Supplementary information, Data S1 Sequences and material and methods

TALE-Ctrl

```
   1  GTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTG
  61  ATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGT
 121  GCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATC
 181  TGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGAC
 241  ATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCAT
 301  ATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACG
 361  ACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTT
 421  TCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAG
 481  TGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGC
 541  ATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAG
 601  TCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGT
 661  TTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC
 721  ACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGG
 781  GCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGCGCGTTTTGCCTGTACTGGGTCT
 841  CTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
 901  AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGAC
 961  TCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGC
1021  GCCCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTC
1081  GGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAA
1141  TTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGG
1201  GGAGAATTAGATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATA
1261  AATTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCC
1321  TGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGA
1381  CAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATC
1441  AAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACA
1501  AAAGTAAGACCACCGCACAGCAAGCGGCCGGCCGCGCTGATCTTCAGACCTGGAGGAGGA
1561  GATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCA
1621  TTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTG
1681  GGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCG
1741  TCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAAC
1801  AATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATC
1861  AAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTG
1921  GGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGT
1981  TGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGA
2041  GAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAA
2101  GAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTT
2161  AACATAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTA
2221  GGTTTAAGAATAGTTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCA
2281  CCATTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATA
2341  GAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCGGCA
2401  CTGCGTGCGCCAATTCTGCAGACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGG
2461  GGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATACA
2521  AACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGGA
2581  CAGCAGAGATCCAGTTTGGTTAGTACCGGGCCCGCTCTAGACGATGTACGGGCCAGATAT
2641  ACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTT
2701  CATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGA
2761  CCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCA
2821  ATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCA
2881  GTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGG
2941  CCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATC
3001  TACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGT
3061  GGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGT
3121  TTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTG
3181  ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTA
3241  ACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACC
3301  CAAGCTGGCTAGCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCATGAGGACCAGGC
3361  TGCCATCTCCCCCTGCCCCTTCCCCCGCATTTAGCGCTGGGAGCTTTAGCGACCTGCTTA
3421  GGCAGTTCGACCCCAGCTTGTTCAACACCAGCCTGTTTGACAGCCTGCCTCCCTTCGGAG
```

```
3481   CGCACCACACCGAAGCCGCCACCGGCGAGTGGGACGAGGTGCAAAGCGGCCTGAGGGCAG
3541   CGGACGCTCCTCCGCCAACCATGAGGGTGGCAGTGACAGCAGCTAGGCCCCCTCGGGCAA
3601   AACCTGCACCCAGGAGAAGGGCTGCCCAACCCAGCGACGCGAGTCCAGCCGCACAGGTGG
3661   ACCTCAGGACGCTGGGCTACAGCCAGCAACAGCAAGAGAAGATCAAGCCCAAAGTAAGGA
3721   GCACCGTGGCCCAGCACCACGAGGCCCTGGTGGGTCACGGCTTCACCCACGCGCATATCG
3781   TTGCTCTGAGCCAACATCCCGCAGCTCTGGGTACCGTTGCGGTGAAGTATCAGGACATGA
3841   TCGCGGCACTGCCTGAAGCTACACACGAAGCCATAGTGGGCGTTGGCAAGCAGTGGAGCG
3901   GTGCCAGAGCGCTTGAGGCACTGTTGACGGTGGCTGGCGAGCTGAGGGGACCGCCACTGC
3961   AACTGGACACCGGCCAACTGCTGAAGATCGCCAAGAGGGGAGGCGTGACGGCGGTGGAGG
4021   CCGTGCATGCCTGGAGGAATGCCCTGACCGGCGCGCCCCTGAACCTGACACCAGAGCAAG
4081   TAGTGGCTATTGCAAGTAACATCGGTGGCAAACAAGCGCTGGAGACCGTGCAGAGGCTCC
4141   TTCCGGTGCTCTGCCAAGCACACGGTTTGACCCCCGAACAGGTTGTAGCCATAGCTTCTA
4201   ACAACGGAGGTAAGCAGGCACTGGAAACCGTGCAGCGCCTGCTCCCAGTACTGTGTCAGG
4261   CTCATGGGCTCACTCCGGAACAGGTGGTCGCAATCGCGAGCAACGGCGGCGGCAAGCAAG
4321   CCCTTGAGACAGTCCAAAGACTTTTGCCTGTCCTTTGTCAGGCGCATGGCCTTACGCCTG
4381   AGCAAGTCGTTGCGATCGCCTCCCACGACGGCGGAAAACAGGCTTTGGAAACCGTGCAGC
4441   GGTTGCTGCCCGTTTTGTGCCAAGCCCACGGATTGACCCCCGAACAGGTTGTAGCCATAG
4501   CTTCTAACATCGGAGGTAAGCAGGCACTGGAAACCGTGCAGCGCCTGCTCCCAGTACTGT
4561   GTCAGGCTCATGGGCTCACTCCGGAACAGGTGGTCGCAATCGCGAGCAACGGCGGCGGCA
4621   AGCAAGCCCTTGAGACAGTCCAAAGACTTTTGCCTGTCCTTTGTCAGGCGCATGGCCTTA
4681   CGCCTGAGCAAGTCGTTGCGATCGCCTCCCACGACGGCGGAAAACAGGCTTTGGAAACCG
4741   TGCAGCGGTTGCTGCCCGTTTTGTGCCAAGCCCACGGATTGACCCCCGAACAGGTTGTAG
4801   CCATAGCTTCTAACAACGGAGGTAAGCAGGCACTGGAAACCGTGCAGCGCCTGCTCCCAG
4861   TACTGTGTCAGGCTCATGGGCTTACGCCTGAGCAAGTCGTTGCGATCGCCTCCCACGACG
4921   GCGGAAAACAGGCTTTGGAAACCGTGCAGCGGTTGCTGCCCGTTTTGTGCCAAGCCCACG
4981   GACTCACTCCGGAACAGGTGGTCGCAATCGCGAGCAACGGCGGCGGCAAGCAAGCCCTTG
5041   AGACAGTCCAAAGACTTTTGCCTGTCCTTTGTCAGGCGCATGGCCTGACACCAGAGCAAG
5101   TAGTGGCTATTGCAAGTAACATCGGTGGCAAACAAGCGCTGGAGACCGTGCAGAGGCTCC
