and grow.

Analysis of plasmids from bacterial colonies

Plasmids were isolated from 2-3 bacterial colonies per mutation for Sanger sequencing as described previously [17,18]. The resulting DNA sequences were compared with that of the wild-type template for the presence of the desired mutations with the aid of the sequence alignment tools in the SnapGene software package. The corresponding chromatograms were manually inspected for sequencing quality. The resulting DNA sequences and their corresponding chromatograms were also examined manually for potential unwanted mutations.

Plasmid sequencing and analysis

Plasmid isolation, sequencing and sequence analysis were performed as described above [17]. Briefly, after isolation with a Qiagen QIAprep® Spin Miniprep Kit, plasmids were sequenced at Genome Quebec technological service center. For sequencing reactions, plasmids were denatured at 96°C, after which primer annealing was carried out at lower temperatures not exceeding 50°C to allow hybridization of sequencing primers. Samples from the sequencing reactions were analyzed on 96-capillary array DNA Sequencer 3730XL (Applied Biosystems). The resulting sequences and chromatograms were transferred to Snapgene 8.0.1 for alignment analysis and manual inspection.

Estimation of misincorporation rate

DNA sequence alignment tools in the SnapGene software package were used to analyze Sanger DNA sequencing data. Sequence data quality was manually assessed from the corresponding sequencing chromatograms, and the aligned sequences were visually inspected to identify unwanted mutations. To calculate the misincorporation rate, the number of unwanted mutations was divided by the total number of the sequenced nucleotides whose corresponding sequencing chromatograms were inspected and determined to be of high quality.

References

1. C. A. Hutchison, 3rd, S. Phillips, M. H. Edgell, S. Gillam, P. Jahnke and M. Smith (1978) Mutagenesis at a specific position in a DNA sequence. *J Biol Chem* **253**, 6551-6560.
2. A. Henrie, S. E. Hemphill, N. Ruiz-Schultz, B. Cushman, M. T. DiStefano, D. Azzariti, S. M. Harrison, H. L. Rehm and K. Eilbeck (2018) ClinVar Miner: Demonstrating utility of a Web-based tool for viewing and filtering ClinVar data. *Hum Mutat* **39**, 1051-1060.
3. E. Cerami, J. Gao, U. Dogrusoz, B. E. Gross, S. O. Sumer, B. A. Aksoy, A. Jacobsen, C. J. Byrne, M. L. Heuer, E. Larsson, Y. Antipin, B. Reva, A. P. Goldberg, C. Sander and N. Schultz (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404.
4. J. G. Tate, S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, P. Fish, B. Harsha, C. Hathaway, S. C. Jupe, C. Y. Kok, K. Noble, L. Ponting, C. C. Ramshaw, C. E. Rye, H. E. Speedy, R. Stefancsik, S. L. Thompson, S. Wang, S. Ward, P. J. Campbell and S. A. Forbes (2019) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, D941-D947.
5. B. Hu, H. Guo, P. Zhou and Z. L. Shi (2021) Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* **19**, 141-154.
6. P. V. Markov, M. Ghafari, M. Beer, K. Lythgoe, P. Simmonds, N. I. Stilianakis and A. Katzourakis (2023) The evolution of SARS-CoV-2. *Nat Rev Microbiol* **21**, 361-379.
7. D. G. L. Thean, H. Y. Chu, J. H. C. Fong, B. K. C. Chan, P. Zhou, C. C. S. Kwok, Y. M. Chan, S. Y. L. Mak, G. C. G. Choi, J. W. K. Ho, Z. Zheng and A. S. L. Wong (2022) Machine learning-coupled combinatorial mutagenesis enables resource-efficient engineering of CRISPR-Cas9 genome editor activities. *Nat Commun* **13**, 2219.
8. G. D. Martyn, G. Veggiani, U. Kusebauch, S. R. Morrone, B. P. Yates, A. U. Singer, J. Tong, N. Manczyk, G. Gish, Z. Sun, I. Kurinov, F. Sicheri, M. F. Moran, R. L. Moritz and S. S. Sidhu (2022) Engineered SH2 Domains for Targeted Phosphoproteomics. *ACS Chem Biol* **17**, 1472-1484.
9. Y. Bian, L. Li, M. Dong, X. Liu, T. Kaneko, K. Cheng, H. Liu, C. Voss, X. Cao, Y. Wang, D. Litchfield, M. Ye, S. S. Li and H. Zou (2016) Ultra-deep tyrosine phosphoproteomics enabled by a phosphotyrosine superbinder. *Nat Chem Biol* **12**, 959-966.
10. T. N. Starr, N. Czudnochowski, Z. Liu, F. Zatta, Y. J. Park, A. Addetia, D. Pinto, M. Beltramello, P. Hernandez, A. J. Greaney, R. Marzi, W. G. Glass, I. Zhang, A. S. Dingens, J. E. Bowen, M. A. Tortorici, A. C. Walls, J. A. Wojcechowskyj, A. De Marco, L. E. Rosen, J. Zhou, M. Montiel-Ruiz, H. Kaiser, J. R. Dillen, H. Tucker, J. Bassi, C. Silacci-Fregni, M. P. Housley, J. di Iulio, G. Lombardo, M. Agostini, N. Sprugasci, K. Culap, S. Jaconi, M. Meury, E. Dellota, Jr., R. Abdelnabi, S. C. Foo, E. Cameroni, S. Stumpf, T. I. Croll, J. C. Nix, C. Havenar-Daughton, L. Piccoli, F. Benigni, J. Neyts, A. Telenti, F. A. Lempp, M. S. Pizzuto, J. D. Chodera, C. M. Hebner, H. W. Virgin, S. P. J. Whelan, D. Veesler, D. Corti, J. D. Bloom and G. Snell (2021) SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature* **597**, 97-102.
11. S. L. Lisanza, J. M. Gershon, S. W. K. Tipps, J. N. Sims, L. Arnoldt, S. J. Hendel, M. K. Simma, G. Liu, M. Yase, H. Wu, C. D. Tharp, X. Li, A. Kang, E. Brackenbrough, A. K. Bera, S. Gerben, B. J. Wittmann, A. C. McShan and D. Baker (2024) Multistate and functional protein design

