

ENGINEERED DNA MOLECULE FOR CODING RNA

TECHNICAL FIELD

The present application relates to the field of biotechnology, specifically to an RNA comprising a poly (A) tail. The poly (A) tail makes the replication of the DNA encoding the RNA in prokaryotic system more stable, and can be used to regulate the expression level of the RNA in an eukaryotic cell.

BACKGROUND ART

The primary structure of a translatable mRNA drug molecule consists of a 5' cap structure, a 5' non-coding region (5' UTR), a coding region, a 3' non-coding region, and a polyadenosine tail (poly (A) tail). The known functions of the poly (A) tail include maintaining the *in vivo* stability of mRNA molecules and participating in the initiation of protein translation, the latter of which is achieved through the interaction between the poly (A) tail binding protein (PABP) and the translation initiation complex. In an eukaryotic cell, the poly (A) tail is synthesized by a post-transcriptional modification under the action of typical poly (A) polymerase.

The first step in preparing mRNA drugs *in vitro* is that the mRNA drugs are synthesized through *in vitro* transcription (IVT) by using a linearized plasmid containing the designed product sequence as a template, and the poly (A) tail is usually added downstream of the 3' UTR in a co-transcriptional manner. To achieve co-transcriptional addition of poly (A), the corresponding poly (dA:dT) sequence needs to be included in the template plasmid. However, the poly (dA:dT) repeat sequence in the plasmid is unstable during replication in *E. coli*, with deletion mutations often occur in such sequence, leading to the shortening of poly(dA:dT). This phenomenon is not conducive to the preparation process of *in vitro* transcription template plasmids through large scale fermentation, and the poly (A) truncation has a significant impact on the *in vivo* stability and biological activity of mRNA.

SUMMARY

The present application provides a new poly (A) tail to enhance its conservation during preparation process *in vitro*. Furthermore, a method for regulating the expression level of RNA in an eukaryotic cell based on the poly (A) tail is also provided.

Specifically, the first aspect of the present application provides an engineered DNA molecule capable of being replicated in a cell, comprising a polyadenosine tail (Poly A tail) coding sequence, wherein the poly (A) tail coding sequence comprises:

- a single element **a** and at least one element **b**, and at least one element **c**;
- a single element **a** and at least one element **b**, and at least one element **d**; or

a single element **a** and at least one element **b**, and at least one element **c** and at least one element **d**,

in the poly (A) tail coding sequence, the element **a** consists of a plurality of consecutive adenine (A) nucleotides, and the length range of the element **a** is ≥ 20 nt;

the element **b** consists of a plurality of consecutive A nucleotides, and the length range of the element **b** is $3 \text{ nt} \leq \mathbf{b} < 20 \text{ nt}$;

the element **c** consists of a non-A nucleotide, and the nucleotide is selected from T, C and G nucleotides;

the element **d** consists of any two or more consecutive nucleotides, and the nucleotides are selected from A, T, C and G nucleotides, wherein the nucleotides at the 5' terminus and 3' terminus of the element **d** are not A nucleotides, and the element **d** does not comprise 3 or more consecutive A nucleotides; and the length range of the element **d** is $2 \text{ nt} \leq \mathbf{d} \leq 20 \text{ nt}$;

wherein the element **a** and element **b** are not adjacent, and the element **c** and element **d** are not adjacent, and

the poly (A) tail coding sequence does not comprise any two elements **b** that are adjacent to each other, does not comprise any two elements **c** that are adjacent to each other, and does not comprise any two elements **d** that are adjacent to each other.

In some embodiments, the poly (A) tail coding sequence further comprises a single one element **e**, wherein the element **e** consists of one or two consecutive A, which is located at the 3' terminus of the poly (A) tail coding sequence and is adjacent to the element **d** or element **c**.

In some embodiments, the poly (A) tail coding sequence does not comprise other elements except for element **a**, element **b**, element **c**, and element **d**.

