bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Generation of highly diverse peptide library by linear-double-stranded DNA based AND gate genetic circuit in mammalian cells

Shuai Li [1,2#*], Weijun Su [3#], Chunze Zhang [4]

1. Department of Breast Cancer Pathology and Research Laboratory, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer; Key Laboratory of Cancer Prevention and Therapy, Tianjin; Tianjin's Clinical Research Center for Cancer, 300060, China;

2. State Key Laboratory of Medicinal Chemical Biology, Nankai University, Tianjin 300071, China;

3. School of Medicine, Nankai University, Tianjin 300071, China;

4. Department of Colorectal Surgery, Tianjin Union Medical Center, Tianjin 300121, China.

# Shuai Li and Weijun Su contributed equally to this work.

* To whom correspondence should be addressed. Tel: +86 22 23340123; Fax: +86 22 23340123;

Email: shuaili@tmu.edu.cn & shuai.li2001@gmail.com.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

DNA-encoded peptide libraries are ideal functional peptide discovery platforms for their extremely large capacity. However, it's still difficult to build high content peptide library in intact mammalian cells, which offer advantages associated with appropriate protein modification, proper protein folding, and natural status of membrane protein. Our previous work established linear-double-stranded DNAs (ldsDNAs) as innovative biological parts to implement AND gate genetic circuits in mammalian cell line. In the current study, we employ ldsDNA with terminal NNK degenerate codons as AND gate input to build highly diverse peptide library in mammalian cells. This **l**dsDNA-**b**ased **A**ND **g**ate (LBAG) peptide strategy is easy to conduct, only PCR reaction and cell transfection experiments are needed. High-throughput sequencing (HTS) results reveal that our new LBAG strategy could generate peptide library with both amino acid sequence and peptide length diversities. Moreover, by a mammalian cell two-hybrid system, we pan an MDM2 protein interacting peptide through the LBAG peptide library. Our work establishes ldsDNA as biological parts for building highly diverse peptide library in mammalian cells.

Keywords

peptide library, linear-double-stranded DNA (ldsDNA), AND gate, degenerate codon.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Different state-of-the-art display technologies, e.g., phage, ribosome, mRNA, bacterial, and yeast-display, have been developed to identify functional peptides (1-7). Through construction and screen vast peptide libraries, these display technologies physically couple between phenotype (high affinity high selectivity binding) and genotype (DNA sequence). Currently, plasmids and PCR products have been explored as genetic mediums to store highly diverse coding information of the peptide library (8,9). These genetic materials are introduced into phage-host bacteria, yeast, or test tubes to transcript-translate into peptides. However, it's still a big challenge to build high content peptide library in intact mammalian cells. Some attempts have been made to introduce DNA-encoded peptide library into mammalian cells using episomal-, viral- or transposon-mediated gene transfer (10-17). While, the peptide library capacity is constantly lost during plasmid extraction, lentivirus generation, transfection/infection process. This leads to small library size relative to other display technologies.

We previously applied linear-double-stranded DNA (ldsDNA, or PCR amplicon) as novel biological parts to implement Boolean logic AND gate genetic circuits in mammalian cells (18). Via splitting essential gene expression cassette into signal-mute ldsDNA, our strategy achieves two or three-input AND gate calculation with a low noise-signal ratio both *in vitro* and *in vivo*. This split-relink process is similar to V(D)J recombination, which serves as the fundamental molecular mechanism to generate highly diverse repertoires of antigen-receptors (antibodies and T-cell receptors) in mammalian (19,20). In the present study, we employ this **l**dsDNA-**b**ased **A**ND-**g**ate (LBAG) genetic circuit to generate high content peptide library.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Generation of peptide library by ldsDNA-based AND-gate genetic circuit

Our previous study demonstrated that ldsDNAs (PCR amplicons), split from intact gene expression cassettes (containing promoter, gene coding sequence and polyA signal), could undergo reconnection to form one output-signal-generating molecule in cultured cells (18). Here we hypothesize that ldsDNAs with terminal NNK degenerate codons could generate highly diverse peptide library through **l**dsDNA-**b**ased **A**ND-**g**ate (LBAG) calculation in mammalian cells (Figure 1).

Taking pEGFP-C1 plasmid as PCR template, primers with multiple tandem NNK-trinucleotides were used to generate ldsDNAs with terminal degenerate codons (Supplementary Table 1). Three kinds of ldsDNAs, containing CMV promoter, Kozak sequence and different number of NNK-trinucleotides, were produced (namely CMV-Kozak-$<NNK>_5$, CMV-Kozak-$<NNK>_{10}$, CMV-Kozak-$<NNK>_{15}$) (for sequences see Supplementary Materials). Meanwhile, two kinds of ldsDNAs, containing terminal NNK-trinucleotides (or not), GFP coding sequence and SV40 polyA signal (PAS), were amplified (namely $<NNK>_5$-GFP-SV40 PAS, $<NNK>_0$-GFP-SV40 PAS). Six different combinations of these two subgroup ldsDNAs were introduced into HEK293T cells to construct the AND-gate genetic circuits. The six combinations are as following: 1, CMV-Kozak-$<NNK>_5$ with $<NNK>_5$-GFP-SV40 PAS (5X+5X for short); 2, CMV-Kozak-$<NNK>_{10}$ with $<NNK>_5$-GFP-SV40 PAS (10X+5X); 3, CMV-Kozak-$<NNK>_{15}$ with $<NNK>_5$-GFP-SV40 PAS (15X+5X); 4, CMV-Kozak-$<NNK>_5$ with $<NNK>_0$-GFP-SV40 PAS (5X+0X); 5,

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

CMV-Kozak-<NNK>$_{10}$ with <NNK>$_0$-GFP-SV40 PAS (10X+0X); 6, CMV-Kozak-<NNK>$_{15}$ with <NNK>$_0$-GFP-SV40 PAS (15X+0X).

When AND-gate genetic circuits are built up in cells, peptide library would be generated surrounding the junctions formed between these two subtype ldsDNAs (Figure 1). Forty-eight hours after transfection, total RNAs were extracted then cDNAs were synthesized via reverse transcription. Nucleotide sequences surrounding the junctions formed by two ldsDNA-inputs were PCR amplified and then subjected to pair-end high-throughput sequencing (HTS). The HTS statistics for all sequencing libraries are shown in Table 1.

Reading frame phase of ldsDNA-based peptide library

To display or tether library peptides, oligopeptides are usually fused to specific anchor proteins. For example in phage display experiment, the peptide library is fused to an open reading frame (ORF) thereby connecting peptide to phage coat protein (1). Our previous research demonstrated that the ldsDNA could be subject to terminal nucleotide(s) deletion which may lead to reading frame shift. Here we analyze the reading frame phase characteristics of the LBAG peptide libraries (Figure 2). Libraries generated by dual-side (_X+_X) terminal-NNKs-ldsDNAs, e.g., CMV-Kozak-<NNK>$_5$ with <NNK>$_5$-GFP-SV40 PAS (5X+5X), show a similar reading frame phase distribution. The average in-frame (3n) ratio ranging from 45% to 47%. While libraries generated by single-side (_X+0X) terminal-NNKs-ldsDNAs, e.g., CMV-Kozak-<NNK>$_5$ with <NNK>$_0$-GFP-SV40 PAS (5X+0X), show dramatic dynamics of reading frame phase distribution (Figure 2). 5X+0X peptide libraries reach the maximum 69% in-frame rate compared with 63% of 10X+0X and 53% of 15X+0X,

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

respectively. Then, the in-frame nucleotide sequences (3n) are translated into polypeptides for further analysis.

Length dynamic and amino acid composition of ldsDNA-based peptide library

We next analyze the polypeptide length distributions of different LBAG peptide libraries (peptides with pre-stop codon are excluded). Dual-side terminal-NNKs-ldsDNAs strategies, including 5X+5X, 10X+5X and 15X+5X, show a more dynamic length distribution compared with that from single-side terminal-NNKs-ldsDNAs strategies (5X+0X, 10X+0X and 15X+0X) (Figure 3). In addition, for single-side terminal-NNKs-ldsDNAs strategies, most peptides have a length equal to the theoretical polypeptide length generated by direct ligation (supposing no insertion or deletion happened) (Figure 3D-F). For example, in the 5X+0X group, over 81% of the peptide length is five amino acids.

The amino acid composition of different LBAG peptide libraries is in line with the theoretical NNK amino acid composition (pre-stop codon is included in the analysis) (Table 2). However, Proline (P) residue always shows a higher percentage than the theoretical ratio, particularly when the peptide chain getting longer. Meanwhile, the percentage of each nucleotide at each position of peptide library codons is showed in Supplementary Table 2. The nucleotide distributions are in line with NNK degenerate base ratio.

The capacity of the LBAG peptide library

Peptide library capacity could be represented by its coverage ability on amino acid combination possibilities. Here, all eighteen LBAG peptide libraries cover all 400 possible dual-amino-acid combinations (2-mer). For triple-amino-acid (3-mer), the library coverage is ranging from 64% to 98% (Figure 4A). For quadruple-amino-acid (4-mer), the library coverage is ranging from 5.7% to 32.1%

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Figure 4B). The peptide library content, defined as the total amino acid number coded by a library, serves as a determinate factor of library capacity (Figure 4).