5161   TTCCGGTGCTCTGCCAAGCACACGGTTTGACCCCCGAACAGGTTGTAGCCATAGCTTCTA
5221   ACGGCGGAGGTAAGCAGGCACTGGAAACCGTGCAGCGCCTGCTCCCAGTACTGTGTCAGG
5281   CTCATGGGCTCACTCCGGAACAGGTGGTCGCAATCGCGAGCAACGGCGGCGGCAAGCAAG
5341   CCCTTGAGACAGTCCAAAGACTTTTGCCTGTCCTTTGTCAGGCGCATGGCCTGACACCAG
5401   AGCAAGTAGTGGCTATTGCAAGTAACATCGGTGGCAAACAAGCGCTGGAGACCGTGCAGA
5461   GGCTCCTTCCGGTGCTCTGCCAAGCACACGGTCTCACTCCGGAACAGGTGGTCGCAATCG
5521   CGAGCCACGACGGCGGCAAGCAAGCCCTTGAGACAGTCCAAAGACTTTTGCCTGTCCTTT
5581   GTCAGGCGCATGGCCTTACGCCTGAGCAAGTCGTTGCGATCGCCTCCCACGACGGCGGAA
5641   AACAGGCTTTGGAAACCGTGCAGCGGTTGCTGCCCGTTTTGTGCCAAGCCCACGGACTGA
5701   CACCAGAGCAAGTAGTGGCTATTGCAAGTAACATCGGTGGCAAACAAGCGCTGGAGAGCA
5761   TCGTGGCCCAGCTGTCTCGGCCCGACCCTGCCCTCGCCGCTCTGACCAACGACCACCTGG
5821   TGGCCCTGGCTTGCCTCGGGGGCAGGCCAGCTCTTGACGCCGTGAAGAAGGGCCTTCCTC
5881   ACGCCCCAGCCCTGATCAAGCGGACCAACAGAAGGATTCCCGAGAGGACATCACATCGAG
5941   TGGCAGATCACGCGCAAGTGGTCCGCGTGCTCGGATTCTTCCAGTGTCACTCCCACCCCG
6001   CACAAGCGTTCGATGACGCCATGACTCAATTTGGTATGTCGAGACACGGACTGCTGCAGC
6061   TCTTTCGTAGAGTCGGTGTCACAGAACTCGAGGCCCGCTCGGGCACACTGCCTCCCGCCT
6121   CCCAGCGGTGGGACAGGATTCTCCAAGCGAGCGGTATGAAACGCGCGAAGCCTTCACCTA
6181   CGTCAACTCAGACACCTGACCAGGCGAGCCTTCATGCGTTCGCAGACTCGCTGGAGAGGG
6241   ATTTGGACGCGCCCTCGCCCATGCATGAAGGGGACCAAACTCGCGCGTCAGCTAGCCCCA
6301   AGAAGAAGAGAAAGGTGGAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTG
6361   ATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGG
6421   ATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTCGACC
6481   TGGACATGCTGATTAACTCTAGAGGCAGTGGAGAGGGCAGAGGAAGTCTGCTAACATGCG
6541   GTGACGTCGAGGAGAATCCTGGCCCAGTGAGCAAGGGCGAGGAGGATAACATGGCCATCA
6601   TCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCG
6661   AGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGG
6721   TGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACG
6781   GCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCC
6841   CCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGA
6901   CCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCA
6961   ACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCG
7021   AGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGA
7081   AGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGC
7141   AGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACT
7201   ACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACG
```

```
7261   AGCTGTACAAGTAACATGTTTAAGGGTTCCGGTTCCACTAGGTACAATTCGATATCAAGC
7321   TTATCGATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACT
7381   ATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTG
7441   CTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATG
7501   AGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAA
7561   CCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCC
7621   CCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGG
7681   CTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTT
7741   GGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTT
7801   CGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTC
7861   CGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCATC
7921   GATACCGTCGACCTCGATCGAGACCTAGAAAAACATGGAGCAATCACAAGTAGCAATACA
7981   GCAGCTACCAATGCTGATTGTGCCTGGCTAGAAGCACAAGAGGAGGAGGAGGTGGGTTTT
8041   CCAGTCACACCTCAGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGC
8101   CACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTCACTCCCAACGAAGACAAGAT
8161   ATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACA
8221   CCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAGTT
8281   GAGCAAGAGAAGGTAGAAGAAGCCAATGAAGGAGAGAACACCCGCTTGTTACACCCTGTG
8341   AGCCTGCATGGGATGGATGACCCGGAGAGAGAAGTATTAGAGTGGAGGTTTGACAGCCGC
8401   CTAGCATTTCATCACATGGCCCGAGAGCTGCATCCGGACTGTACTGGGTCTCTCTGGTTA
8461   GACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAA
8521   TAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAAC
8581   TAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGCATGTGAGCAAAA
8641   GGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
8701   CGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACA
8761   GGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCG
8821   ACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCT
8881   CATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGT
8941   GTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAG
9001   TCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGC
9061   AGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTAC
9121   ACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGA
9181   GTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGC
9241   AAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG
9301   GGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCA
9361   AAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGT
9421   ATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCA
9481   GCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACG
9541   ATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCA
9601   CCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGT
9661   CCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGT
9721   AGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCA
9781   CGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACA
9841   TGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGA
9901   AGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACT
9961   GTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGA
10021  GAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCG
10081  CCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTC
10141  TCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGA
10201  TCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAT
10261  GCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTT
10321  CAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGT
10381  ATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAC
```