using RoseTTAFold sequence space diffusion. *Nat Biotechnol* 10.1038/s41587-024-02395-w.
12. T. Kortemme (2024) De novo protein design-From new structures to programmable functions. *Cell* **187**, 526-544.
13. T. A. Kunkel (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci U S A* **82**, 488-492.
14. M. P. Weiner, G. L. Costa, W. Schoettlin, J. Cline, E. Mathur and J. C. Bauer (1994) Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. *Gene* **151**, 119-123.
15. C. L. Fisher and G. K. Pei (1997) Modification of a PCR-based site-directed mutagenesis method. *Biotechniques* **23**, 570-571, 574.
16. S. Li and M. F. Wilkinson (1997) Site-directed mutagenesis: a two-step method using PCR and DpnI. *Biotechniques* **23**, 588-590.
17. N. Mousavi, Zhou, E., Razavi, A., Ebrahimi, E., Varela-Castillo, P. and Yang, X. J. (2025) P3 site-directed mutagenesis: An efficient method based on primer pairs with 3'-overhangs. *J. Biol. Chem.*, ePub (also avaliable at bioRxiv: https://www.biorxiv.org/content/10.1101/2024.1110.1118.615663v615662).
18. N. Mousavi, Zhou, E., Razavi, A., Ebrahimi, E., Varela-Castillo, P. and Yang, X. J. (2025) Efficient site-directed mutagenesis mediated by primer pairs with 3'-overhangs. *Current Protcols in Protein Sciences*, in press.
19. L. Zheng, U. Baumann and J. L. Reymond (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* **32**, e115.
20. H. Liu and J. H. Naismith (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* **8**, 91.
21. H. H. Hogrefe, C. J. Hansen, B. R. Scott and K. B. Nielson (2002) Archaeal dUTPase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation. *Proc Natl Acad Sci U S A* **99**, 596-601.
22. Y. Wang, D. E. Prosen, L. Mei, J. C. Sullivan, M. Finney and P. B. Vander Horn (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Res* **32**, 1197-1207.
23. B. D. Strahl and C. D. Allis (2000) The language of covalent histone modifications. *Nature* **403**, 41-45.
24. B. A. Nacev, L. Feng, J. D. Bagert, A. E. Lemiesz, J. Gao, A. A. Soshnev, R. Kundra, N. Schultz, T. W. Muir and C. D. Allis (2019) The expanding landscape of 'oncohistone' mutations in human cancers. *Nature* **567**, 473-478.
25. K. A. Thompson, B. Wang, W. S. Argraves, F. G. Giancotti, D. P. Schranck and E. Ruoslahti (1994) BR140, a novel zinc-finger protein with homology to the TAF250 subunit of TFIID. *Biochem. Biophys. Res. Commun.* **198**, 1143-1152.
26. X. J. Yang (2015) MOZ and MORF acetyltransferases: molecular interaction, animal development and human disease. *Biochim Biophys Acta* **1853**, 1818-1826.
27. K. Yan, J. Rousseau, K. Machol, L. A. Cross, K. E. Agre, C. F. Gibson, A. Goverde, K. L. Engleman, H. Verdin, E. Baere, L. Potocki, D. Zhou, M. Cadieux-Dion, G. A. Bellus, M. D. Wagner, R. J. Hale, N. Esber, A. F. Riley, B. D. Solomon, M. T. Cho, H. Porter, B. C. Lanpher, A. M. Lewis, J. Savatt, I. Thiffault, B. Callewaert, P. M. Campeau and X. J. Yang (2020)

Deficient histone H3 propionylation by BRPF1-KAT6 complexes in neurodevelopmental disorders and cancer. *Sci. Adv.* **6**, eaax0021.

28. F. Huang, A. Paulson, A. Dutta, S. Venkatesh, M. Smolle, S. M. Abmayr and J. L. Workman (2014) Histone acetyltransferase Enok regulates oocyte polarization by promoting expression of the actin nucleation factor spire. *Genes Dev* **28**, 2750-2763.
29. Y. Doyon, C. Cayrou, M. Ullah, A. J. Landry, V. Cote, W. Selleck, W. S. Lane, S. Tan, X. J. Yang and J. Cote (2006) ING tumor suppressors are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol. Cell* **21**, 51-64.
30. B. J. Klein, K. L. Cox, S. M. Jang, J. Cote, M. G. Poirier and T. G. Kutateladze (2020) Molecular Basis for the PZP Domain of BRPF1 Association with Chromatin. *Structure* **28**, 105-110 e103.
31. K. Yan, J. Rousseau, R. O. Littlejohn, C. Kiss, A. Lehman, J. R. Rosenfeld, C. T. R. Stumpel, A. P. Stegmann, L. Robak, F. Scaglia, T. T. Nguyen, H. Fu, N. F. Ajeawung, M. V. Camurri, L. Li, A. Gardham, B. Panis, M. Almannai, M. J. Sacoto, B. Baskin, C. Ruivenkamp, D. Study, C. Study, M. T. Cho, T. Potjer, G. W. Santen, M. J. Parker, N. Canham, M. McKinnon, L. Potocki, J. MacKenzie, E. R. Roeder, P. M. Campeau and X. J. Yang (2017) Mutations in the chromatin regulator gene BRPF1 causes syndromic intellectual disability and deficient histone acetylation. *Am. J. Hum. Genet.* **100**, 91-104.
32. F. Mattioli, E. Schaefer, A. Magee, P. Mark, G. M. Mancini, K. Dieterich, G. Von Allmen, M. Alders, C. Coutton, M. van Slegtenhorst, G. Vieville, M. Engelen, J. M. Cobben, J. Juusola, A. Pujol, J. L. Mandel and A. Piton (2017) Mutations in Histone Acetylase Modifier BRPF1 Cause an Autosomal-Dominant Form of Intellectual Disability with Associated Ptosis. *Am J Hum Genet* **100**, 105-116.
33. S. Baker, J. Murrell, A. Nesbitt, K. Pechter, J. Balciuniene, X. Zhao, Z. Yu, E. Denenberg, E. DeChene, A. Wilkens, E. Bhoj, Q. Guan, M. Dulik, L. Conlin, A. Abou Tayoun, M. Luo, C. Wu, K. Cao, M. Sarmady, E. Bedoukian, J. Tarpinian, L. Medne, C. Skraban, M. Deardorff, I. Krantz, B. Krock and A. Santani (2019) Automated clinical exome reanalysis reveals novel diagnoses. *J Mol Diagn* **21**, 38-48.
34. N. Pode-Shakked, O. Barel, B. Pode-Shakked, A. Eliyahu, A. Singer, O. Nayshool, N. Kol, A. Raas-Rothschild, E. Pras and M. Shohat (2019) BRPF1-associated intellectual disability, ptosis, and facial dysmorphism in a multiplex family. *Mol Genet Genomic Med* 10.1002/mgg3.665, e665.
35. Y. Mishima, S. Miyagi, A. Saraya, M. Negishi, M. Endoh, T. A. Endo, T. Toyoda, J. Shinga, T. Katsumoto, T. Chiba, N. Yamaguchi, I. Kitabayashi, H. Koseki and A. Iwama (2011) The Hbo1-Brd1/Brpf2 complex is responsible for global acetylation of H3K14 and required for fetal liver erythropoiesis. *Blood* **118**, 2443-2453.
36. Y. Feng, A. Vlassis, C. Roques, M. E. Lalonde, C. Gonzalez-Aguilera, J. P. Lambert, S. B. Lee, X. Zhao, C. Alabert, J. V. Johansen, E. Paquet, X. J. Yang, A. C. Gingras, J. Cote and A. Groth (2016) BRPF3-HBO1 regulates replication origin activation and histone H3K14 acetylation. *EMBO J* **35**, 176-192.
37. K. Yan, L. You, C. Degerny, M. Ghorbani, X. Liu, L. Chen, L. Li, D. Miao and X. J. Yang (2016) The chromatin regulator BRPF3 preferentially activates the HBO1 acetyltransferase but is dispensable for mouse development and survival. *J Biol Chem* **291**, 2647-2663.