In some embodiments, the poly (A) tail coding sequence does not comprise other elements except for element **a**, element **b**, element **c**, element **d** and element **e**.

In some embodiments, the poly (A) tail coding sequence comprises at least 2 elements **d**. In some embodiments, the poly (A) tail coding sequence comprises at least 2 elements **c**. In some embodiments, the poly (A) tail coding sequence comprises at least one element **d** and one element **c**.

In some embodiments, the number of the element **b** is 2-10, for example, 3, 4, 5, 6, 7, 8, or 9.

In some embodiments, the number of the element **c** is 0-10, for example, 1, 2, 3, 4, 5, 6, 7, 8, or 9.

In some embodiments, the number of the element **d** is 0-5, for example, 1, 2, 3, or 4.

In some embodiments, when element **c** and element **d** are exist simultaneously, the total number of elements **c** and elements **d** is 2-15, for example, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14.

In some embodiments, the number of the element **a** is 1, the number of the element **b** is 3, the number of the element **c** is 2, and the number of the element **d** is 1.

In some embodiments, the number of the element **a** is 1, the number of the element **b** is 4, the

number of the element **c** is 4, and the number of the element **d** is 1.

In some embodiments, the number of the element **a** is 1, the number of the element **b** is 5, the number of the element **c** is 4, and the number of the element **d** is 1.

In some embodiments, the number of the element **a** is 1, the number of the element **b** is 3, the number of the element **c** is 3, and the number of the element **d** is 1.

In some embodiments, the number of the element **a** is 1, the number of the element **b** is 3, the number of the element **c** is 2, and the number of the element **d** is 1.

In some embodiments, the length range of the element **a** is ≤ 80 nt. In some embodiments, the element **a** is 21 nt, 22 nt, 23 nt, 24 nt, 25 nt, 26 nt, 27 nt, 28 nt, 29 nt, 30 nt, 31 nt, 32 nt, 33 nt, 34 nt, 35 nt, 36 nt, 37 nt, 38 nt, 39 nt, 40 nt, 41 nt, 42 nt, 43 nt, 44 nt, 45 nt, 46 nt, 47 nt, 48 nt, 49 nt, 50 nt, 51 nt, 52 nt, 53 nt, 54 nt, 55 nt, 56 nt, 57 nt, 58 nt, 59 nt, 60 nt, 61 nt, 62 nt, 63 nt, 64 nt, 65 nt, 66 nt, 67 nt, 68 nt, 69 nt, 70 nt, 71 nt, 72 nt, 73 nt, 74 nt, 75 nt, 76 nt, 77 nt, 78 nt, or 79 nt.

In some embodiments, the element **b** is 3 nt, 4 nt, 5 nt, 6 nt, 7 nt, 8 nt, 9 nt, 10 nt, 11 nt, 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, or 19 nt in length.

In some embodiments, the length range of the element **d** is $2 \text{ nt} \leq d \leq 20 \text{ nt}$, 3-18 nt, 5-16 nt, 4-10 nt, or 6-12 nt, for example, 2 nt, 3 nt, 4 nt, 5 nt, 6 nt, 7 nt, 8 nt, 9 nt, 10 nt, 11 nt, 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt or 20 nt, preferably 6 nt.