2622-3276 bp: CMV promoter

3348-4064 bp: TALE N-terminus

4065-5756 bp: TALE repeat array (NI NN NG HD NI NG HD NN HD NG NI NG NG NI HD HD NI)

5757-6290 bp: TALE C-terminus

6291-6332 bp: NLS

6333-6497 bp: VP64

6504-6566 bp: 2A peptide
6567-7274 bp: mCherry

RVD Library Entry Vector

The following sequence was inserted into the TA-cloning site of pMD19-T vector (Takara, Inc.).

```
   1    TAGCTATACGTCTCATTGACCCCCGAACAGGTTGTAGCCATAGCTTCTAAGTCTTCAGAG
  61    ACGCTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGTTAAG
 121    CTATCAACAAGTTTGTACAAAAAAGCTGAACGAGAAACGTAAAATGATATAAATATCAAT
 181    ATATTAAATTAGATTTTGCATAAAAAACAGACTACATAATACTGTAAAACACAACATATC
 241    CAGTCACTATGAATCAACTACTTAGATGGTATTAGTGACCTGTAGTCGACCGACAGCCTT
 301    CCAAATGTTCTTCGGGTGATGCTGCCAACTTAGTCGACCGACAGCCTTCCAAATGTTCTT
 361    CTCAAACGGAATCGTCGTATCCAGCCTACTCGCTATTGTCCTCAATGCCGTATTAAATCA
 421    TAAAAAGAAATAAGAAAAAGAGGTGCGAGCCTCTTTTTTGTGTGACAAAATAAAAACATC
 481    TACCTATTCATATACGCTAGTGTCATAGTCCTGAAAATCATCTGCATCAAGAACAATTTC
 541    ACAACTCTTATACTTTTCTCTTACAAGTCGTTCGGCTTCATCTGGATTTTCAGCCTCTAT
 601    ACTTACTAAACGTGATAAAGTTTCTGTAATTTCTACTGTATCGACCTGCAGACTGGCTGT
 661    GTATAAGGGAGCCTGACATTTATATTCCCCAGAACATCAGGTTAATGGCGTTTTTGATGT
 721    CATTTTCGCGGTGGCTGAGATCAGCCACTTCTTCCCCGATAACGGAGACCGGCACACTGG
 781    CCATATCGGTGGTCATCATGCGCCAGCTTTCATCCCCGATATGCACCACCGGGTAAAGTT
 841    CACGGGAGACTTTATCTGACAGCAGACGTGCACTGGCCAGGGGGATCACCATCCGTCGCC
 901    CGGGCGTGTCAATAATATCACTCTGTACATCCACAAACAGACGATAACGGCTCTCTCTTT
 961    TATAGGTGTAAACCTTAAACTGCATTTCACCAGCCCCTGTTCTCGTCAGCAAAAGAGCCG
1021    TTCATTTCAATAAACCGGGCGACCTCAGCCATCCCTTCCTGATTTTCCGCTTTCCAGCGT
1081    TCGGCACGCAGACGACGGGCTTCATTCTGCATGGTTGTGCTTACCAGACCGGAGATATTG
1141    ACATCATATATGCCTTGAGCAACTGATAGCTGTCGCTGTCAACTGTCACTGTAATACGCT
1201    GCTTCATAGCATACCTCTTTTTGACATACTTCGGGTATACATATCAGTATATATTCTTAT
1261    ACCGCAAAAATCAGCGCGCAAATACGCATACTGTTATCTGGCTTTTAGTAAGCCGGATCC
1321    ACGCGGCGTTTACGCCCCCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATT
1381    CTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGC
1441    ACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCC
1501    ATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAA
1561    AACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACA
1621    TCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGAT
1681    GAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATC
1741    ACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGGCGGGCA
1801    AGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAG
1861    GCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCC
1921    TCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTT
1981    TTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGT
2041    AGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCAT
2101    TTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTAT
2161    TCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCT
2221    GCCAACTTAGTCGACTACAGGTCACTAATACCATCTAAGTAGTTGATTCATAGTGACTGG
2281    ATATGTTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTAATTTAATATA
2341    TTGATATTTATATCATTTTACGTTTCTCGTTCAGCTTTCTTGTACAAAGTGGTTGATCTA
2401    GAGGGCCCGCGGTTCGAACGTCTCTGAAGACAAGGAGGTAAGCAGGCACTGGAAACCGTG
2461    CAGCGCCTGCTCCCAGTACTGTGTCAGGCTCATGGGTGAGACGTATAGCTA
```