38. Y. Tao, C. Zhong, J. Zhu, S. Xu and J. Ding (2017) Structural and mechanistic insights into regulation of HBO1 histone acetyltransferase activity by BRPF2. *Nucleic Acids Res* **45**, 5707-5719.
39. X. J. Yang (2004) The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic Acids Res.* **32**, 959-976.
40. C. D. Allis, S. L. Berger, J. Cote, S. Dent, T. Jenuwien, T. Kouzarides, L. Pillus, D. Reinberg, Y. Shi, R. Shiekhattar, A. Shilatifard, J. Workman and Y. Zhang (2007) New nomenclature for chromatin-modifying enzymes. *Cell* **131**, 633-636.
41. Y. Yan, N. A. Barlev, R. H. Haley, S. L. Berger and R. Marmorstein (2000) Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases. *Mol. Cell* **6**, 1195-1205.
42. L. Li, M. Ghorbani, M. W. Hubshman, J. Rousseau, I. Thiffault, R. E. Schnur, C. Breen, R. Oegema, J. M. Weiss, Q. Waisfisz, S. Welner, H. Kingston, E. M. Boon, L. Basel-Salmon, O. Konen, H. Goldberg, L. Bazaq, S. Tzur, X. Bi, M. Bruccoleri, K. McWalter, M. Cho, M. Scarano, S. S. Brooks, S. S. Hughes, K. L. I. van Gassen, J. M. van Hagen, T. K. Pandita, P. B. Agrawal, P. M. Campeau and X. J. Yang (2020) Lysine acetyltransferase 8 is involved in cerebral development and syndromic intellectual disability. *J. Clin. Investig.* **130**, 1431-1445.
43. S. Yang, Y. Yu, Y. Xu, F. Jian, W. Song, A. Yisimayi, P. Wang, J. Wang, J. Liu, L. Yu, X. Niu, J. Wang, Y. Wang, F. Shao, R. Jin, Y. Wang and Y. Cao (2024) Fast evolution of SARS-CoV-2 BA.2.86 to JN.1 under heavy immune pressure. *Lancet Infect Dis* **24**, e70-e72.
44. D. Agudelo, A. Duringer, L. Bozoyan, C. C. Huard, S. Carter, J. Loehr, D. Synodinou, M. Drouin, J. Salsman, G. Dellaire, J. Laganiere and Y. Doyon (2017) Marker-free coselection for CRISPR-driven genome editing in human cells. *Nat Methods* **14**, 615-620.
45. O. R. Collaboration (2016) The ORFeome Collaboration: a genome-scale human ORF-clone resource. *Nat Methods* **13**, 191-192.
46. M. Dalvai, J. Loehr, K. Jacquet, C. C. Huard, C. Roques, P. Herst, J. Cote and Y. Doyon (2015) A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. *Cell Rep* **13**, 621-633.
47. K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260.
48. M. Ullah, N. Pelletier, L. Xiao, S. P. Zhao, K. Wang, C. Degerny, S. Tahmasebi, C. Cayrou, Y. Doyon, S. L. Goh, N. Champagne, J. Cote and X. J. Yang (2008) Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes. *Mol Cell Biol* **28**, 6828-6843.
49. I. Kitabayashi, Y. Aikawa, L. A. Nguyen, A. Yokoyama and M. Ohki (2001) Activation of AML1-mediated transcription by MOZ and inhibition by the MOZ-CBP fusion protein. *EMBO J.* **20**, 7184-7196.
50. N. Champagne, N. R. Bertos, N. Pelletier, A. H. Wang, M. Vezmar, Y. Yang, H. H. Heng and X. J. Yang (1999) Identification of a human histone acetyltransferase related to monocytic leukemia zinc finger protein. *J. Biol. Chem.* **274**, 28528-28536.

Figure Legends

Fig. 1 Primer design for P3a mutagenesis and efficient generation of histone H3 mutants

A. Completely complementary primer pair for the QuickChange™ mutagenesis method [14,15]. Two red asterisks mark the mutation sites within the primer pair. The primers are 20-25 nucleotides (nt) in length and completely complementary to each other.

B. Partially complementary primer pair with 3′-overhangs. Two red asterisks mark the mutation sites within the primer pair. Liu & Naismith utilized primers with the average length of ~45 nucleotides (ranging from 35 to 55 nucleotides) in length, with special instructions on primer design [20]. We have simplified the primer design and reduced the length of the two primers to ~30 nucleotides, with 9-bp complementary regions flanking a single-nucleotide mutation site [17]. Shorter primers reduce cost and also minimize unwanted mutations introduced by the impurity of primers resulting from their chemical synthesis [17,18].