The asymmetry of input ldsDNA end processing during AND gate calculation

As shown by Figure 2 and 3, single-side (_X+0X) and dual-side (_X+_X) LBAG peptide library strategies show obvious differences in both reading frame phase distributions and peptide length dynamics. To further investigate these differences and figure out whether or not the up-stream and down-stream ldsDNAs have the same chance of end processing, we PCR-generated two pairs of AND gate ldsDNAs: pair 1, CMV-Kozak-FLAG-NNN with GFP-SV40 PAS; and pair 2, CMV-Kozak-FLAG with NNN-GFP-SV40 PAS (N for randomized nucleotides, for ldsDNA sequences see Supplementary Materials). After cell transfection, RNA extraction, cDNA synthesis, and PCR amplification, high-throughput sequencing (HTS) was performed to characterize the oligonucleotide sequences formed by AND gate calculation. About 67% of the sequences formed by CMV-Kozak-FLAG-NNN and GFP-SV40 PAS contains intact random nucleotides (NNN) (Figure 5). While for CMV-Kozak-FLAG and NNN-GFP-SV40 PAS group, about 57% of oligonucleotide contains triple nucleotides (Figure 5). Thus, the down-stream ldsDNAs (NNN-GFP-SV40 PAS) are more likely subject to end processing than up-stream ldsDNAs (CMV-Kozak-FLAG-NNN) in AND gate relink process in cells. This may be the reason why single-side (_X+0X) strategy generate higher in-frame rate and less variability among the peptide length than dual-side (_X+_X) strategy.

In addition, through the above terminal-triple-random-nucleotide-ldsDNA AND gate calculation, we also evaluated the nucleotide preferences of LBAG calculation. In both CMV-Kozak-FLAG-NNN + GFP-SV40 PAS group and CMV-Kozak-FLAG + NNN-GFP-SV40 PAS group, Guanine (G) is

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

always the unfavorable end nucleotide (Supplementary Table 3). Furthermore, up-stream or down-stream ldsDNAs containing two or three terminal Guanines are hard to connect to the corresponding ldsDNAs (Supplementary Table 3).

Using LBAG peptide library to identify MDM2 protein interacting peptide

The E3 ubiquitin ligase MDM2 negatively regulates the activity of the tumor suppressor protein p53 through interacting with the N-terminal transactivation domain of p53. Structurally, the residues F19, W23 and L26 in p53, which insert deep into the MDM2 hydrophobic cleft, are critical in MDM2-p53 interaction (21). Also, the non-peptide inhibitors of the p53-MDM2 interaction have to contain mimics of these amino acids (22-24). Pazgier *et al.* screened a duodecimal peptide phage display library to identify MDM2 (residues 25-109) interacting peptides (25). They found 14 out of 15 MDM2 binding clones containing the same conserved residues: **F**XXX**W**XX**L**, which in line with the MDM2-p53 binding pattern. Shiheido *et al.* employed mRNA display to identify MDM2 interacting peptides (26,27). Similarly, they found more than half of all peptides retained the three hydrophobic residues corresponding to F19, W23 and L26 of wild-type p53. Here, to test the performance of our LBAG peptide library strategy, we plan to identify the MDM2 interacting peptides in mammalian cells.

To detect protein-protein interaction in mammalian cells, we employed the two-hybrid system, which has been adapted for use in mammalian cells (28,29). The system contains three plasmids: pACT, which expresses VP16 activation domain, pBIND, which expresses GAL4 DNA-binding domain, pG5egfp, which contains five GAL4 binding sites upstream of a minimal TATA box and EGFP ORF. In the present study, MDM2 (residues 25-109) is subcloned into pACT plasmid (namely pACT-MDM2). By PCR amplification, the LBAG peptide library is fused to GAL4 DNA-binding

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

domain through an XTEN linker (Figure 6A). To achieve high library content, we here used single-side LBAG strategy.

Two plasmids (pACT-MDM2, pG5egfp) and two ldsDNAs (CMV-GAL4-XTEN-<NNK>$_{15}$, SV40 PAS) were cotransfected into HEK293T cells. The peptide library will be formed through the AND gate calculation between CMV-GAL4-XTEN-<NNK>$_{15}$ and SV40 PAS (Figure 6A). If the peptide interacts with MDM2, the activation domain VP16 will be recruited to GAL4 binding sites, thereby activating the GFP reporter gene expression (Figure 6B). Here fluorescence-activated cell sorting (FACS) was used to isolate the GFP positive cells. High-throughput sequencing of the peptide library in these cells identified peptide S**F**VIR**W**TA**L**WPGPRS among the top hits (23 reads, rank fourth) (Supplementary Table 4). This peptide contains the conserved residues **F**XXX**W**XX**L**, which is critical to interact with MDM2.

Discussion

Affinity-based display technologies, e.g., phage, ribosome, mRNA, bacterial, and yeast-display, have been developed to screen peptide libraries to identify protein-binding peptides. Peptide library size is a determinate factor of a success screen. For phage and yeast-display, the complexity of peptide library stores in plasmid (1,2,7). Currently, the phage display library size could reach 10^9 (9,30). While for ribosome and mRNA-display, PCR products serve as genetic medium to store library coding information. Ribosome and mRNA-display could achieve extremely large library size (>10^{13}) (5,31). In the current research, we provide a new strategy to build high content peptide library in cultured cell line. The protocol is easy to conduct, only containing two steps: PCR amplification and cell

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

transfection. Since degenerate trinucleotides is present at the 5' end of the primer, the library complexity keeps on increasing during PCR amplification process. Theoretically, no specific nucleotide sequence is enriched during PCR reaction. PCR products are then transfected into cells to conduct AND-gate linkage to generate full gene expression cassette (promoter-CDS-PAS). For our strategy, no particular attention needs to be paid to maintain the library abundance. The size of ldsDNA-based library needs to be further investigated. Determining the peptide sequence diversity in each cell by single-cell-sequencing will provide crucial information for estimating the whole library size.

For the current protocol, there are still several issues that need to be further addressed to increase library capacity. First of all, the high in-frame ratio would improve library capacity. Our results reveal that peptide libraries generated by single-side NNKs-ldsDNA show higher in-frame ratios than that from dual-side NNKs-ldsDNAs (Figure 2). Besides, chemical modified primers could be used to generate nuclease resistance ldsDNAs to prevent reading frame shift (32,33). Another major obstacle is the pre-stop codon, which causes premature termination (34). For NNK degenerate codon, the percentage of stop-codon in each amino acid position is 3.125% (1/32) and will be accumulated when the peptide is getting longer. When the polypeptide length reaches 20 amino acids, about 47% of the sequences contain pre-stop codon. One potential solution to avoid stop codon is to synthesize PCR primer using trimer phosphoramidites (35,36).

In the present study, we tested the feasibility of our LBAG peptide library strategy through the identification of the MDM2 interacting peptide (Figure 6). Indeed, one peptide (S**F**VIR**W**TA**L**WPGPRS), which contains three crucial residues mediating MDM2-p53 interaction, is among the top hits (Supplementary Table 4). However, in phage display and mRNA display

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

experiments, multiple peptides with **F**XXX**W**XX**L** pattern were identified (25-27). This may come from the enrichment of interacting peptides during rounds of panning in phage and mRNA display process. For the current protocol, four components, including two plasmids (pACT-MDM2, pG5egfp) and two ldsDNAs (CMV-GAL4-XTEN-<NNK>$_{15}$, SV40 PAS), were transient cotransfected into one HEK293T cell to conduct AND gate calculation and assemble the two-hybrid system. In the future application, to save the cell transfection capacity, the plasmids should be preset into cells by generating a stable cell line. Thereby, only ldsDNAs need to be transfected into cells.

Here by terminal-NNKs-**l**dsDNA-**b**ased **A**ND-**g**ate (LBAG) genetic circuit, we develop a novel method to generate highly diverse peptide library in mammalian cells. Our protocol is easy to conduct only PCR-reaction and cell-transfection are needed. High-throughput sequencing results reveal that LBAG peptide library is characterized by both amino acids sequence diversity and peptide length dynamics. We also identified a peptide with conserved residues which are critical to bind to MDM2. Our new method may have great application potential in therapeutic peptide development.

Materials and Methods

Cell culture

The HEK293T cell line was maintained in Dulbecco's modified Eagle's medium (DMEM) (Thermo Fisher Scientific, Hudson, NH, USA) containing 10% fetal bovine serum (FBS) with 1% penicillin-streptomycin solution at 37 degree with 5% CO_2.

ldsDNA synthesis

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

KOD-Plus-Ver.2 DNA polymerase (Toyobo, Osaka, Japan) was used to amplify ldsDNAs (PCR amplicons) using the pEGFP-C1 (Clontech, Mountain View, CA, USA) plasmid as the template. PCR products underwent agarose electrophoresis and gel-purification to remove plasmid template and free dNTPs (Universal DNA purification kit, Tiangen, Beijing, China). The amount of PCR products was determined by the OD 260 absorption value. PCR primer sequences and ldsDNA sequences were present in Supplementary Table 1 and Supplementary Materials, respectively.

ldsDNA transfection

HEK293T cells were seeded into 6-well plates the day before transfection (60% - 70% confluency). 500 ng/well of each input amplicons (total 1000 ng/well) were transfected into cells with Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA) reagent following standard protocol.

RNA extraction, cDNA synthesis and PCR amplification

Forty-eight hours after transfection, total RNAs were extracted by RNAiso Plus (TaKaRa, Beijing, China) reagent. cDNAs were synthesized using the PrimeScript ™ RT reagent Kit with gDNA Eraser (TaKaRa) following standard protocol. KOD-Plus-Ver.2 DNA polymerase (Toyobo) was used to amplify the corresponding cDNA sequences surrounding ldsDNA junctions.