In some embodiments, the length of the poly (A) tail coding sequence is greater than 40 nt, for example, 41 nt, 42 nt, 43 nt, 44 nt, 45 nt, 46 nt, 47 nt, 48 nt, 49 nt, 50 nt, 51 nt, 52 nt, 53 nt, 54 nt, 55 nt, 56 nt, 57 nt, 58 nt, 59 nt, 60 nt, 61 nt, 62 nt, 63 nt, 64 nt, 65 nt, 66 nt, 67 nt, 68 nt, 69 nt, 70 nt, 71 nt, 72 nt, 73 nt, 74 nt, 75 nt, 76 nt, 77 nt, 78 nt, 79 nt, 80 nt, 81 nt, 82 nt, 83 nt, 84 nt, 85 nt, 86 nt, 87 nt, 88 nt, 89 nt, 90 nt, 91 nt, 92 nt, 93 nt, 94 nt, 95 nt, 96 nt, 97 nt, 98 nt, 99 nt, 100 nt, 101 nt, 102 nt, 103 nt, 104 nt, 105 nt, 106 nt, 107 nt, 108 nt, 109 nt, 110 nt, 111 nt, 112 nt, 113 nt, 114 nt, 115 nt, 116 nt, 117 nt, 118 nt, 119 nt, 120 nt, 121 nt, 122 nt, 123 nt, 124 nt, 125 nt, 126 nt, 127 nt, 128 nt, 129 nt, 130 nt, 131 nt, 132 nt, 133 nt, 134 nt, 135 nt, 136 nt, 137 nt, 138 nt, 139 nt, 140 nt, 141 nt, 142 nt, 143 nt, 144 nt, 145 nt, 146 nt, 147 nt, 148 nt, 149 nt, 150 nt, 151 nt, 152 nt, 153 nt, 154 nt, 155 nt, 156 nt, 157 nt, 158 nt, 159 nt, 160 nt, 161 nt, 162 nt, 163 nt, 164 nt, 165 nt, 166 nt, 167 nt, 168 nt, 169 nt, 170 nt, 171 nt, 172 nt, 173 nt, 174 nt, 175 nt, 176 nt, 177 nt, 178 nt, 179 nt, 180 nt, 181 nt, 182 nt, 183 nt, 184 nt, 185 nt, 186 nt, 187 nt, 188 nt, 189 nt, 190 nt, 191 nt, 192 nt, 193 nt, 194 nt, 195 nt, 196 nt, 197 nt, 198 nt, 199 nt, 200 nt, 201 nt, 202 nt, 203 nt, 204 nt, 205 nt, 206 nt, 207 nt, 208 nt, 209 nt, 210 nt, 211 nt, 212 nt, 213 nt, 214 nt, 215 nt, 216 nt, 217 nt, 218 nt, 219 nt, 220 nt, 221 nt, 222 nt, 223 nt, 224 nt, 225 nt, 226 nt, 227 nt, 228 nt, 229 nt, 230 nt, 231 nt, 232 nt, 233 nt, 234 nt, 235 nt, 236 nt, 237 nt, 238 nt, 239 nt, 240 nt, 241 nt, 242 nt, 243 nt, 244 nt, 245 nt, 246 nt, 247 nt, 248 nt, 249 nt, 250 nt, 251 nt, 252 nt, 253 nt, 254 nt, 255 nt, 256 nt, 257 nt, 258 nt, 259 nt, 260 nt, 261 nt, 262 nt, 263 nt, 264 nt, 265 nt, 266 nt, 267 nt, 268 nt, 269 nt, 270 nt, 271 nt, 272 nt, 273 nt, 274 nt, 275 nt, 276 nt, 277 nt, 278 nt, 279 nt, 280 nt, 281 nt, 282 nt, 283 nt, 284 nt, 285 nt, 286 nt, 287 nt, 288 nt, 289 nt, 290 nt, 291 nt, 292 nt, 293 nt, 294 nt, 295 nt, 296 nt,