C. The two primers for engineering the mutation can carry a mismatch at the same position or two slightly different positions as indicated with red and green asterisks. In these cases, the primers are slightly longer than 30 nucleotides, with 9-10 bp complementary regions flanking a single-nucleotide mutation site. With such design, two different mutants can be generated with one pair of primers [17].

D. A pair of 'handshaking' primers to introduce deletion (denoted with two red triangles) via a pair of primers possessing two arms complementary to the flanking regions. The deletion can be from a few base pairs to several kb. Typically, this is impossible or highly inefficient for classical site-directed mutagenesis methods. Conceptually, the primer pairs in strategies depicted in panels D-E form cassettes for replacing the corresponding regions in the wild-type backbone.

E. A pair of 'handshaking' primers to introduce insertion (denoted with red solid lines) through a pair of primers containing two arms complementary to the regions (indicated by solid dark lines) flanking an insertion site. This is impossible or typically inefficient for classical site-directed mutagenesis methods.

F. Efficiency for generating different histone H3 mutants. The notation "K27R;K27M" refers to generating two mutants K27R and K27M, whereas the term "1;4/6" denotes, respectively, one and four K27R and K27M mutants out of 6 colonies sequenced. See Fig. S1 for details on P3a mutagenesis and about how the primers are designed.

G. Sequence chromatograms or sequences of 10 representative plasmids analyzed for engineering histone H3 mutants. The parental plasmid encodes *Xenopus* histone H3. During sequencing, we noticed that the plasmid used for mutagenesis contains silent mutations compared to the sequence that we used to design the primers, which is expected to result in more mismatches between the primers and template. Amazingly, despite these mismatches, the primers successfully introduced the designed mutations and corrected some of the silent mutations.

Fig. 2 Engineering missense or nonsense epigenetic regulator mutants by P3a site-specific mutagenesis

A. Efficiency in engineering eleven BRPF1 mutants. The notation 'Y543A;W544A' refers to generating the two mutants Y543A and W544A with just one pair of primers (see Fig. 1C), whereas the term 'K390A/R392A' denotes the double mutant containing both substitutions.

B. Efficiency in engineering eight BRPF2 mutants. Among them, G38R is to investigate a polymorphic substitution, whereas H275P is to repair an unexpected substitution in an

expression plasmid that we have in the lab. The remaining mutants are located at a KAT7-binding site mapped by structural analysis [38].

C. A representative sequence chromatogram and the sequences of 15 plasmids analyzed for engineering BRPF2 mutations. The plasmids were subject to Sanger sequencing and the resulting data were analyzed via SnapGene.

D. Efficiency in generating eight KAT7 mutants. The first two alter a zinc finger, whereas the remaining 6 mutants are located at a BRPF2-binding site mapped by crystal structural analysis [38].

E. Efficiency in engineering fourteen p300 mutants. The notation 'C1204S;C1204R' refers to generating two mutants Y543A and W544A with just one pair of primers (see Fig. 1C). The asterisks in S1726*, R1830* and Q1845* denote stop codons. See Fig. S2 for mutagenesis efficiency for four other epigenetic regulators.

Fig. 3 Rapid construction of SARS-CoV-2 spike mutants by P3a site-directed mutagenesis

A. Efficiency for generating seven different spike mutants. The first three were generated with the coding sequence for the D614G spike protein, whereas the last four were generated with the coding sequence for the Omicron spike protein. The asterisk indicates that among the 7 colonies analyzed, the L455S/F456L mutant plasmid is predominant. We sequenced plasmids from 8 more colonies; among them, five are for the L455S/F456L mutant plasmid, one carries the double mutation but with deletion of an A just upstream from the mutation, one is wild-type and the 8th could not be sequenced. We reasoned that primer imbalance might have caused the problem (due to different quality), so we adjusted their ratio so that the concentration of the primer for the F456L mutation is four times of that for the L455S/F456L double mutation. The

mutagenesis experiment was repeated under the adjusted primer condition and 4 of the resulting colonies were sequenced. One of them carries the F456L mutant plasmid, another is for L455S/F456L and the remaining two are wild-type. Notably, perhaps due to 3′->5′ exonuclease activity of the polymerase used, mutagenesis was also successful even though one primer for the R436R substitution has a 2-bp mismatch at the 3′-end with the expression plasmid for the Omicron spike protein.

B. Sequence chromatograms of representative plasmids sequenced for engineering SARS-CoV-2 spike mutants. The top chromatogram is the wild-type as the control and the remaining four are candidates from P3a site-directed mutagenesis. The sequencing results show that three are the L455S/F456L mutant and one contains a mixture of the wild-type and this mutant.

Fig. 4 Seamless deletion of small and large DNA fragments by P3a cassette mutagenesis

A. Efficiency of generating three BRPF1 deletion mutants and five other truncation or internal deletion mutants. The primer design for the other mutants is similar to what is illustrated for the dN131 mutant of BRPF1 in panels B-C.

B. Sequence of the region encoding the N-terminal 131 residues of human BRPF1. The 393-bp DNA fragment to be deleted is highlighted in blue.

C. Sequence of the deletion mutant and the two partially complementary primers (dN131-F and dN131-R) that were used for P3a cassette mutagenesis. The notation 'dN131' refers to the deletion of the N-terminal 131 residues.