Two-hybrid system in mammalian cells

The two-hybrid system used in this study was developed from commercial CheckMate™ Mammalian Two-Hybrid System (Promega, Madison, WI, USA) with some modifications. Briefly, MDM2 (residues 25-109) was subcloned into pACT (pACT empty vector, GenBank Accession Number AF264723) vector through XbaI and KpnI restriction sites (namely pACT-MDM2). EGFP sequence was cloned into pG5luc vector (pG5luc empty vector, GenBank Accession Number AF264724) to

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

replace luciferase sequence through HindIII and FseI restriction sites (namely pG5egfp). Two ldsDNAs CMV-GAL4-XTEN-<NNK>$_{15}$ and SV40 PAS were PCR-generated using the pBIND (pBIND empty vector, GenBank Accession Number AF264722) plasmid as the template. pACT-MDM2 (500ng/well), pG5egfp (500ng/well), ldsDNA CMV-GAL4-XTEN-<NNK>$_{15}$ (500ng/well) and ldsDNA SV40 PAS (250ng/well) were transient cotransfected into 6-well plated HEK293T cells using Lipofectamine 2000. The experiment could be scaled up by transfected more 6-well plated cells. Forty-eight hours after transfection, FACS was used to isolate the GFP positive cells.

Library construction and high-throughput sequencing

Library construction and high-throughput sequencing were conduct by Novogene (Beijing, China). Sequencing libraries were generated using TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, San Diego, CA, USA) following the manufacturer's recommendations and index codes were added. The library quality was assessed on the Qubit Fluorometer and Agilent Bioanalyzer 2100 system. At last, the libraries were sequenced on an Illumina HiSeq 2500 platform and 250 bp paired-end reads were generated.

Data analysis

Paired-end reads from the original DNA fragments were merged using FLASH, a very fast and accurate analysis tool which was designed to merge paired-end reads when there are overlaps between reads 1 and reads 2. Paired-end reads were assigned to each sample according to the unique barcodes. Reads were filtered by QIIME quality filters. Nucleotide sequences generated by ldsDNA-based AND-gate genetic circuits were extracted by in-house Perl scripts. High-throughput sequencing data

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

are deposited on the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) under the accession number GEO: GSE134671 (reviewer access token: kpujmequhnspngx).

Acknowledgements

This work was supported by the National Natural Science Foundation of China [31870860, 31400673 to S.L., 31971388, 81402407 to W.S.]; the Tianjin Research Program of Application Foundation and Advanced Technology [14JCQNJC09800 to S.L., 15JCQNJC11700 to W.S.]; the State Key Laboratory of Medicinal Chemical Biology (Nankai University) [2018103 to S.L.]; and the Key Project of National Health and Family Planning Commission of Tianjin [2017057 to C.Z.]. The authors have declared that no competing interest exists.

Author Contributions

S.L. conceived and supervised the study. S.L. and W.S. designed and performed the experiments. S.L. and W.S. analyzed the data. W.S. and S.L. prepared the figures. S.L. and W.S. wrote the paper. All authors discussed the results and reviewed the manuscript.

Competing Interests

The authors have declared that no competing interest exists.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure Legends

Figure 1. Schematic showing the rationale of generating highly diverse peptide library by ldsDNA-based AND-gate (LBAG) calculation. ldsDNAs containing CMV promoter and tandem NNK degenerate codons are cotransfected with ldsDNAs containing NNK codons, GFP coding region, and polyA signal sequence. The AND-gate calculation is conducted between two kinds of ldsDNA to form peptide library.

Figure 2. Reading frame character of nucleotide sequences formed by different libraries. Pair-end high-throughput sequencing data is processed to extract the nucleotide sequences formed by ldsDNA-based AND-gate. The nucleotide sequence length is divided by three to show its reading frame character. The average percentage of different phase of different libraries is presented. All data are displayed as the mean ± SD, n = 3.

Figure 3. The peptide length distributions of six different peptide libraries generated by ldsDNA-based AND-gate genetic circuits. In-frame (3n) nucleotide sequences are translated into polypeptide sequences. Then the length of polypeptide sequence in each library is summarized. The average percentage of different polypeptide length is presented. All data are displayed as the mean ± SD, n = 3.

Figure 4. Peptide library coding capacity on covering triple-amino-acid (A) or quadruple-amino-acid (B) peptide possibilities. Peptide libraries are scanned by all triple or quadruple amino-acid combinations, respectively. The library content (defined as the total amino acid number coded by a library) and the percentage of coverage are presented.

Figure 5. Reading frame character of nucleotide sequences formed by ldsDNAs with terminal-triple-random-nucleotides. Two pairs of AND gate ldsDNAs were PCR-generated and transfected into HEK293T cells (pair 1, CMV-Kozak-FLAG-NNN with GFP-SV40 PAS; pair 2, CMV-Kozak-FLAG with NNN-GFP-SV40 PAS). Pair-end high-throughput sequencing data is processed to extract the nucleotide sequences formed by ldsDNA-based AND-gate. The average percentage of different phase of different libraries is presented. All data are displayed as the mean ± SD, n = 4.

Figure 6. Schematic showing the rationale of generating LBAG peptide library for two-hybrid system in mammalian cells. (A) ldsDNA CMV-GAL4-XTEN-$<NNK>_{15}$ (containing CMV promoter, GAL4-XTEN coding sequence, tandem NNK degenerate codons) is cotransfected with ldsDNA SV40 PAS. The AND-gate calculation is conducted between two ldsDNAs to form peptide library. (B) Schematic representation of the mammalian two-hybrid system. The pG5egfp plasmid contains five GAL4 binding sites upstream of a minimal TATA box, which in turn is upstream of the GFP gene.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Interaction between the two test proteins, GAL4-peptide and MDM2-VP16, results in an expression of GFP protein.

References

1. Smith, G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science, 228, 1315-1317.
2. Scott, J.K. and Smith, G.P. (1990) Searching for peptide ligands with an epitope library. Science, 249, 386-390.
3. Mattheakis, L.C., Bhatt, R.R. and Dower, W.J. (1994) An in vitro polysome display system for identifying ligands from very large peptide libraries. Proceedings of the National Academy of Sciences of the United States of America, 91, 9022-9026.
4. Hanes, J. and Pluckthun, A. (1997) In vitro selection and evolution of functional proteins by using ribosome display. Proceedings of the National Academy of Sciences of the United States of America, 94, 4937-4942.
5. Roberts, R.W. and Szostak, J.W. (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. Proceedings of the National Academy of Sciences of the United States of America, 94, 12297-12302.
6. Georgiou, G., Stathopoulos, C., Daugherty, P.S., Nayak, A.R., Iverson, B.L. and Curtiss, R., 3rd. (1997) Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. Nat Biotechnol, 15, 29-34.
7. Boder, E.T. and Wittrup, K.D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. Nat Biotechnol, 15, 553-557.
8. Henninot, A., Collins, J.C. and Nuss, J.M. (2018) The Current State of Peptide Drug Discovery: Back to the Future? J Med Chem, 61, 1382-1414.
9. Obexer, R., Walport, L.J. and Suga, H. (2017) Exploring sequence space: harnessing chemical and biological diversity towards new peptide leads. Curr Opin Chem Biol, 38, 52-61.
10. Beerli, R.R., Bauer, M., Buser, R.B., Gwerder, M., Muntwiler, S., Maurer, P., Saudan, P. and Bachmann, M.F. (2008) Isolation of human monoclonal antibodies by mammalian cell display. Proceedings of the National Academy of Sciences of the United States of America, 105, 14336-14341.
11. Waldmeier, L., Hellmann, I., Gutknecht, C.K., Wolter, F.I., Cook, S.C., Reddy, S.T., Grawunder, U. and Beerli, R.R. (2016) Transpo-mAb display: Transposition-mediated B cell display and functional screening of full-length IgG antibody libraries. mAbs, 8, 726-740.
12. Bowers, P.M., Horlick, R.A., Neben, T.Y., Toobian, R.M., Tomlinson, G.L., Dalton, J.L., Jones, H.A., Chen, A., Altobell, L., 3rd, Zhang, X. et al. (2011) Coupling mammalian cell surface display with somatic hypermutation for the discovery and maturation of human antibodies. Proceedings of the National Academy of Sciences of the United States of America, 108, 20455-20460.
13. Ho, M., Nagata, S. and Pastan, I. (2006) Isolation of anti-CD22 Fv with high affinity by Fv display on human cells. Proceedings of the National Academy of Sciences of the United States of America, 103, 9637-9642.
14. Zhang, H., Sturchler, E., Zhu, J., Nieto, A., Cistrone, P.A., Xie, J., He, L., Yea, K., Jones, T., Turn, R. et al. (2015) Autocrine selection of a GLP-1R G-protein biased agonist with potent antidiabetic effects. Nat Commun, 6, 8918.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