297 nt, 298 nt, 299 nt, 300 nt, 301 nt, 302 nt, 303 nt, 304 nt, 305 nt, 306 nt, 307 nt, 308 nt, 309 nt, 310 nt, 311 nt, 312 nt, 313 nt, 314 nt, 315 nt, 316 nt, 317 nt, 318 nt, 319 nt, 320 nt, 321 nt, 322 nt, 323 nt, 324 nt, 325 nt, 326 nt, 327 nt, 328 nt, 329 nt, 330 nt, 331 nt, 332 nt, 333 nt, 334 nt, 335 nt, 336 nt, 337 nt, 338 nt, 339 nt, 340 nt, 341 nt, 342 nt, 343 nt, 344 nt, 345 nt, 346 nt, 347 nt, 348 nt, 349 nt, 350 nt, 351 nt, 352 nt, 353 nt, 354 nt, 355 nt, 356 nt, 357 nt, 358 nt, 359 nt, 360 nt, 361 nt, 362 nt, 363 nt, 364 nt, 365 nt, 366 nt, 367 nt, 368 nt, 369 nt, 370 nt, 371 nt, 372 nt, 373 nt, 374 nt, 375 nt, 376 nt, 377 nt, 378 nt, 379 nt, 380 nt, 381 nt, 382 nt, 383 nt, 384 nt, 385 nt, 386 nt, 387 nt, 388 nt, 389 nt, 390 nt, 391 nt, 392 nt, 393 nt, 394 nt, 395 nt, 396 nt, 397 nt, 398 nt, 399 nt, or 400 nt, etc.

In some embodiments, 50% or more of the polynucleotides of element **a** are located in the 5' portion or 3' portion of the poly (A) tail coding sequence. In some embodiments, 50% or more of the polynucleotides of element **a** are located in the 5' portion of the poly (A) tail coding sequence. In some embodiments, 50% or more of the polynucleotides of element **a** are located in the 3' portion of the poly (A) tail coding sequence. In some embodiments, the number of nucleotides of element **a** located at the 3' portion of the poly (A) tail coding sequence is equal to the number of nucleotides located at the 5' portion of the poly (A) tail coding sequence.

In some embodiments, the element **c** is G, C or T.

In some embodiments, the element **d** comprises a palindromic sequence. The element **d** is a palindromic sequence. In some embodiments, the element **d** comprises a sequence selected from the group consisting of GATATC (SEQ ID NO: 15), GTATAC (SEQ ID NO: 16), GAATCT (SEQ ID NO: 17), GCATATGACT (SEQ ID NO: 18), and GATATCGTATAC (SEQ ID NO: 19). In some embodiments, the element **d** is a sequence selected from the group consisting of GATATC (SEQ ID NO: 15), GTATAC (SEQ ID NO: 16), GAATCT (SEQ ID NO: 17), GCATATGACT (SEQ ID NO: 18), and GATATCGTATAC (SEQ ID NO: 19). In some embodiments, the element **d** comprises a polynucleotide sequence represented by SEQ ID NO: 15. In some embodiments, the polynucleotide sequence of element **d** is represented by SEQ ID NO: 15.

In some embodiments, the nucleotide at the 3' terminus of the poly (A) tail coding sequence is A. In some embodiments, the nucleotide at the 3' terminus of the poly (A) tail coding sequence is G. In some embodiments, the nucleotide at the 3' terminus of the poly (A) tail coding sequence is C. In some embodiments, the nucleotide at the 3' terminus of the poly (A) tail coding sequence is T.

In some embodiments, the 3' portion of the poly (A) tail coding sequence comprises one or more non-A nucleotides. In some embodiments, 1/2 portion of the poly (A) tail coding sequence near the 3' terminus comprises one or more non-A nucleotides. In some embodiments, 1/3 portion of the poly (A) tail coding sequence near the 3' terminus comprises one or more non-A nucleotides. In some embodiments, 1/4 portion of the poly (A) tail coding sequence near the 3' terminus comprises one or more non-A nucleotides.

In some embodiments, the structure of the poly (A) tail coding sequence is:

element **a**-element **c**-element **b**-element **c**-element **b**-element **c**-element **b**-element **c**-element **b**;

element **b**-element **c**-element **b**-element **c**-element **a**-element **d**-element **b**-element **c**-element **b**-element **c**-element **b**;

element **b**-element **c**-element **b**-element **c**-element **b**-element **d**-element **a**-element **c**;

element **a**-element **d**-element **b**-element **c**-element **b**-element **c**-element **b**; or

element **b**-element **c**-element **b**-element **c**-element **b**-element **d**-element **a**.