Fig. 5 Seamless epitope tagging and untagging via P3a cassette mutagenesis

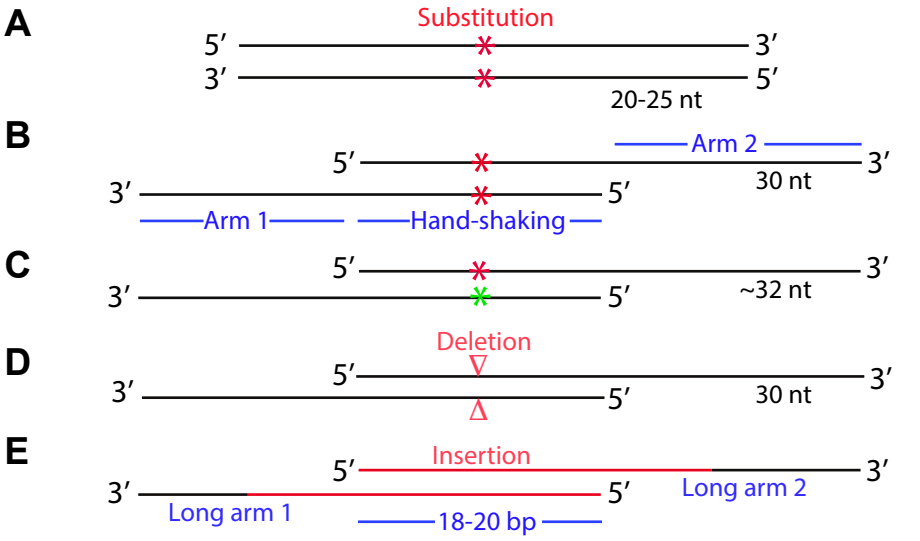
A. Efficiency of epitope tagging into mammalian expression vectors for 8 epigenetic regulators and epitope removal for two expression vectors (deletion of an HA tag from the CBP expression plasmid encoding both FLAG and HA tags, and removal of a 6xHis tag from the p300 expression vector).

B. DNA sequence encoding the N-terminal region of human p300. A 92-bp fragment to be deleted during P3a mutagenesis for introducing the coding sequence for a FLAG tag is highlighted in blue.

C. Sequence of the resulting mutant and the two partially complementary primers (p300-F and p300-R) that were used for P3a site-directed mutagenesis. The two primers were designed to delete the 92-bp fragment (highlighted in panel B) and insert the coding sequence for the FLAG tag.

D. Sanger sequence chromatogram of one mutant plasmid analyzed for the FLAG-coding region (top). Sequences of two other mutant candidates are at the bottom two lines. The results indicate that all three plasmids carry the FLAG coding sequence properly inserted.

Fig. 1



F

Mutation (histone H3)	Mutants/colonies sequenced (%)
K9R	3/3 (100%)
K14R	3/3 (100%)
K18R	3/3 (100%)
K23R	2/3 (66.7%)
K27R; K27M	1;4/6 (83.3%)
G34A; G34R	1;1/6 (33.3%)
K36R; K36M	2;3/6 (83.3%)
Total	23/30 (76.7%)

G

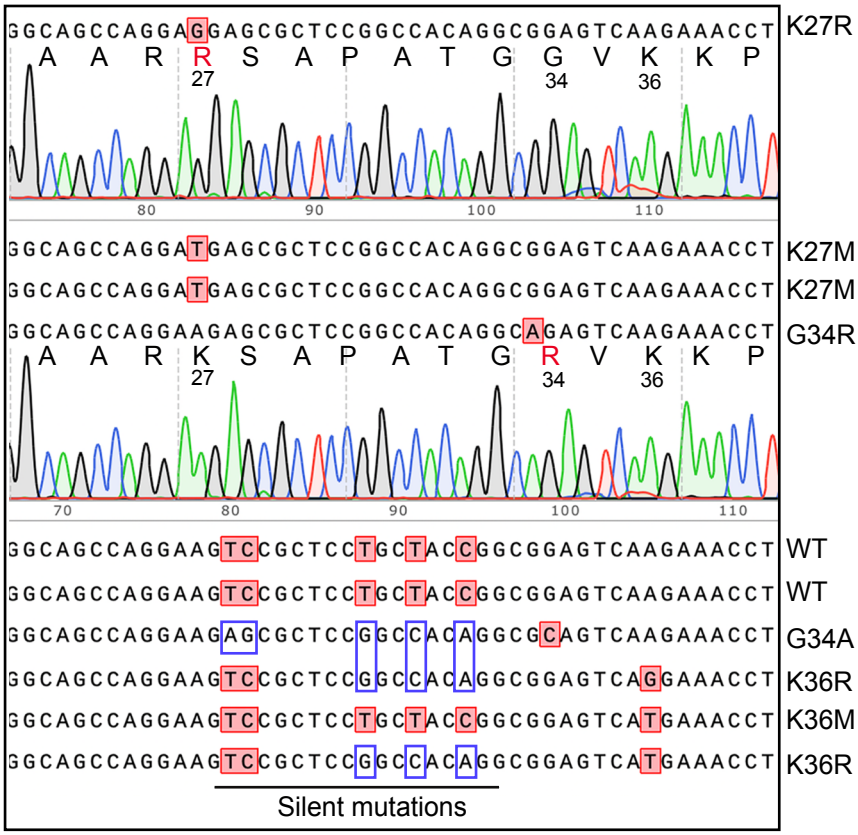


Fig. 2

A

Mutation (BRPF1)	Mutants/colonies sequenced (%)
Y215A	3/3 (100%)
W224A	3/3 (66.7%)
F245A	3/3 (100%)
P334A	3/3 (100%)
R381A	1/3 (33.3%)
K383A	3/3 (100%)
K390A	2/3 (66.7%)
R392A	3/3 (100%)
K390A/R392A	3/3 (100%)
Y543A;W544A	3;1/4 (100%)
Total	28/31 (90.3%)

B

Mutation (BRPF2)	Mutants/colonies sequenced (%)
G38R	2/3 (66.7%)
I53A	3/3 (66.7%)
I57A	3/3 (100%)
I59A	3/3 (33.3%)
I60A	3/3 (100%)
I61A	2/3 (66.7%)
H275P	3;1/4 (100%)
P275A	3/3 (66.7%)
Total	23/25 (92.0%)

D

Mutation (KAT7)	Mutants/colonies sequenced (%)
C185S	3/3 (100%)
C209S	3/3 (66.7%)
K432R	2/3 (100%)
K537A	3/3 (100%)
I539A	3/3 (100%)
I541A	3/3 (100%)
V574A	3/3 (100%)
K576A	3/3 (100%)
Total	23/24 (95.8%)

E

Mutation (p300)	Mutants/colonies sequenced (%)
C1204S;C1204R	2;1/5 (60.0%)
E1242A;E1242K	2;2/4 (100%)
D1384N;D1384H	2;0/3 (66.7%)
D1399N;D1399Y	4;5.5/11 (86.3%)
Y1414C	3.5/3 (100%)
W1466C;Y1467N	2;0/3 (66.7%)
S1726*	2/3 (66.7%)
R1830*	3/3 (100%)
Q1845*	3/3 (100%)
Total	32/38 (84.2%)