15. Xie, J., Zhang, H., Yea, K. and Lerner, R.A. (2013) Autocrine signaling based selection of combinatorial antibodies that transdifferentiate human stem cells. Proceedings of the National Academy of Sciences of the United States of America, 110, 8099-8104.
16. Stepanov, A.V., Markov, O.V., Chernikov, I.V., Gladkikh, D.V., Zhang, H., Jones, T., Sen'kova, A.V., Chernolovskaya, E.L., Zenkova, M.A., Kalinin, R.S. et al. (2018) Autocrine-based selection of ligands for personalized CAR-T therapy of lymphoma. Sci Adv, 4, eaau4580.
17. Zhang, H., Yea, K., Xie, J., Ruiz, D., Wilson, I.A. and Lerner, R.A. (2013) Selecting agonists from single cells infected with combinatorial antibody libraries. Chem Biol, 20, 734-741.
18. Li, S., Su, W. and Zhang, C. (2019) Linear double-stranded DNAs as innovative biological parts to implement genetic circuits in mammalian cells. FEBS J, 286, 2341-2354.
19. Bassing, C.H., Swat, W. and Alt, F.W. (2002) The mechanism and regulation of chromosomal V(D)J recombination. Cell, 109 Suppl, S45-55.
20. Schatz, D.G. (2004) V(D)J recombination. Immunol Rev, 200, 5-11.
21. Kussie, P.H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A.J. and Pavletich, N.P. (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science, 274, 948-953.
22. Chene, P. (2003) Inhibiting the p53-MDM2 interaction: an important target for cancer therapy. Nature reviews. Cancer, 3, 102-109.
23. Vassilev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C. et al. (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science, 303, 844-848.
24. Vassilev, L.T. (2007) MDM2 inhibitors for cancer therapy. Trends in molecular medicine, 13, 23-31.
25. Pazgier, M., Liu, M., Zou, G., Yuan, W., Li, C., Li, J., Monbo, J., Zella, D., Tarasov, S.G. and Lu, W. (2009) Structural basis for high-affinity peptide inhibition of p53 interactions with MDM2 and MDMX. Proceedings of the National Academy of Sciences of the United States of America, 106, 4665-4670.
26. Shiheido, H., Takashima, H., Doi, N. and Yanagawa, H. (2011) mRNA display selection of an optimized MDM2-binding peptide that potently inhibits MDM2-p53 interaction. PloS one, 6, e17898.
27. Nagata, T., Shirakawa, K., Kobayashi, N., Shiheido, H., Tabata, N., Sakuma-Yonemura, Y., Horisawa, K., Katahira, M., Doi, N. and Yanagawa, H. (2014) Structural basis for inhibition of the MDM2:p53 interaction by an optimized MDM2-binding peptide selected with mRNA display. PloS one, 9, e109163.
28. Dang, C.V., Barrett, J., Villa-Garcia, M., Resar, L.M., Kato, G.J. and Fearon, E.R. (1991) Intracellular leucine zipper interactions suggest c-Myc hetero-oligomerization. Molecular and cellular biology, 11, 954-962.
29. Fearon, E.R., Finkel, T., Gillison, M.L., Kennedy, S.P., Casella, J.F., Tomaselli, G.F., Morrow, J.S. and Van Dang, C. (1992) Karyoplasmic interaction selection strategy: a general strategy to detect protein-protein interactions in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America, 89, 7958-7962.
30. Smith, G.P. and Petrenko, V.A. (1997) Phage Display. Chem Rev, 97, 391-410.
31. Lamla, T. and Erdmann, V.A. (2003) Searching sequence space for high-affinity binding peptides using ribosome display. J Mol Biol, 329, 381-388.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

32. Rahal, E.A., Henricksen, L.A., Li, Y., Turchi, J.J., Pawelczak, K.S. and Dixon, K. (2008) ATM mediates repression of DNA end-degradation in an ATP-dependent manner. DNA Repair (Amst), 7, 464-475.

33. Chang, H.H., Watanabe, G., Gerodimos, C.A., Ochi, T., Blundell, T.L., Jackson, S.P. and Lieber, M.R. (2016) Different DNA End Configurations Dictate Which NHEJ Components Are Most Important for Joining Efficiency. J Biol Chem, 291, 24377-24389.

34. Cho, G., Keefe, A.D., Liu, R., Wilson, D.S. and Szostak, J.W. (2000) Constructing high complexity synthetic libraries of long ORFs using in vitro selection. J Mol Biol, 297, 309-319.

35. Janczyk, M., Appel, B., Springstubbe, D., Fritz, H.J. and Muller, S. (2012) A new and convenient approach for the preparation of beta-cyanoethyl protected trinucleotide phosphoramidites. Org Biomol Chem, 10, 1510-1513.

36. Suchsland, R., Appel, B. and Muller, S. (2018) Preparation of trinucleotide phosphoramidites as synthons for the synthesis of gene libraries. Beilstein J Org Chem, 14, 397-406.

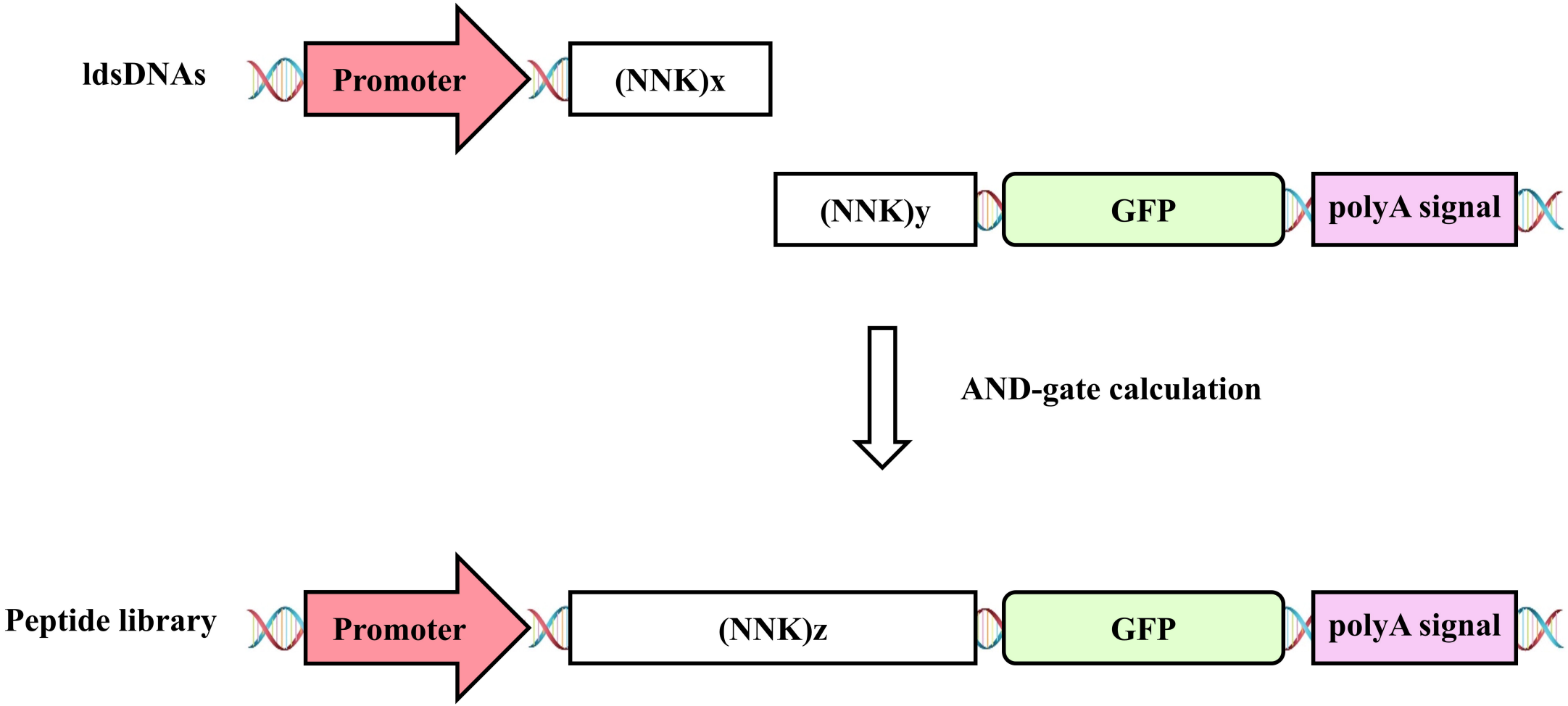
ldsDNAs
Promoter
(NNK)x
(NNK)y
GFP
polyA signal
AND-gate calculation
Peptide library
Promoter
(NNK)z
GFP
polyA signal

Figure 1

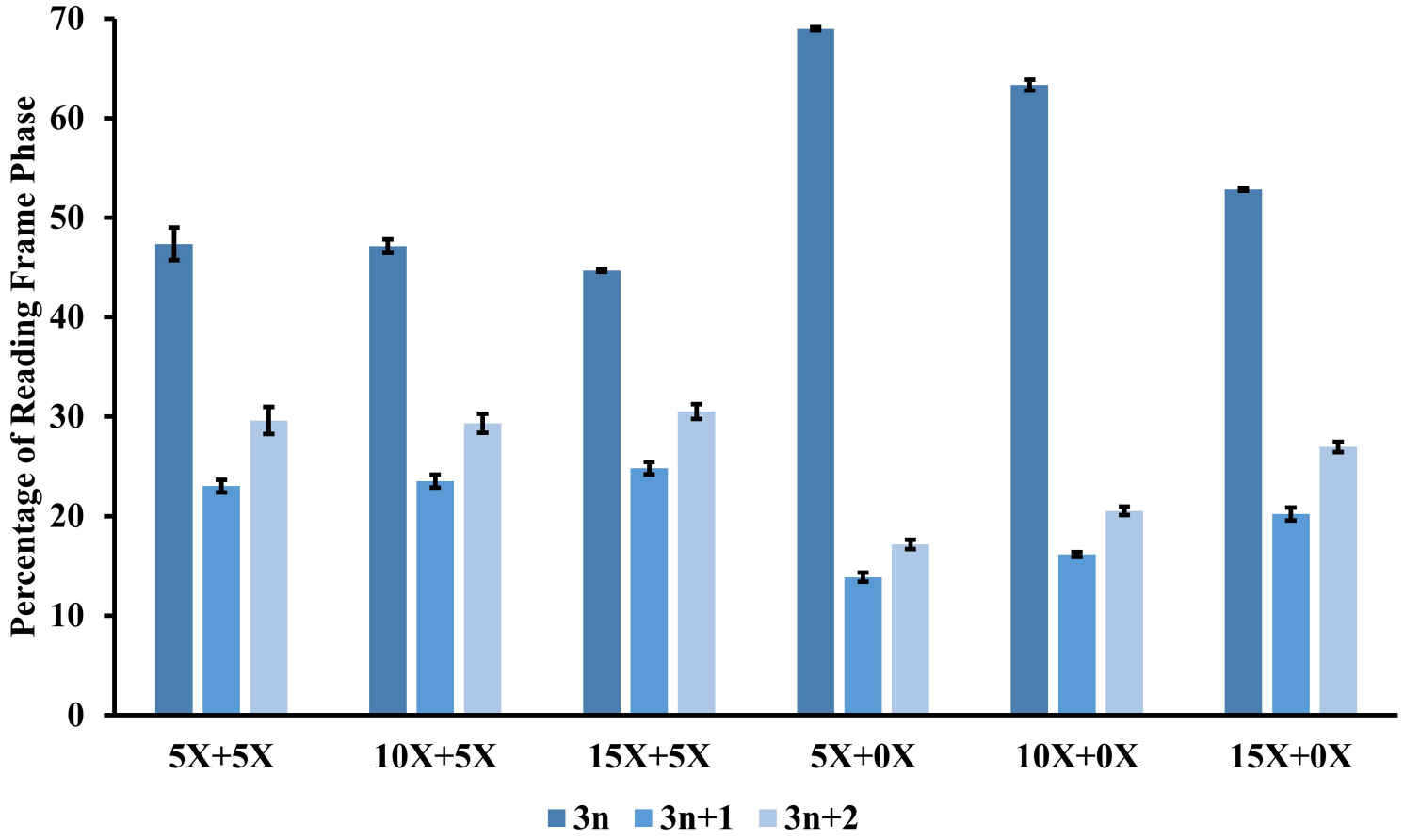
Percentage of Reading Frame Phase
70
60
50
40
30
20
10
0
5X+5X
10X+5X
15X+5X
5X+0X
10X+0X
15X+0X
3n
3n+1
3n+2