16-48 bp: N-terminus of the TALE repeats
57-2425 bp: ccdB cassette
2434-2496 bp: C-terminus of the TALE repeats

RVD Library Vector

```
   1  GTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTG
  61  ATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGT
 121  GCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATC
 181  TGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGAC
 241  ATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCAT
 301  ATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACG
 361  ACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTT
 421  TCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAG
 481  TGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGC
 541  ATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAG
 601  TCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGT
 661  TTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC
 721  ACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGG
 781  GCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGCGCGTTTTGCCTGTACTGGGTCT
 841  CTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTT
 901  AAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGAC
 961  TCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGC
1021  GCCCGAACAGGGACTTGAAAGCGAAAGGGAAACCAGAGGAGCTCTCTCGACGCAGGACTC
1081  GGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAA
1141  TTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGG
1201  GGAGAATTAGATCGCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATA
1261  AATTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCC
1321  TGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGA
1381  CAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATC
1441  AAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAACA
1501  AAAGTAAGACCACCGCACAGCAAGCGGCCGGCCGCGCTGATCTTCAGACCTGGAGGAGGA
1561  GATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCA
1621  TTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAAGAGCAGTG
1681  GGAATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCG
1741  TCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAAC
1801  AATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATC
1861  AAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTG
1921  GGGATTTGGGGTTGCTCTGGAAAACTCATTTGCACCACTGCTGTGCCTTGGAATGCTAGT
1981  TGGAGTAATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGA
2041  GAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAAACCAGCAA
2101  GAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGGCAAGTTTGTGGAATTGGTTT
2161  AACATAACAAATTGGCTGTGGTATATAAAATTATTCATAATGATAGTAGGAGGCTTGGTA
2221  GGTTTAAGAATAGTTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCA
2281  CCATTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATA
2341  GAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATCGGCA
2401  CTGCGTGCGCCAATTCTGCAGACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGG
2461  GGGGATTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATACA
2521  AACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGGA
2581  CAGCAGAGATCCAGTTTGGTTAGTACCGGGCCCGCTCTAGACGATGTACGGGCCAGATAT
2641  ACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTT
2701  CATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGA
2761  CCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCA
2821  ATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCA
2881  GTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGG
2941  CCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATC
3001  TACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGT
3061  GGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGT
3121  TTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTG
3181  ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTA
3241  ACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACC
3301  CAAGCTGGCTAGCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCATGAGGACCAGGC
3361  TGCCATCTCCCCCTGCCCCTTCCCCCGCATTTAGCGCTGGGAGCTTTAGCGACCTGCTTA
3421  GGCAGTTCGACCCCAGCTTGTTCAACACCAGCCTGTTTGACAGCCTGCCTCCCTTCGGAG
3481  CGCACCACACCGAAGCCGCCACCGGCGAGTGGGACGAGGTGCAAAGCGGCCTGAGGGCAG
3541  CGGACGCTCCTCCGCCAACCATGAGGGTGGCAGTGACAGCAGCTAGGCCCCCTCGGGCAA
```

```
3601   AACCTGCACCCAGGAGAAGGGCTGCCCAACCCAGCGACGCGAGTCCAGCCGCACAGGTGG
3661   ACCTCAGGACGCTGGGCTACAGCCAGCAACAGCAAGAGAAGATCAAGCCCAAAGTAAGGA
3721   GCACCGTGGCCCAGCACCACGAGGCCCTGGTGGGTCACGGCTTCACCCACGCGCATATCG
3781   TTGCTCTGAGCCAACATCCCGCAGCTCTGGGTACCGTTGCGGTGAAGTATCAGGACATGA
3841   TCGCGGCACTGCCTGAAGCTACACACGAAGCCATAGTGGGCGTTGGCAAGCAGTGGAGCG
3901   GTGCCAGAGCGCTTGAGGCACTGTTGACGGTGGCTGGCGAGCTGAGGGGACCGCCACTGC
3961   AACTGGACACCGGCCAACTGCTGAAGATCGCCAAGAGGGGAGGCGTGACGGCGGTGGAGG
4021   CCGTGCATGCCTGGAGGAATGCCCTGACCGGCGCGCCCCTGAACAGAGACGCTGGCTTAT
4081   CGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGTTAAGCTATCAACAAGT
4141   TTGTACAAAAAAGCTGAACGAGAAACGTAAAATGATATAAATATCAATATATTAAATTAG
4201   ATTTTGCATAAAAAACAGACTACATAATACTGTAAAACACAACATATCCAGTCACTATGA
4261   ATCAACTACTTAGATGGTATTAGTGACCTGTAGTCGACCGACAGCCTTCCAAATGTTCTT
4321   CGGGTGATGCTGCCAACTTAGTCGACCGACAGCCTTCCAAATGTTCTTCTCAAACGGAAT
4381   CGTCGTATCCAGCCTACTCGCTATTGTCCTCAATGCCGTATTAAATCATAAAAAGAAATA
4441   AGAAAAAGAGGTGCGAGCCTCTTTTTTGTGTGACAAAATAAAAACATCTACCTATTCATA
4501   TACGCTAGTGTCATAGTCCTGAAAATCATCTGCATCAAGAACAATTTCACAACTCTTATA
4561   CTTTTCTCTTACAAGTCGTTCGGCTTCATCTGGATTTTCAGCCTCTATACTTACTAAACG
4621   TGATAAAGTTTCTGTAATTTCTACTGTATCGACCTGCAGACTGGCTGTGTATAAGGGAGC
4681   CTGACATTTATATTCCCCAGAACATCAGGTTAATGGCGTTTTTGATGTCATTTTCGCGGT
4741   GGCTGAGATCAGCCACTTCTTCCCCGATAACGGAGACCGGCACACTGGCCATATCGGTGG
4801   TCATCATGCGCCAGCTTTCATCCCCGATATGCACCACCGGGTAAAGTTCACGGGAGACTT
4861   TATCTGACAGCAGACGTGCACTGGCCAGGGGGATCACCATCCGTCGCCCGGGCGTGTCAA
4921   TAATATCACTCTGTACATCCACAAACAGACGATAACGGCTCTCTCTTTTATAGGTGTAAA
4981   CCTTAAACTGCATTTCACCAGCCCCTGTTCTCGTCAGCAAAAGAGCCGTTCATTTCAATA
5041   AACCGGGCGACCTCAGCCATCCCTTCCTGATTTTCCGCTTTCCAGCGTTCGGCACGCAGA
5101   CGACGGGCTTCATTCTGCATGGTTGTGCTTACCAGACCGGAGATATTGACATCATATATG
5161   CCTTGAGCAACTGATAGCTGTCGCTGTCAACTGTCACTGTAATACGCTGCTTCATAGCAT
5221   ACCTCTTTTTGACATACTTCGGGTATACATATCAGTATATATTCTTATACCGCAAAAATC
5281   AGCGCGCAAATACGCATACTGTTATCTGGCTTTTAGTAAGCCGGATCCACGCGGCGTTTA
5341   CGCCCCCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGG
5401   AAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCT
5461   TGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACG
5521   TTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCA
5581   ATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATAT
5641   ATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCA
5701   GTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCG
5761   TCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATA
5821   AAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCC
5881   AGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCT
5941   TTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTA
6001   GCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATT
6061   TCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAG
6121   TTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGA
6181   TCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGCCAACTTAGTC
6241   GACTACAGGTCACTAATACCATCTAAGTAGTTGATTCATAGTGACTGGATATGTTGTGTT
6301   TTACAGTATTATGTAGTCTGTTTTTTTATGCAAAATCTAATTTAATATATTGATATTTATA
6361   TCATTTTACGTTTCTCGTTCAGCTTTCTTGTACAAAGTGGTTGATCTAGAGGGCCCGCGG
6421   TTCGAACGTCTCTAGCATCGTGGCCCAGCTGTCTCGGCCCGACCCTGCCCTCGCCGCTCT
6481   GACCAACGACCACCTGGTGGCCCTGGCTTGCCTCGGGGGCAGGCCAGCTCTTGACGCCGT
6541   GAAGAAGGGCCTTCCTCACGCCCCAGCCCTGATCAAGCGGACCAACAGAAGGATTCCCGA
6601   GAGGACATCACATCGAGTGGCAGATCACGCGCAAGTGGTCCGCGTGCTCGGATTCTTCCA
6661   GTGTCACTCCCACCCCGCACAAGCGTTCGATGACGCCATGACTCAATTTGGTATGTCGAG
6721   ACACGGACTGCTGCAGCTCTTTCGTAGAGTCGGTGTCACAGAACTCGAGGCCCGCTCGGG
6781   CACACTGCCTCCCGCCTCCCAGCGGTGGGACAGGATTCTCCAAGCGAGCGGTATGAAACG
6841   CGCGAAGCCTTCACCTACGTCAACTCAGACACCTGACCAGGCGAGCCTTCATGCGTTCGC
6901   AGACTCGCTGGAGAGGGATTTGGACGCGCCCTCGCCCATGCATGAAGGGGACCAAACTCG
6961   CGCGTCAGCTAGCCCCAAGAAGAAGAGAAAGGTGGAGGCCAGCGGTTCCGGACGGGCTGA
7021   CGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCT
7081   TGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGC
7141   CCTTGATGATTTCGACCTGGACATGCTGATTAACTCTAGAGGCAGTGGAGAGGGCAGAGG
7201   AAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGCCCAGTGAGCAAGGGCGAGGA
7261   GGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGT
7321   GAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCA
```

```
7381   GACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTC
7441   CCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTA
7501   CTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGG
7561   CGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGT
7621   GAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGG
7681   CTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAA
7741   GCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAA
7801   GGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCAC
7861   CTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTC
7921   CACCGGCGGCATGGACGAGCTGTACAAGTAACATGTTTAAGGGTTCCGGTTCCACTAGGT
7981   ACAATTCGATATCAAGCTTATCGATAATCAACCTCTGGATTACAAAATTTGTGAAAGATT
8041   GACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCC
8101   TTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTG
8161   GTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCAC
8221   TGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTC
8281   CGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGC
8341   CCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAA
8401   ATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTC
8461   CTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCC
8521   GGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTG
8581   GGCCGCCTCCCCGCATCGATACCGTCGACCTCGATCGAGACCTAGAAAAACATGGAGCAA
8641   TCACAAGTAGCAATACAGCAGCTACCAATGCTGATTGTGCCTGGCTAGAAGCACAAGAGG
8701   AGGAGGAGGTGGGTTTTCCAGTCACACCTCAGGTACCTTTAAGACCAATGACTTACAAGG
8761   CAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTCACT
8821   CCCAACGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTG
8881   ATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCT
8941   ACAAGCTAGTACCAGTTGAGCAAGAGAAGGTAGAAGAAGCCAATGAAGGAGAGAACACCC
9001   GCTTGTTACACCCTGTGAGCCTGCATGGGATGGATGACCCGGAGAGAGAAGTATTAGAGT
9061   GGAGGTTTGACAGCCGCCTAGCATTTCATCACATGGCCCGAGAGCTGCATCCGGACTGTA
9121   CTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACC
9181   CACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGT
9241   TGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTA
9301   GCAGCATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGG
9361   CGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGA
9421   GGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCG
9481   TGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGG
9541   GAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTC
9601   GCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCG
9661   GTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA
9721   CTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGT
9781   GGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAG
9841   TTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCG
9901   GTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATC
9961   CTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTT
10021  TGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTT
10081  TTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCA
10141  GTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCG
10201  TCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATAC
10261  CGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGG
10321  CCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCC
10381  GGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
10441  CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAAC
10501  GATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTC
10561  CTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCAC
10621  TGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACT
10681  CAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAA
10741  TACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTT
10801  CTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCA
10861  CTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAA
10921  AAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATAC
10981  TCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCG
11041  GATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCC
11101  GAAAAGTGCCACCTGAC
```