C

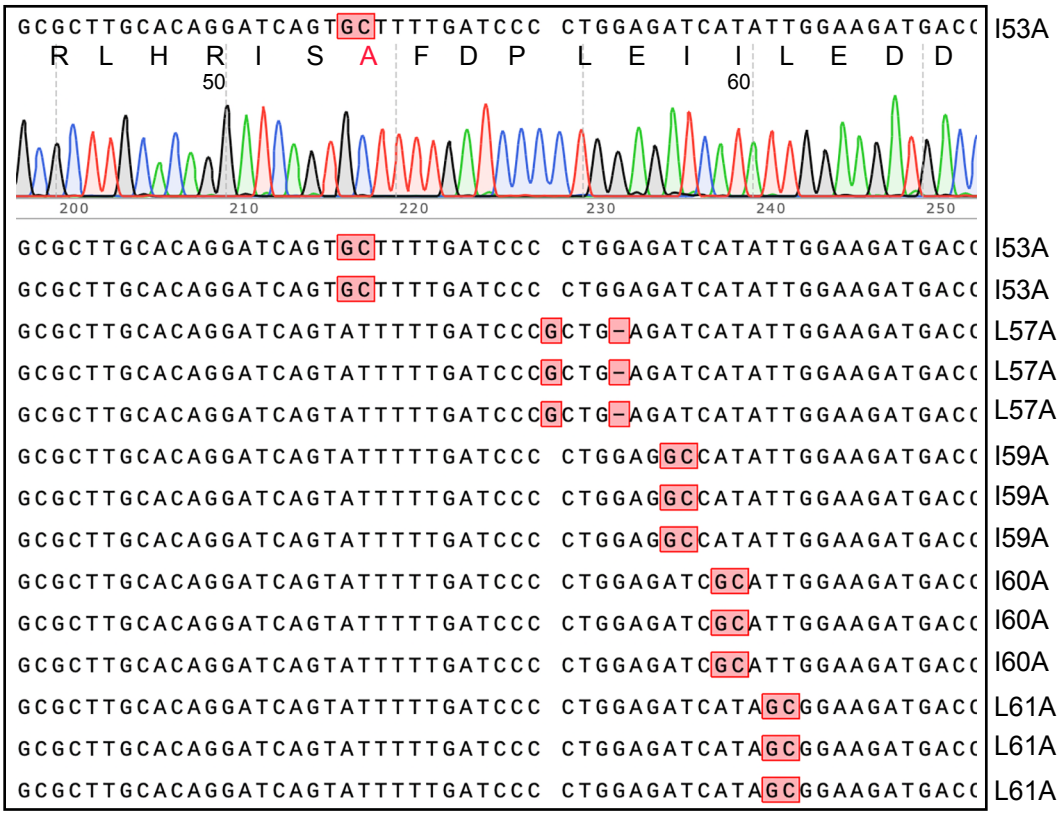


Fig. 3

A

Mutation (spike)	Mutants/colonies sequenced (%)
R346T	3/3 (100%)
Q493E	4/6 (66.7%)
L981F	6/6 (100%)
R346T (JN.1)	4/4 (100%)
L455S/F456L (JN.1)	6.5/7 (92.8%)*
Q493E (JN.1)	3/3 (100%)
V1104L (JN.1)	3/3 (100%)
Total	29.5/32 (92.2%)