In some embodiments, the structure of the poly (A) tail coding sequence is:

element **a**-element **c**-element **b**-element **c**-element **b**-element **c**-element **b**-element **c**-element **b**;
specifically, 60A-G-19A-G-19A-G-19A-G-3A.

In some embodiments, the structure of the poly (A) tail coding sequence is:

7A-C-18A-G-60A-GG-7A-C-18A-G-14A,

19A-G-19A-G-19A-element **d**-60A-G,

60A-element **d**-19A-G-19A-G-17A,

19A-G-19A-G-19A-element **d**-60A,

19A-G-19A-G-19A-element **d**-60A,

19A-C-19A-C-19A-element **d**-60A,

19A-T-19A-T-19A-element **d**-60A,

19A-G-19A-G-19A-element **d**-60A, or

19A-G-19A-G-19A-element **d**-60A; and

wherein the element **d** consists of 6 or 12 nucleotides.

In some embodiments, the poly (A) tail coding sequence is represented by any one of SEQ ID NOs: 1-10.

In some embodiments, the poly (A) tail coding sequence is represented by SEQ ID NO: 3 or SEQ ID NO: 4.

In some embodiments, the engineered DNA molecule is further connected to a gene of interest fragment on the 5' end side of its poly (A) tail coding sequence, wherein the gene of interest fragment and the poly (A) tail coding sequence co-encode RNA. In some embodiments, the engineered DNA molecule is further connected to a gene of interest fragment on the 5' end side of its poly (A) tail coding sequence, wherein the gene of interest fragment and the poly (A) tail coding sequence co-encode mRNA. In some embodiments, the gene of interest fragment comprises a protein coding sequence or a non-protein coding sequence, such as a functional RNA coding sequence. In some embodiments, the gene of interest fragment further comprises a 5'UTR coding sequence on the 5' end side of the protein coding sequence or the functional RNA coding sequence. In some embodiments, the gene of interest fragment further comprises a 3' UTR coding sequence on the 3' end side of the protein coding sequence or the functional RNA coding sequence. In some embodiments, the gene of interest fragment further comprises a 3' UTR coding sequence on the 3'

end side of the protein coding sequence or the functional RNA coding sequence, and further comprises a 5' UTR coding sequence on the 5' end side of the protein coding sequence or the functional RNA coding sequence. In some embodiments, the engineered DNA molecule further comprises a replicon, for example, an origin of replication, such as an ORI. In some embodiments, the engineered DNA molecule further comprises a label gene in order to facilitate screening of cells comprising the engineered DNA molecule, and the label gene is selected from, for example, antibiotic resistance gene, fluorescent protein, and the like. In some embodiments, the DNA molecule further comprises a promoter, which initiates the transcription of RNA co-encoded by the gene of interest fragment and the poly (A) tail coding sequence. In some embodiments, the promoter is a prokaryotic promoter. In some embodiments, the promoter is an eukaryotic promoter. In some embodiments, the DNA molecule further comprises a replicon, for example, an origin of replication, a promoter, a 5' UTR coding sequence, a protein coding sequence, and a 3' UTR coding sequence. In some embodiments, the DNA molecule further comprises a replicon, for example, an origin of replication, a resistance gene, a 5' UTR coding sequence, a protein coding sequence, and a 3' UTR coding sequence. In some embodiments, the DNA molecule further comprises a replicon, for example, an origin of replication, a resistance gene, a promoter, a 5' UTR coding sequence, a protein coding sequence, and a 3' UTR coding sequence. In some embodiments, the protein coding sequence encodes an HPV viral antigen protein. In some embodiments, the HPV protein is derived from HPV type 16 and/or type 18. In some embodiments, the protein coding sequence encodes the E2, E6 or E7 protein of HPV. In some embodiments, the protein coding sequence encodes a fusion protein of the E6 and E7 proteins of HPV. In some embodiments, the protein coding sequence encodes a fusion protein of the E2, E6 and E7 proteins of HPV. In some embodiments, the polypeptide fragments of the fusion protein are derived from HPV type 16 and/or type 18. In some embodiments, the polypeptide fragments of the fusion protein are derived from E2, E6 and E7 proteins of HPV type 16 and/or type 18. In some embodiments, the protein coding sequence encodes a polypeptide represented by SEQ NO: 26 or a conservatively substituted variant thereof.