2622-3276 bp: CMV promoter
3348-4064 bp: TALE N-terminus
4065-6433 bp: ccdB cassette
6434-6967 bp: TALE C-terminus
6968-7009 bp: NLS
7010-7174 bp: VP64
7181-7243 bp: 2A peptide
7244-7951 bp: mCherry

SUPPLEMENTARY METHODS

Artificial system for RVD screening

The artificial screening system was composed of four reporters and a TALE-VP64 expression library in which the RVDs of three consecutive monomers in the middle of an artificial TALE array were encoded by the same 6 randomly synthesized nucleotides (TALE-$(XX')_3$). TALE-$(XX')_3$ contained 14.5 repeats fused with the VP64 trans-activation domain and 2A peptide-linked mCherry. The variable diresidues (XX') for testing were placed in the 7^{th} - 9^{th} repeat modules, and the variable RVD-carrying TALE modules were purposely designed as triplets to augment the DNA binding capability. X and X' represents the 12^{th} and 13^{th} amino acids in the 7^{th} - 9^{th} repeat modules, respectively. Four reporters consist of TALE-$(XX')_3$ binding sites CTGGCCNNNTACGTA, in which N represents A, T, C or G, was located immediately upstream of a minimal CMV promoter (P_{minCMV}) and its downstream EGFP gene. TALE-Ctrl was constructed to have the identical backbone as TALE-$(XX')_3$ except that its TALE repeats (16.5-mers) are different, not matching with any reporters. The mCherry-normalized EGFP level of TALE-Ctrl co-transfected with reporter served as the corresponding basal level. For each sample, EGFP fluorescence intensity was normalized to mCherry intensity. Fold induction is calculated as the result of normalized sample EGFP intensity divided by normalized basal level shown as follows.

$$\textbf{Fold Induction} = \frac{\left(\prod_{i=1}^{n} \textbf{Exp}_{\textbf{EGFP}_i}\right)^{1/n} \Big/ \left(\prod_{i=1}^{n} \textbf{Exp}_{\textbf{mCherry}_i}\right)^{1/n}}{\left(\prod_{i=1}^{n} \textbf{Ctrl}_{\textbf{EGFP}_i}\right)^{1/n} \Big/ \left(\prod_{i=1}^{n} \textbf{Ctrl}_{\textbf{mCherry}_i}\right)^{1/n}}$$