B

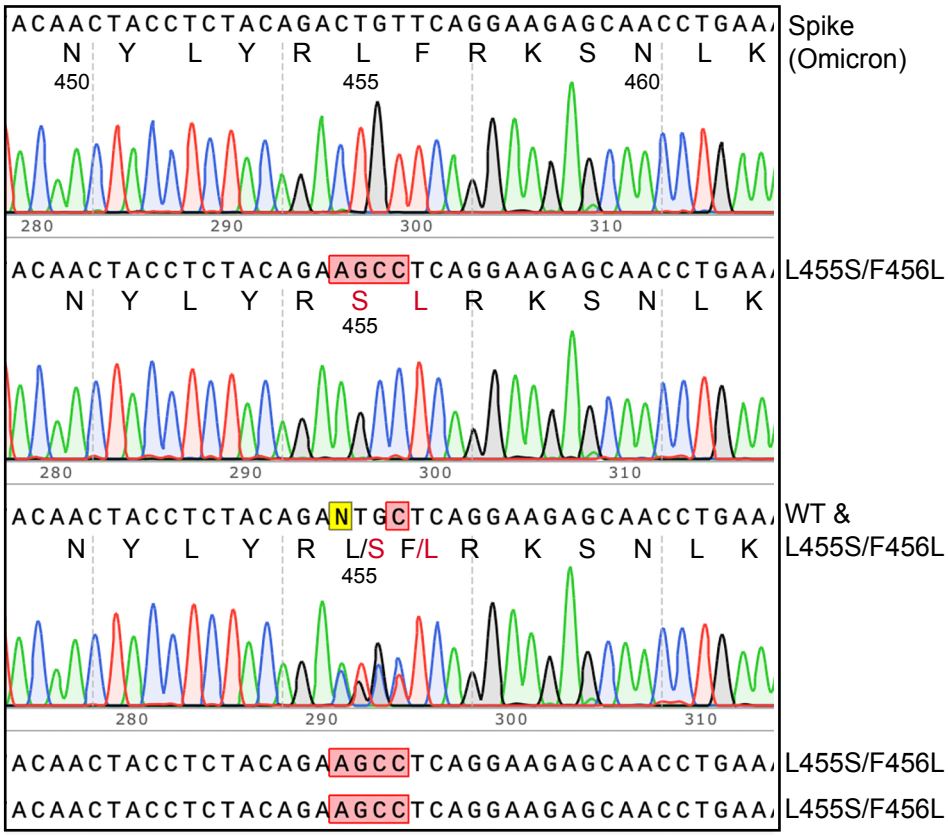
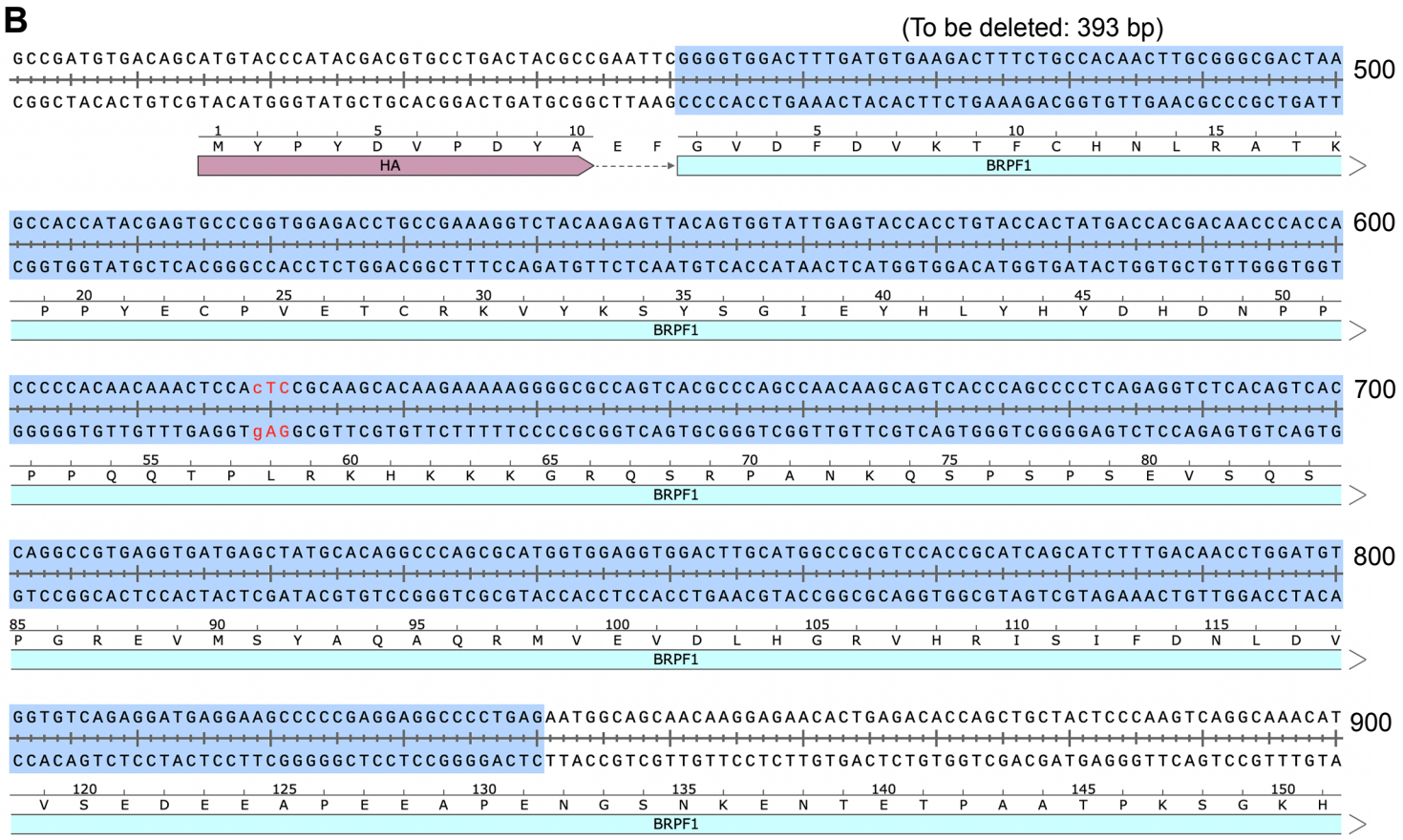


Fig. 4

A

Deletion	Mutants/colonies sequenced (%)
BRPF1-dN131	3/3 (100%)
BRPF1-dN204	3/3 (100%)
BRPF1-d594-604	2/3 (66.7%)
BRPF2-d31-80	3/3 (100%)
JADE2-d65-85	3/6 (50.0%)
p300-dN1301	3/3 (100%)
CBP-dN1068	3/3 (100%)
Total	20/24 (76.9%)

B



C

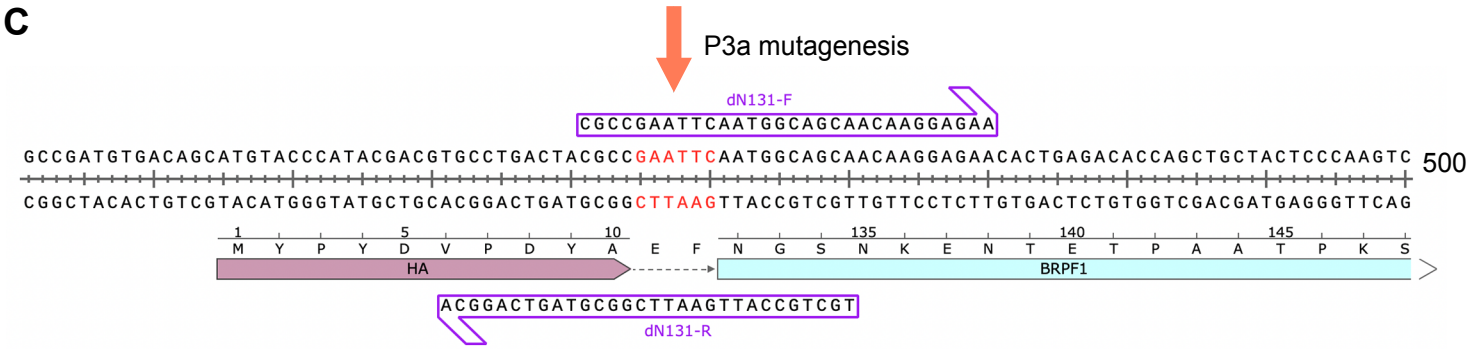
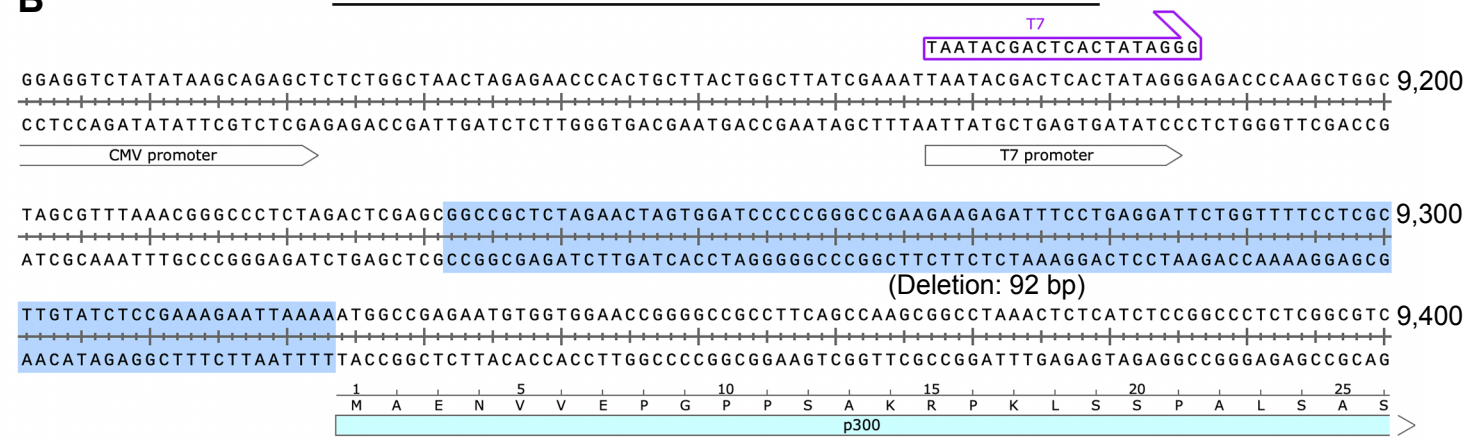