Figure 2

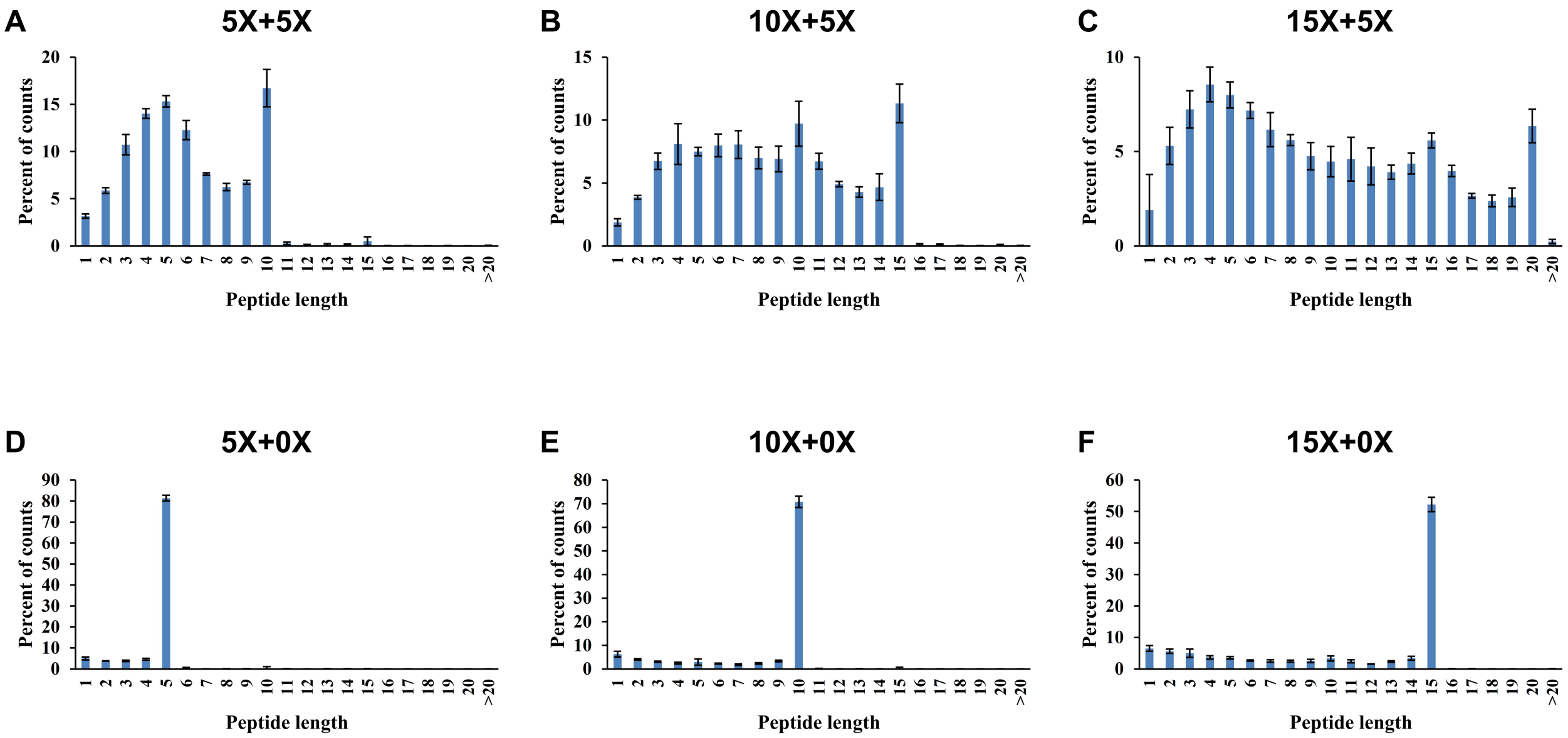
A
5X+5X
B
10X+5X
C
15X+5X
D
5X+0X
E
10X+0X
F
15X+0X
Percent of counts
Peptide length
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 >20

Figure 3

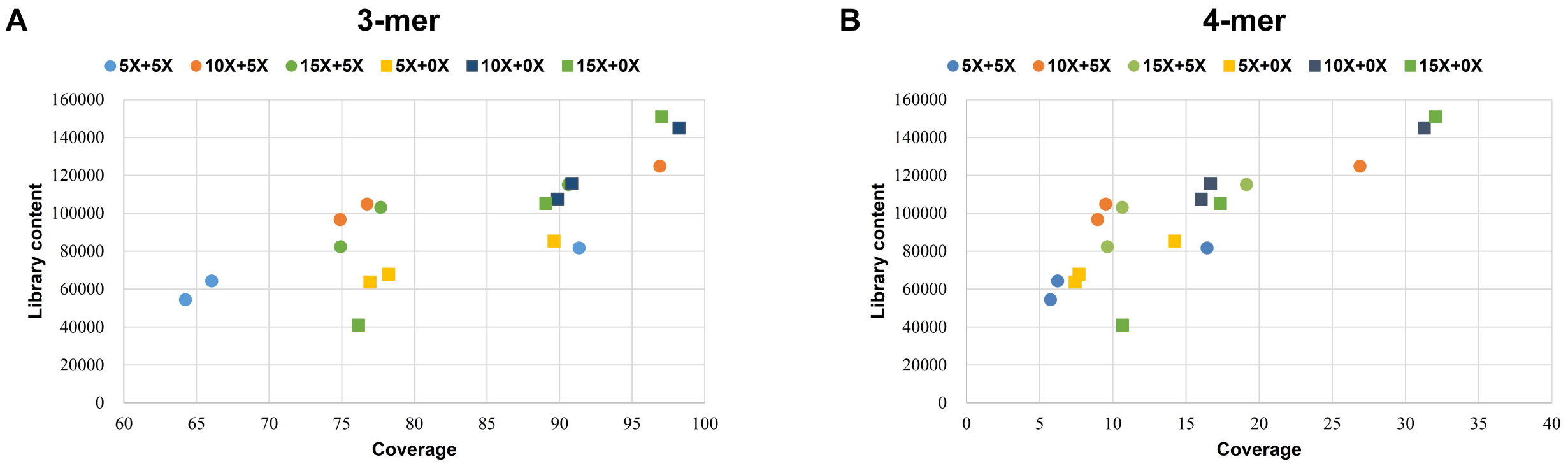
A
3-mer
5X+5X
10X+5X
15X+5X
5X+0X
10X+0X
15X+0X
Library content
Coverage
B
4-mer
5X+5X
10X+5X
15X+5X
5X+0X
10X+0X
15X+0X
Library content
Coverage