In some embodiments, the engineered DNA molecule comprises a polynucleotide sequence represented by any one of SEQ ID NOs: 22-25 or a synonymous mutant thereof, or a polynucleotide sequence sharing 85% or more sequence identity with the polynucleotide sequence represented by any one of SEQ ID NOs: 22-25 or a synonymous mutant thereof.

In some embodiments, the DNA molecule is a DNA plasmid. In some embodiments, the DNA molecule is a linear plasmid or a circular plasmid. In some embodiments, the DNA molecule is single-stranded or double-stranded. In some embodiments, the plasmid is a pUC-, pTZ-, pMB1-, or pCoIE1-based plasmid. In some embodiments, the plasmid is a pUC57 vector-based plasmid.

In some embodiments, the cell is a prokaryotic cell. In some embodiments, the cell is a *recA*-bacterium. In some embodiments, the cell is *Escherichia coli*. In some embodiments, the *E. coli* is selected from the group consisting of K-12 and derivative strains thereof, and B strain and

derivative strains thereof. In some embodiments, the *E. coli* is selected from the group consisting of MG1655, DH5 or DH5 α , DH10B, BL21, DB3.1, HB101, JM109, JM110, MC1061, MG1655, Pir1, Stbl2, Stbl3, Top10, XL1 Blue, XL10 Gold, BLR, HMS174, Tuner, Rostetta2, Lemo21, T7Express, and Origami2.

The second aspect of the present application discloses a cell comprising the DNA molecule of the first aspect. In some embodiments, the DNA molecule of the aforementioned first aspect can be replicated and/or transcribed in the cell. In some embodiments, the cell is a prokaryotic cell, and the DNA molecule of the aforementioned first aspect can be replicated in the prokaryotic cell. In some embodiments, the cell is a recA-bacterium. In some embodiments, the cell is *E. coli*. In some embodiments, the prokaryotic cell is a competent cell. In some embodiments, the prokaryotic cell is an engineered cell. In some embodiments, the prokaryotic cell is an engineered prokaryotic cell. In some embodiments, the cell is *E. coli*, and the *E. coli* is selected from the group consisting of K-12 and derivatives thereof, and B strain and derivatives thereof. In some embodiments, the *E. coli* is selected from the group consisting of MG1655, DH5 or DH5 α , DH10B, BL21, DB3.1, HB101, JM109, JM110, MC1061, MG1655, Pir1, Stbl2, Stbl3, Top10, XL1 Blue, XL10 Gold, BLR, HMS174, Tuner, Rostetta2, Lemo21, T7 Express, and Origami2. In some embodiments, the cell is an eukaryotic cell, and the DNA molecule of the aforementioned first aspect can be transcribed in the eukaryotic cell. In some embodiments, the eukaryotic cell is a mammalian cell. In some embodiments, the eukaryotic cell is selected from the group consisting of a yeast or a mold.

In a third aspect of the present application, a poly (A) tail is provided. The poly (A) tail is:

- (1) obtained by transcription of the engineered DNA molecule according to the first aspect;
- (2) obtained by chemical synthesis, and has the same polynucleotide sequence as the poly (A) tail obtained by transcription of the engineered DNA molecule according to the first aspect; or
- (3) obtained by further modification of the poly (A) tail in the above item (1) or (2).