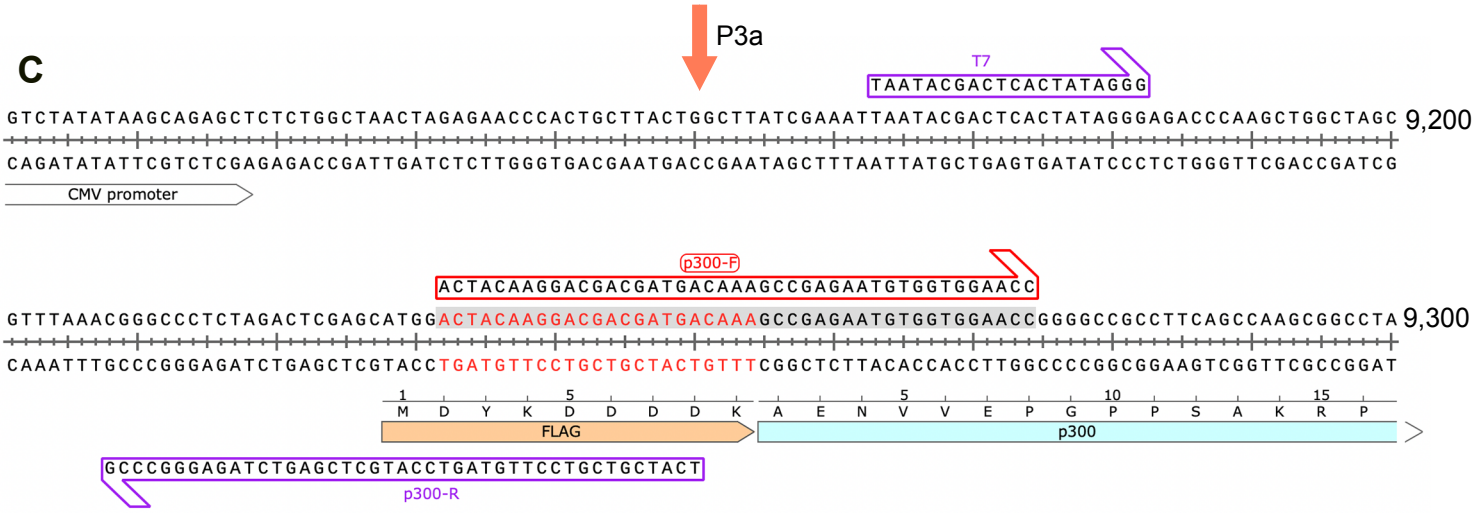
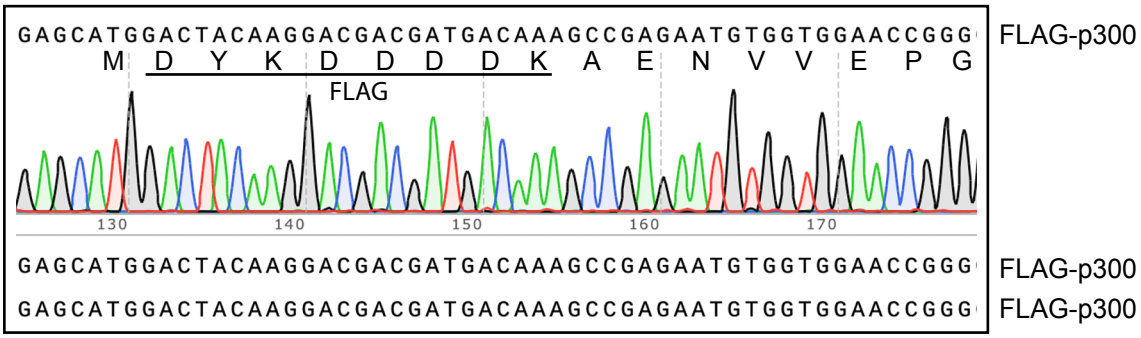
Fig. 5

A

Tagging/untagging	Mutants/colonies sequenced (%)
CBP (HA deletion)	1/3 (33.3%)
p300 (FLAG)	3/3 (100%)
p300 (6His removal)	1/3 (33.3%)
SIN3A (HA)	3/3 (100%)
YEAST2 (HA)	2/3 (100%)
ZZZ3 (HA)	2/3 (66.7%)
PHF6 (HA)	3/3 (100%)
MIER1 (HA)	3/3 (100%)
PHF12 (HA)	2/3 (66.7%)
FLAG-->3xFLAG*	5/6 (83.3%)
Total	25/33 (75.6%)

B





D