Figure 4

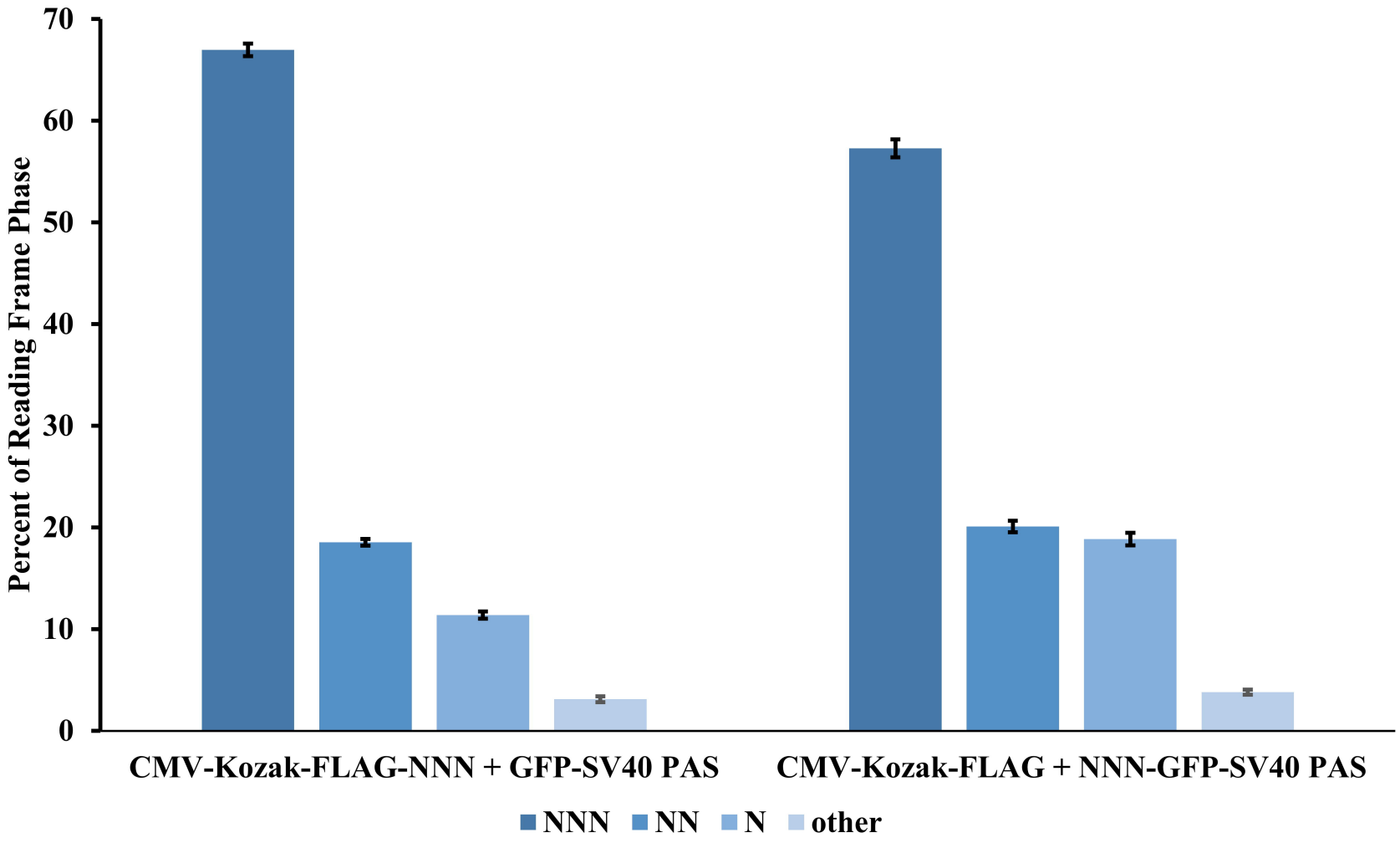
70
60
50
40
30
20
10
0
Percent of Reading Frame Phase
CMV-Kozak-FLAG-NNN + GFP-SV40 PAS
CMV-Kozak-FLAG + NNN-GFP-SV40 PAS
NNN
NN
N
other

Figure 5

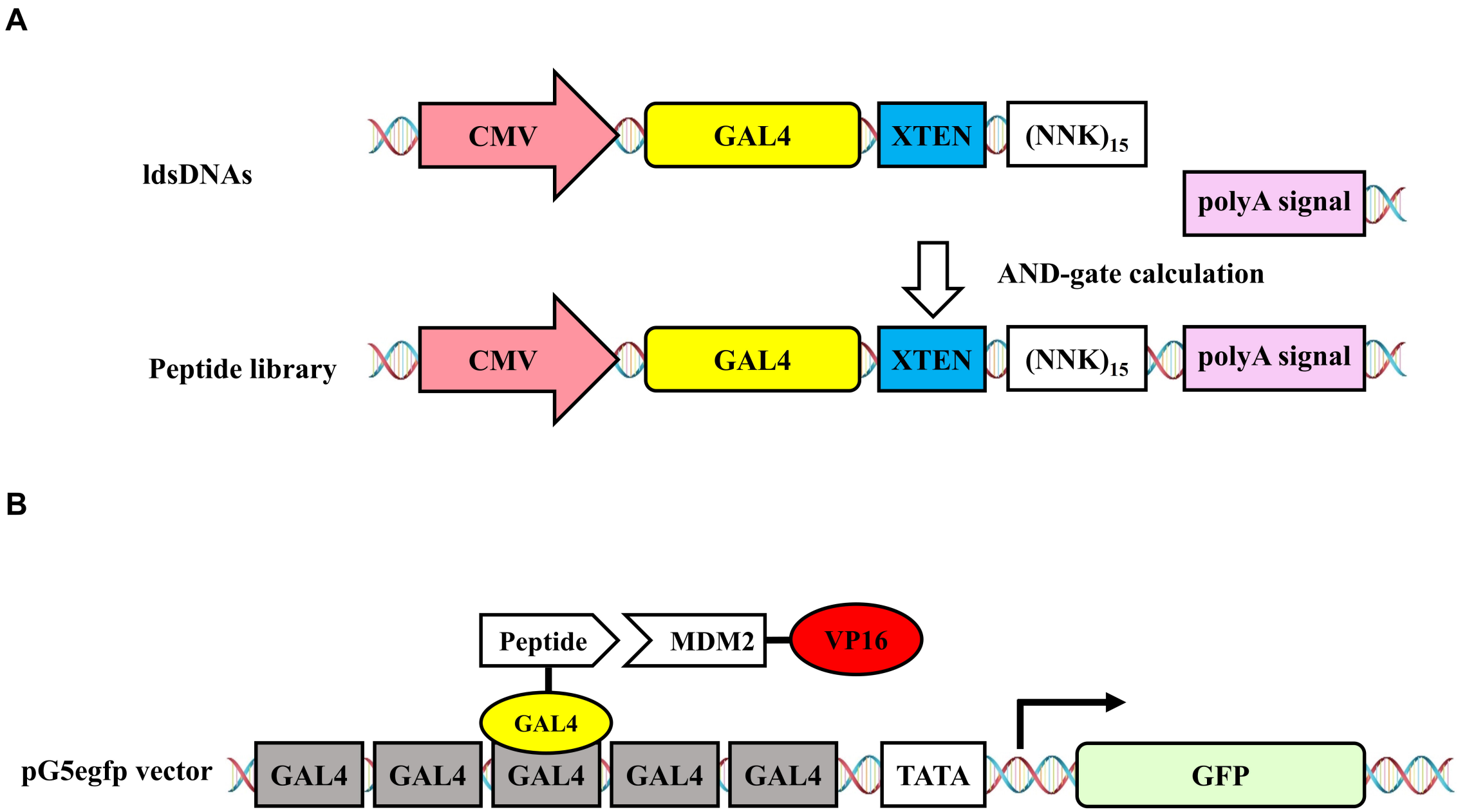
A
ldsDNAs
CMV
GAL4
XTEN
(NNK)15
polyA signal
AND-gate calculation
Peptide library
CMV
GAL4
XTEN
(NNK)15
polyA signal
B
Peptide
MDM2
VP16
GAL4
pG5egfp vector
GAL4
GAL4
GAL4
GAL4
GAL4
TATA
GFP

Figure 6

Table 1. High-throughput sequencing (HTS) statistics

Library	Description	Clean Read count	Read count in analysis
5X+5X (1)	Peptide library generated by ldsDNA CMV-Kozak-<NNK>$_5$ and <NNK>$_5$-GFP-SV40 PAS	79179	32698
5X+5X (2)		75950	25194
5X+5X (3)		61489	23021
10X+5X (1)	Peptide library generated by ldsDNA CMV-Kozak-<NNK>$_{10}$ and <NNK>$_5$-GFP-SV40 PAS	84191	38218
10X+5X (2)		75195	30324
10X+5X (3)		71341	32768
15X+5X (1)	Peptide library generated by ldsDNA CMV-Kozak-<NNK>$_{15}$ and <NNK>$_5$-GFP-SV40 PAS	67121	32273
15X+5X (2)		39153	24083
15X+5X (3)		82312	31784
5X+0X (1)	Peptide library generated by ldsDNA CMV-Kozak-<NNK>$_5$ and <NNK>$_0$-GFP-SV40 PAS	87718	30612
5X+0X (2)		83744	22487
5X+0X (3)		81241	23620
10X+0X (1)	Peptide library generated by ldsDNA CMV-Kozak-<NNK>$_{10}$ and <NNK>$_0$-GFP-SV40 PAS	82228	33130
10X+0X (2)		74850	25512
10X+0X (3)		86618	26676
15X+0X (1)	Peptide library generated by ldsDNA CMV-Kozak-<NNK>$_{15}$ and <NNK>$_0$-GFP-SV40 PAS	85127	33192
15X+0X (2)		30998	9570
15X+0X (3)		75915	23598

Pair-end high-throughput sequencing statistics for all libraries referenced in this study. Sequences formed by ldsDNA-based AND-gate calculation are subject for further analysis.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 2. Percentage of each amino acid of different peptide library