Notes:

$\left(\prod_{i=1}^{n} \textbf{Exp}_{\textbf{EGFP}_i}\right)^{1/n}$: Geometric mean of $\textbf{Exp}_{\textbf{EGFP}}$ of cells from FACS (n = number of cells)

$\textbf{Exp}_{\textbf{EGFP}}$: EGFP intensity of HEK293T cells co-transfected with TALE-$(XX')_m$ plus reporter (m = 3, 6 and 12, corresponding to TALE-$(XX')_3$, TALE-$(XX')_6$ and TALE-$(XX')_{12}$, respectively)

$\text{Exp}_{\text{mCherry}}$: mCherry intensity of HEK293T cells co-transfected with TALE-$(XX')_m$ plus reporter

$\textbf{Ctrl}_{\textbf{EGFP}}$: EGFP intensity of HEK293T cells co-transfected with TALE-Ctrl plus reporter

$\textbf{Ctrl}_{\textbf{mCherry}}$: mCherry intensity of HEK293T cells co-transfected with TALE-Ctrl plus reporter

Construction of the TALE-$(XX')_3$ library

A 102-nt monomer encoding a standard TALE repeat unit was synthesized with the following sequences: 5'-GCTATGGCTACAACCTGTTCGGGGGTCAACCCATGAGCCT GACACAGTACTGGGAGCAGGCGCTGCACGGTTTCCAGTGCCTGCTTACCTCCNNNN NNAGAA-3'. The six random nucleotides (underlined Ns) corresponding to the RVD-encoding region were purposely placed near the 3' end to ensure unbiased oligonucleotide synthesis. This single-stranded DNA was cyclized using linker primer 1 (5'-GAACAGGTT GTAGCCATAGCTTCT-3') and T4 DNA ligase (NEB). Using the agarose gel-purified single stranded circular DNA as template, rolling-circle amplification was conducted with phi29 DNA polymerase (NEB) and primer 1 (30°C, 90 min) followed by primer extension using primer 2 (5'-AGGTTGTAGCCATAGCT-3') and phi29 (30°C, 90 min) to acquire long dsDNA. After ultrasonic shearing (270 W, work 10 s, pause 10 s, 10 cycles) and T4 DNA polymerase (NEB) treatment to blunt the ends of the DNA fragments (12°C, 15 min, with 400 μM dNTP mix), 250-400 bp DNA fragments were harvested by gel purification. These DNA fragments were then cloned into a pre-made entry vector using the ligase-independent cloning (LIC) method to create an entry library. BsmBI digestion of entry

library clones produced ~300 bp DNA fragments that were subsequently cloned, via the Golden Gate approach[1], into a pre-made RVD library vector, which was constructed using the ULtiMATE protocol previously developed by our group[2]. Each plasmid in the final RVD library was verified through sequencing analysis. About 350 kinds of TALE-$(XX')_3$ constructs were obtained from this approach.

Individual construction of TALE-$(XX')_3$

A complementary primer (5'-aaCGTCTCaGTTCGGGGGTCAACCCATGAGCCTGACAC AGTACTGGGAGCAGGCGCTGCACGGTTTCCAGTGCCTGCTT-3') and a specific primer (5'-tCGTCTCaGAACAGGTTGTAGCCATAGCTTCTNNNNNNGGAGGTAAGCAGGCACT GGAA-3'; NNNNNN indicates the RVD codons) were annealed and PCR extended to generate a 102 bp monomer with BsmBI sites at each end. The monomer was ligated via a 6 cycles of the Golden Gate method[1] to generate repeats. The repeat product was PCR amplified using the primers G-lib-F (5'-TAGCTATACGTCTCATTGACCCCCGAACAGGTT GTAGCC-3') and G-lib-R (5'-TAGCTATACGTCTCACCCATGAGCCTGACACAGTACTGG GAGCA-3') and Taq Hifi (Transgen, Inc.). The 3-repeat fragment was gel purified and subsequently cloned into a pre-made RVD library vector via the Golden Gate method[1]. Trans1-T1 competent cells (Transgen, Inc.) were used for bacterial transformation. About 50 kinds of TALE-$(XX')_3$ constructs were obtained from this approach.

Design and construction of TALE-$(XX')_6$ and TALE-$(XX')_{12}$

For TALE-$(XX')_6$, the customized TALEs (17.5-mer) are the same as TALE-$(XX')_3$ except that there are six identical repeats containing the variable RVDs. Accordingly, four reporters were constructed, consisting of TALE-$(XX')_6$ binding sites with six consecutive nucleotides (A, T, C or G) substituted at positions 7 - 12 in front of a minimal CMV promoter and its downstream EGFP gene. The 7^{th} - 9^{th} repeats in TALE-$(XX')_3$ were PCR

amplified using the primers G-lib-seq-F (5'-TCTAGGTACCAAGCCCACGGATTGA-3') and G-lib-seq-R (5'-ATCGATCGTCCGGAGTGAGCCCA-3'). The 300 bp PCR product was purified and further PCR amplified using the primers G-lib-F and G-lib-R. The product of the 6-repeat fragment was gel purified and cloned into a pre-made RVD library vector via the Golden Gate method[1] to construct the final TALE-$(XX')_6$ plasmid.