In some embodiments, the further modification comprises substituting one or more ribonucleotides in the poly (A) tail obtained by the above item (1) or (2) with one or more deoxyribonucleotides. In some embodiments, the one or more ribonucleotides are substituted with deoxyribonucleotides to which they correspond, for example, one or more ribonucleotides A of the poly (A) tail are substituted with deoxyribonucleotides A, one or more ribonucleotides U are substituted with deoxyribonucleotides T, one or more ribonucleotides C are substituted with deoxyribonucleotides C, one or more ribonucleotides G are substituted with deoxyribonucleotides G, or one or more ribonucleotides G are substituted with ribonucleotides I (inosine) or deoxyribonucleotides I. In some embodiments, the modification is a chemical modification. In some embodiments, the modification is base editing. In some embodiments, the modification is deamination treatment of one or more ribonucleotides in the poly (A) tail obtained by the above item (1) or (2).

In some embodiments, the poly (A) tail comprises a polynucleotide sequence selected from

any one of SEQ ID NOs: 1-10. In some embodiments, the polynucleotide sequence of the poly (A) tail is represented by any one of SEQ ID NOs: 1-10.

The present application also provides use of the aforementioned poly (A) tail. In some embodiments, a use of the poly (A) tail for making an RNA molecule more stable is provided, wherein the poly (A) tail is located at the 3' terminus of the RNA, and the more stable means being more stable outside a cell, inside a cell, or in an animal body relative to an RNA molecule comprising an another poly (A) tail. In some embodiments, a use of the poly (A) tail for reducing the stability of an RNA molecule is provided, wherein the poly (A) tail is located at the 3' terminus of the RNA, and the reducing the stability refers to a reduced stability outside a cell, inside a cell, or in an animal body relative to an RNA molecule comprising an another poly (A) tail. In some embodiments, a use of the poly (A) tail for increasing the expression level of an RNA molecule in the same period is provided, wherein the increasing the expression level refers to an increased expression level outside a cell, inside a cell, or in an animal body relative to an RNA molecule comprising an another poly (A) tail. In some embodiments, a use of the poly (A) tail for reducing the expression level of an RNA molecule within the same period is provided, wherein the reducing the expression level refers to a reduced expression level outside a cell, inside a cell, or in an animal body relative to an RNA molecule comprising an another poly (A) tail. In some embodiments, a use of the poly (A) tail for prolonging the expression time of an RNA molecule is provided, wherein the prolonging the expression time refers to a prolonged expression time outside a cell, inside a cell, or in an animal body relative to an RNA molecule comprising an another poly (A) tail. In some embodiments, a use of the poly (A) tail for shortening the expression time of an RNA molecule is provided, wherein the shortening the expression time refers to a shortened expression time outside a cell, inside a cell, or in an animal body relative to an RNA molecule comprising an another poly (A) tail. In some embodiments, a use of the poly (A) tail for extending the half-life of an RNA molecule is provided, wherein the extending the half-life refers to an extended half-life outside a cell, inside a cell, or in an animal body relative to an RNA molecule comprising an another poly (A) tail. In some embodiments, a use of the poly (A) tail for shortening the half-life of an RNA molecule is provided, wherein the shortening the half-life refers to a shortened half-life outside a cell, inside a cell, or in an animal body relative to an RNA molecule comprising an another poly (A) tail.

In some embodiments, the RNA molecule is an mRNA molecule. In some embodiments, the poly (A) tail and the other poly (A) tail are two different poly (A) tails that belong to the poly (A) tail according to the third aspect of the present application. In some embodiments, the poly (A) tail is the poly (A) tail according to the third aspect of the present application, and the other poly (A) tail is a poly (A) tail other than the poly (A) tail according to the third aspect of the present application. In some embodiments, the inside a cell means inside a host cell, and the host cell is an eukaryotic cell. In some embodiments, the host cell is a mammalian cell. In some embodiments, the host cell is a human cell.