Amino acid	NNK	5X+5X	10X+5X	15X+5X	5X+0X	10X+0X	15X+0X
Stop codon	3.125	2.34 ± 0.11	2.46 ± 0.14	2.28 ± 0.15	2.59 ± 0.14	2.47 ± 0.09	2.31 ± 0.13
A (Ala)	6.25	6.16 ± 0.08	6.34 ± 0.18	7.11 ± 0.19	5.43 ± 0.12	6.22 ± 0.13	7.32 ± 0.26
C (Cys)	3.125	2.80 ± 0.10	2.72 ± 0.09	2.74 ± 0.24	2.06 ± 0.17	2.35 ± 0.03	2.16 ± 0.09
D (Asp)	3.125	2.84 ± 0.18	2.96 ± 0.26	2.76 ± 0.07	3.72 ± 0.07	3.13 ± 0.07	2.85 ± 0.05
E (Glu)	3.125	2.74 ± 0.07	2.75 ± 0.05	2.63 ± 0.11	3.67 ± 0.11	3.24 ± 0.12	3.19 ± 0.10
F (Phe)	3.125	3.34 ± 0.18	3.34 ± 0.08	2.88 ± 0.08	1.71 ± 0.05	2.52 ± 0.12	1.88 ± 0.04
G (Gly)	6.25	5.23 ± 0.23	5.28 ± 0.14	5.85 ± 0.08	4.35 ± 0.07	4.89 ± 0.13	5.68 ± 0.14
H (His)	3.125	3.54 ± 0.17	3.41 ± 0.18	3.26 ± 0.20	4.35 ± 0.11	3.72 ± 0.14	3.21 ± 0.10
I (Ile)	3.125	3.45 ± 0.08	3.40 ± 0.38	2.76 ± 0.12	3.49 ± 0.26	3.05 ± 0.12	2.48 ± 0.07
K (Lys)	3.125	3.57 ± 0.16	3.20 ± 0.06	2.70 ± 0.07	5.93 ± 0.17	4.08 ± 0.09	3.32 ± 0.15
L (Leu)	9.375	8.84 ± 0.16	9.35 ± 0.17	8.98 ± 0.28	6.69 ± 0.22	8.63 ± 0.11	7.93 ± 0.12
M (Met)	3.125	2.91 ± 0.09	2.90 ± 0.18	2.68 ± 0.03	3.54 ± 0.10	3.04 ± 0.04	2.59 ± 0.08
N (Asn)	3.125	4.23 ± 0.18	3.55 ± 0.16	2.90 ± 0.07	6.33 ± 0.09	4.37 ± 0.12	3.35 ± 0.12
P (Pro)	6.25	8.16 ± 0.15	9.20 ± 0.52	10.40 ± 0.28	7.34 ± 0.10	9.41 ± 0.16	11.32 ± 0.11
Q (Gln)	3.125	3.13 ± 0.11	3.12 ± 0.04	3.17 ± 0.09	4.13 ± 0.11	3.89 ± 0.02	3.82 ± 0.09
R (Arg)	9.375	8.34 ± 0.28	8.47 ± 0.24	9.79 ± 0.23	8.74 ± 0.08	9.18 ± 0.25	10.68 ± 0.29
S (Ser)	9.375	10.30 ± 0.16	9.76 ± 0.30	9.68 ± 0.20	8.54 ± 0.14	8.63 ± 0.17	8.90 ± 0.32
T (Thr)	6.25	7.24 ± 0.17	6.64 ± 0.27	6.94 ± 0.12	8.88 ± 0.21	7.27 ± 0.17	7.58 ± 0.15
V (Val)	6.25	5.66 ± 0.24	5.56 ± 0.19	5.31 ± 0.21	3.59 ± 0.11	4.58 ± 0.07	4.51 ± 0.21
W (Trp)	3.125	2.30 ± 0.06	2.70 ± 0.31	2.60 ± 0.04	1.96 ± 0.07	2.33 ± 0.05	2.36 ± 0.07
Y (Tyr)	3.125	2.89 ± 0.11	2.89 ± 0.21	2.59 ± 0.07	2.96 ± 0.11	2.99 ± 0.14	2.57 ± 0.16

In-frame (3n) nucleotide sequences generated by ldsDNA-based AND-gate are translated into polypeptide sequences. The total number of each amino acid in each

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

library is summarized. The average percentage of different amino acid is presented. All data are displayed as the mean ± SD, n = 3. The theoretical amino acid percentage of NNK degenerate codon is also presented.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Table 1. List of PCR primers used in study.

Amplicon	Forward primer	Reverse primer
CMV-Kozak-<NNK>$_5$	CTTTCCTGCGTTATCCCCTGATTCT	MNNMNNMNNMNNMNNCACCATGGTGGCGACCGGTAGCGCT
CMV-Kozak-<NNK>$_{10}$	CTTTCCTGCGTTATCCCCTGATTCT	MNNMNNMNNMNNMNNMNNMNNMNNMNNMNNCACCATGGTGGCGACCGGTAGCGCT
CMV-Kozak-<NNK>$_{15}$	CTTTCCTGCGTTATCCCCTGATTCT	MNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNCACCATGGTGGCGACCGGTAGCGCT
<NNK>$_5$-GFP-SV40 PAS	NNKNNKNNKNNKNNKATGGTGAGCAAGGGCGAGGAGCTGT	CGGCCTATTGGTTAAAAAATGAGCT
<NNK>$_0$-GFP-SV40 PAS	ATGGTGAGCAAGGGCGAGGAGCTGT	CGGCCTATTGGTTAAAAAATGAGCT
CMV-Kozak-FLAG-NNN	CTTTCCTGCGTTATCCCCTGATTCT	NNNTTTATCGTCATCGTCTTTGTAGTCCACCATGGTGGCGACCGGTAGCGCT
GFP-SV40 PAS	A*TGGTGAGCAAGGGCGAGGAGCTGT	CGGCCTATTGGTTAAAAAATGAGCT
CMV-Kozak-FLAG	CTTTCCTGCGTTATCCCCTGATTCT	T*TTATCGTCATCGTCTTTGTAGTCCACCATGGTGGCGACCGGTAGCGCT
NNN-GFP-SV40	NNNATGGTGAGCAAGGGCGAGGAGCTGT	CGGCCTATTGGTTAAAAAATGAGCT
Amplicon for PE-sequencing library construction	[Barcode]TCAGATCCGCTAGCGCTACC	[Barcode]GAACTTCAGGGTCAGCTTGC
CMV-GAL4-XTEN-<NNK>$_{15}$	TACGGTTCCTGGCCTTTTGCTGGCC	MNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNMNNGGATTCGGGTGTGGCGGACTCTGAG
SV40 PAS	CTCTAGATAACTGAGGTACCTGAATAACTAAGGCCGCTT	ATACCGCATCAGGAAATTGTAAGCG

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Amplicon for PE-sequencing library construction (MDM2 interacting peptide library)	[Barcode]GGTCAAAGACAGTTGACTGT	[Barcode]ACCCTCACTAAAGGGAAGCG

N = T or C or A or G; K = T or G; M = C or A; * = phosphorothioate bond.

Supplementary Table 2. Percentage of each nucleotide at each position of peptide library codons

Codon position	Nucleotide	5X+5X	10X+5X	15X+5X	5X+0X	10X+0X	15X+0X
	T	23.69 ± 0.30	24.31 ± 0.45	22.86 ± 0.35	17.96 ± 0.45	21.21 ± 0.21	19.67 ± 0.36
	C	26.43 ± 0.30	28.23 ± 0.59	30.39 ± 0.29	25.64 ± 0.34	29.42 ± 0.28	31.93 ± 0.41
	A	27.26 ± 0.23	24.58 ± 0.35	23.09 ± 0.13	35.64 ± 0.22	27.31 ± 0.27	24.84 ± 0.15
First	G	22.63 ± 0.18	22.89 ± 0.30	23.66 ± 0.24	20.75 ± 0.11	22.06 ± 0.24	23.55 ± 0.42
	T	24.20 ± 0.40	24.55 ± 0.67	22.60 ± 0.10	19.01 ± 0.02	21.82 ± 0.24	19.39 ± 0.35
	C	28.78 ± 0.19	29.43 ± 0.80	31.66 ± 0.35	26.53 ± 0.31	28.91 ± 0.45	32.50 ± 0.29
	A	25.23 ± 0.38	24.29 ± 0.39	22.19 ± 0.34	33.66 ± 0.07	27.87 ± 0.39	24.55 ± 0.50
Second	G	21.79 ± 0.29	21.74 ± 0.10	23.55 ± 0.09	20.80 ± 0.35	21.40 ± 0.33	23.55 ± 0.13
	T	52.50 ± 0.42	51.13 ± 0.24	47.80 ± 0.04	49.64 ± 0.74	48.56 ± 0.34	45.07 ± 0.44
	C	0.65 ± 0.04	0.66 ± 0.06	1.51 ± 0.03	0.48 ± 0.05	0.31 ± 0.03	0.83 ± 0.08
	A	0.86 ± 0.10	0.61 ± 0.04	1.03 ± 0.08	0.42 ± 0.03	0.36 ± 0.04	0.60 ± 0.05
Third	G	45.98 ± 0.31	47.60 ± 0.29	49.66 ± 0.12	49.46 ± 0.71	50.77 ± 0.33	53.50 ± 0.33

The total number of each nucleotide at each position of peptide library codons is summarized. The average percentage of different nucleotide at each codon base is presented. All data are displayed as the mean ± SD, n = 3.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Table 3. Percentage of triple-nucleotides formed by LBAG calculation

Triple nucleotides	CMV-Kozak-FLAG-NNN + GFP-SV40 PAS	CMV-Kozak-FLAG + NNN-GFP-SV40 PAS
AAA	2.00 ± 0.21	1.93 ± 0.52
AAC	3.29 ± 0.11	4.82 ± 0.31
AAG	2.22 ± 0.23	0.28 ± 0.06
AAT	2.92 ± 0.20	2.78 ± 0.11
ACA	2.18 ± 0.33	1.19 ± 0.26
ACC	3.74 ± 0.58	2.14 ± 0.21
ACG	1.68 ± 0.12	0.13 ± 0.11
ACT	2.53 ± 0.09	1.31 ± 0.18
AGA	0.85 ± 0.10	0.59 ± 0.12
AGC	1.48 ± 0.16	1.58 ± 0.20
AGG	0.76 ± 0.14	0.07 ± 0.01
AGT	0.68 ± 0.12	0.88 ± 0.25
ATA	1.77 ± 0.18	1.76 ± 0.32
ATC	3.15 ± 0.17	5.06 ± 0.63
ATG	1.86 ± 0.15	0.21 ± 0.04
ATT	2.88 ± 0.19	2.27 ± 0.24
CAA	1.84 ± 0.30	2.52 ± 0.24
CAC	2.45 ± 0.37	4.58 ± 0.77
CAG	1.54 ± 0.34	0.27 ± 0.11
CAT	2.33 ± 0.37	3.37 ± 0.96
CCA	1.53 ± 0.33	2.13 ± 0.14
CCC	1.89 ± 0.09	4.29 ± 0.10
CCG	1.17 ± 0.26	0.29 ± 0.07
CCT	2.11 ± 0.21	2.18 ± 0.39
CGA	0.79 ± 0.07	0.53 ± 0.17
CGC	0.93 ± 0.13	1.08 ± 0.25
CGG	0.53 ± 0.04	0.12 ± 0.06
CGT	0.60 ± 0.05	0.74 ± 0.30
CTA	1.50 ± 0.32	2.48 ± 0.25
CTC	2.11 ± 0.26	5.69 ± 0.43
CTG	1.34 ± 0.25	0.48 ± 0.22
CTT	1.93 ± 0.14	2.94 ± 0.25
GAA	1.19 ± 0.22	0.48 ± 0.13
GAC	1.81 ± 0.36	1.15 ± 0.24
GAG	1.18 ± 0.06	0.07 ± 0.02
GAT	1.58 ± 0.17	0.75 ± 0.22
GCA	1.24 ± 0.24	0.40 ± 0.04
GCC	2.00 ± 0.20	0.92 ± 0.16
GCG	0.92 ± 0.10	0.11 ± 0.09
GCT	1.68 ± 0.32	0.50 ± 0.07