For TALE-$(XX')_{12}$, the customized TALEs (15.5-mer) contained four ACTC-targeting TALE repeat modules followed by 11.5 repeat units with variable RVDs to be tested. Accordingly, four reporters were constructed, consisting of TALE-$(XX')_{12}$ binding sites with 12 consecutive nucleotides (A, T, C or G) substituted at positions 5 - 16 in front of a minimal CMV promoter and its downstream EGFP gene. TALE-$(XX')_{12}$ plasmids were constructed with a similar method using TALE-$(XX')_6$ as a PCR template and applying the ULtiMATE protocol[2].

Backbone plasmid	10 ng
6-repeat fragment	5 ng
10X Tango buffer (Thermo)	1 µl
10 mM ATP	1 µl
BsmBI (Thermo)	0.75 µl
T4 ligase (NEB)	0.25 µl
ddH_2O	up to 10 µl

37°C	5 min	\
		| 10 cycles
16°C	5 min	/
37°C	5 min	

Cell culture, transfection and flow cytometric analysis

HEK293T cells (from Stanley Cohen lab at Stanford University) were cultured in DMEM medium with 10% FBS and 1% penicillin-streptomycin at 37°C and 5% CO_2. Cells were

seeded 24 h prior to transfection in 24-well plates at a density of 10^5 cells per well. The cells in each well were co-transfected with 0.2 μg TALE-$(XX')_3$ plasmid and 0.4 μg reporter plasmid using polyethylenimine (PEI)[3]. At 48 h post-transfection, the cells were collected and analyzed on a BD LSR II flow cytometer (BD Biosciences). Lasers with wavelengths of 488 nm and 561 nm were used to quantify EGFP and mCherry protein expression, respectively. At least 10,000 events were collected from every sample to obtain sufficient data for analysis. Majority of cells showed mCherry fluorescence intensity of 5 x 10^4 – 5 x 10^5, thus were gated for further analysis. The binding efficiencies and specificities of the variable RVDs (XX') in each customized TALE were assayed by comparing the fold induction of EGFP reporters with the basal level of EGFP in HEK293T cells transfected with the reporter plasmid and a customized TALE plasmid containing unmatched TALE repeats (Supplementary Sequences). The EGFP fluorescence intensity assayed from FACS analysis was normalized to the corresponding mCherry fluorescence intensity.

Generation of a heat map illustrating the base-preference of RVDs

The heat map was generated from library screening of TALE-$(XX')_3$ using four reporters (3A, 3T, 3C, and 3G), thereby reflecting the base preference of 400 RVDs. EGFP activities from different reporters are coded by different colors representing the reporter identities (3A, green; 3T, red; 3C, blue; and 3G, yellow). The brightness of the colors indicates the fold induction of the reporters by TALE-$(XX')_3$ compared to their basal levels.

Criteria for selection of novel RVDs from TALE-$(XX')_3$ for intensive study

The standards are as follows: (1) the highest fold induction among four reporters is at least equal to NK for G recognition; and (2) the second highest fold induction is lower than half of the highest one. In addition to these simple rules, we have also included groups of

RVDs that displayed some unique patterns, i.e. seven RVDs ended with Ala (KA, CA, FA, YA, RA, PA, and AA) that showed preference for T-recognition, and RVDs recognizing both A and G (NN and HN).

Statistical analysis

For comparison between the induction level of TALE-$(XX')_{12}$ and the basal level (Fig. 1f), two-sample, one-tailed t-test was performed assuming unequal variance. **P*<0.05, ***P*<0.01, and ****P*<0.005.

Construction of TALENs

TALENs with natural RVDs (i.e. NI, NG, HD and NN) were constructed using ULtiMATE system as previously described[2]. For TALEN repeats using novel RVDs KN (in place of NN), NH (in place of NN) or RG (in place of NG), TALE monomers containing new RVDs were individually synthesized. The final assembly of these TALENs constructs was conducted using the same ULtiMATE protocol as above.

Assessment of TALENs-mediated indels

HEK293T cells were seeded in 6-well plates at a density of 3 x 10^5 cells per well and incubated at 37°C with 5% CO_2. For each well, a pair of TALEN plasmids and pmaxGFP (Lonza Group Ltd.) were co-transfected at a ratio of 9:9:2 (0.9 μg : 0.9 μg : 0.2 μg) using PEI method. The transfected cells were cultured at 37°C for one day followed by 3 days of incubation at 30°C (cold shock) before flow cytometric sorting for GFP positive cells. TALENs-targeting regions were PCR-amplified from the genome DNA of the isolated GFP positive cells. The TALENs-mediated indels were analyzed by mismatch-sensitive T7 endonuclease I (T7E1; New England Biolabs) as described previously[4].