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

GGA	0.38 ± 0.09	0.20 ± 0.03
GGC	0.65 ± 0.15	0.47 ± 0.10
GGG	0.29 ± 0.06	0.02 ± 0.02
GGT	0.33 ± 0.07	0.20 ± 0.03
GTA	0.80 ± 0.19	0.58 ± 0.20
GTC	1.35 ± 0.47	0.92 ± 0.17
GTG	0.78 ± 0.11	0.10 ± 0.09
GTT	1.25 ± 0.04	0.51 ± 0.02
TAA	1.69 ± 0.14	2.12 ± 0.21
TAC	2.65 ± 0.49	3.70 ± 0.25
TAG	1.24 ± 0.13	0.32 ± 0.08
TAT	1.95 ± 0.27	3.04 ± 0.25
TCA	1.51 ± 0.38	2.38 ± 0.25
TCC	2.47 ± 0.13	3.97 ± 0.17
TCG	1.45 ± 0.22	0.23 ± 0.07
TCT	1.86 ± 0.18	2.48 ± 0.24
TGA	0.74 ± 0.18	0.96 ± 0.12
TGC	1.03 ± 0.03	1.60 ± 0.25
TGG	0.62 ± 0.15	0.16 ± 0.06
TGT	0.57 ± 0.06	1.11 ± 0.30
TTA	1.03 ± 0.14	1.86 ± 0.36
TTC	1.99 ± 0.26	4.85 ± 0.44
TTG	1.37 ± 0.05	0.33 ± 0.24
TTT	1.85 ± 0.15	2.82 ± 0.28

The total number of each triple-nucleotides formed by LBAG calculation is summarized. The average percentage of different triple-nucleotides is presented. All data are displayed as the mean ± SD, n = 4.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Table 4. Potential MDM2 interacting peptide

No	Peptide sequence	Read counts
1	C A G M I K Q S Q L R H P K S	47
2	K L R I MM Q A F R H V Y W N	39
3	D R Q T N Y G N Q L P A R D V	25
4	S **F** V I R **W** T A **L** W P G P R S	23
5	M Y S K W D P M R H D F S E T	21
6	H S L D L Q N Q I Q T C E E W	20
7	S T P W L A M M L R L G P F N	17
8	N S C R L M L H P Q P M R P L	16
9	P T S R E F C N A I H A C L V	14
10	K Y F C M P E T R M H P R R T	14

High-throughput sequencing decoded top ten hit amino acid sequences in GFP positive cells obtained from the LBAG mammalian two-hybrid system. Three crucial residues (Phe, Trp and Leu) mediating MDM2-p53 interaction are shown in underline bold.

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary materials

ldsDNA sequences in study:

>CMV-Kozak-<NNK>$_5$

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCATTAGT
TATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCC
GCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCG
CCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTC
CATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCC
CGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTAC
ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAA
TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC
GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT
ATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACCGGTCGCC
ACCATGGTGNNKNNKNNKNNKNNK

>CMV-Kozak-<NNK>$_{10}$

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCATTAGT
TATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCC
GCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCG
CCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTC
CATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCC
CGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTAC
ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAA
TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC
GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT
ATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACCGGTCGCC
ACCATGGTGNNKNNKNNKNNKNNKNNKNNKNNKNNKNNK

>CMV-Kozak-<NNK>$_{15}$

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCATTAGT
TATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCC
GCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCG
CCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTC
CATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCC
CGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTAC

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAA
TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC
GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT
ATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACCGGTCGCC
ACCATGGTGNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKN
NK

>$<NNK>_5$-GFP-SV40 PAS

NNKNNKNNKNNKNNKATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGT
GCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCC
GGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGC
ACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACG
GCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAA
GTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGAC
GGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC
AAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAG
AAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGC
GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTG
CTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCA
ACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCA
CTCTCGGCATGGACGAGCTGTACAAGTCCGGACTCAGATCTCGAGCTCAAGCTTC
GAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGATCTAGATAACTG
ATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCC
ACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTG
TTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAA
ATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTA
TCTTAACGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTT
GTTAAATCAGCTCATTTTTTAACCAATAGGCCG

>$<NNK>_0$-GFP-SV40 PAS

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAG
CTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC
GATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC
CCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAG
CCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAA
GGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACC
CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAG
GGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC
TACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAG

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

GTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGAC
CACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACC
ACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGA
GCTGTACAAGTCCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGAC
GGTACCGCGGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATA
CCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAAC
CTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAA
TGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCA
CTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAACGCGTAAATTG
TAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTT
TTTAACCAATAGGCCG

>CMV-Kozak-FLAG-NNN

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCATTAGT
TATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCC
GCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCG
CCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTC
CATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCC
CGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTAC
ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAA
TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC
GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT
ATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACCGGTCGCC
ACCATGGTGGACTACAAAGACGATGACGATAAANNN

>GFP-SV40 PAS

A*TGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAG
CTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC
GATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC
CCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAG
CCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAA
GGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACC
CGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAG
GGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC
TACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAG
GTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGAC
CACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACC

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGA
GCTGTACAAGTCCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGAC
GGTACCGCGGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAGCCATA
CCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAAC
CTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAA
TGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCA
CTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAACGCGTAAATTG
TAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTT
TTTAACCAATAGGCCG

* = phosphorothioate bond

>CMV-Kozak-FLAG

CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCATGCATTAGT
TATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCC
GCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCG
CCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTC
CATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCC
CGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTAC
ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAA
TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC
GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCT
ATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACCGGTCGCC
ACCATGGTGGACTACAAAGACGATGACGATAA*A

* = phosphorothioate bond

>NNN-GFP-SV40 PAS

NNNATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG
GGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGC
TGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTT
CAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCC
GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAG
ACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG
AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC
AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATC
AAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCC

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

GACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGAC
AACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGC
GATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGG
ACGAGCTGTACAAGTCCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGT
CGACGGTACCGCGGGCCCGGGATCCACCGGATCTAGATAACTGATCATAATCAG
CCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCC
TGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCT
TATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTT
TTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAACGCGTA
AATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCT
CATTTTTTAACCAATAGGCCG

> CMV-GAL4-XTEN-<NNK>$_{15}$

TACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGGCTCGACAGATCTTCAAT
ATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTAT
TGGCCATTGCATACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATG
TCCAATATGACCGCCATGTTGGCATTGATTATTGACTAGTTATTAATAGTAATCAA
TTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTAC
GGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATA
ATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGG
TGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCC
AAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCC
CAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCAT
CGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGG
TTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTT
TTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTGCGATCGCCCG
CCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCA
GAGCTCGTTTAGTGAACCGTCAGATCACTAGAAGCTTTATTGCGGTAGTTTATCA
CAGTTAAATTGCTAACGCAGTCAGTGCTTCTGACACAACAGTCTCGAACTTAAGC
TGCAGTGACTCTCTTAAGGTAGCCTTGCAGAAGTTGGTCGTGAGGCACTGGGCAG
GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCTTG
TCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACA
TCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCCAGTTCAATTACAGCTCTTAA
GGCTAGAGTACTTAATACGACTCACTATAGGCTAGCCAGCTTGAAGCAAGCCTCC
TGAAAGATGAAGCTACTGTCTTCTATCGAACAAGCATGCGATATTTGCCGACTTA
AAAAGCTCAAGTGCTCCAAAGAAAAACCGAAGTGCGCCAAGTGTCTGAAGAACA
ACTGGGAGTGTCGCTACTCTCCCAAAACCAAAAGGTCTCCGCTGACTAGGGCACA
TCTGACAGAAGTGGAATCAAGGCTAGAAAGACTGGAACAGCTATTTCTACTGATT
TTTCCTCGAGAAGACCTTGACATGATTTTGAAAATGGATTCTTTACAGGATATAA
AAGCATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAAAGATGCCGTCAC
AGATAGATTGGCTTCAGTGGAGACTGATATGCCTCTAACATTGAGACAGCATAGA
ATAAGTGCGACATCATCATCGGAAGAGAGTAGTAACAAAGGTCAAAGACAGTTG
ACTGTATCGCCGGAATTCCCGGGGATCCGTCGACTTGACGCGTTGATATCATCTA

bioRxiv preprint doi: https://doi.org/10.1101/392498; this version posted September 19, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

GAAGCGGCAGCGAGACTCCCGGGACCTCAGAGTCCGCCACACCCGAATCCNNKN
NKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNNK

>SV40 PAS

CTCTAGATAACTGAGGTACCTGAATAACTAAGGCCGCTTCCCTTTAGTGAGGGTT
AATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACT
AGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATT
TGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTT
ATGTTTCAGGTTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCT
ACAAATGTGGTAAAATCCGATAAGGATCGATCCGGGCTGGCGTAATAGCGAAGA
GGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACG
CGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGA
CCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTC
ACGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGG
GTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGAT
GGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGG
AGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCC
TATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTT
AAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACG
CTTACAATTTCCTGATGCGGTAT