The present application also provides a DNA fragment or a hybrid molecule fragment of DNA and RNA for encoding the poly (A) tail according to the third aspect of the present application, and a use of the DNA fragment or the hybrid molecule fragment of DNA and RNA for making the replication of the DNA molecule or the hybrid molecule of DNA and RNA for encoding an RNA more conservative in a host cell, and in said use, wherein the DNA fragment or the hybrid molecule fragment of DNA and RNA for encoding the poly (A) tail according to the third aspect of the present application is located on the 3' end side of the RNA coding sequence in the DNA molecule or the hybrid molecule of DNA and RNA. In some embodiments, the host cell is a prokaryotic cell. In some embodiments, the host cell is a *recA*-bacterium. In some embodiments, the host cell is *E. coli*. In some embodiments, the *E. coli* is selected from the group consisting of K-12 and derivative strains thereof, and B strain and derivative strains thereof. In some embodiments, the *E. coli* is selected from the group consisting of MG1655, DH5 or DH5 α , DH10B, BL21, DB3.1, HB101, JM109, JM110, MC1061, MG1655, Pir1, Stbl2, Stbl3, Top10, XL1Blue, XL10Gold, BLR, HMS174, Tuner, Rostetta2, Lemo21, T7Express, and Origami2.

The fourth aspect of the present application also provides an RNA molecule, wherein it comprises the poly (A) tail according to the third aspect. In some embodiments, the RNA molecule is an mRNA molecule. In some embodiments, the RNA molecule is:

- (1) obtained by transcription of the engineered DNA molecule according to the first aspect;
- (2) obtained by chemical synthesis, and has the same polynucleotide sequence as the RNA molecule of the above item (1); or
- (3) obtained by further modification of the RNA molecule in above item (1) or (2).

In some embodiments, the further modification comprises substituting one or more ribonucleotides in the RNA molecule obtained by the above (1) or (2) with one or more deoxyribonucleotides. In some embodiments, the one or more ribonucleotides are substituted with deoxyribonucleotides to which they correspond, for example, one or more ribonucleotides A of the RNA molecule are substituted with deoxyribonucleotides A, one or more ribonucleotides U are substituted with deoxyribonucleotides T, one or more ribonucleotides C are substituted with deoxyribonucleotides C, one or more ribonucleotides G are substituted with deoxyribonucleotides G, or one or more ribonucleotides G are substituted with ribonucleotides I (inosine) or deoxyribonucleotides I. In some embodiments, the modification is a chemical modification. In some embodiments, the modification is base editing. In some embodiments, the modification is deamination treatment of one or more ribonucleotides in the RNA molecule obtained by the above (1) or (2). In some embodiments, the further modification is a post-transcriptional modification. In some embodiments, the further modification comprises capping treatment. In some embodiments, the further modification comprises splicing. In some embodiments, the further modification comprises splicing and capping treatment.

In some embodiments, the RNA molecule comprises coding RNA or non-coding RNA

(ncRNA). In some embodiments, the RNA molecule is a pre-mRNA. In some embodiments, the RNA is a mature mRNA. In some embodiments, the RNA molecule is a long non-coding RNA (lncRNA). In some embodiments, the RNA molecule further comprises a 5'-cap structure. In some embodiments, the polynucleotide sequence of the RNA molecule is represented by any one of SEQ ID NOs: 22-25. In some embodiments, the RNA molecule comprises a polynucleotide sequence represented by any one of SEQ ID NOs: 22-25.

In addition, the present application also provides a hybrid molecule of DNA and RNA, which carries the same genetic information as the engineered DNA molecule according to the first aspect, the same genetic information as the poly (A) tail according to the third aspect, or the same genetic information as the RNA molecule according to the fourth aspect.

In addition, the present application also provides a nucleic acid molecule library. In some embodiments, the nucleic acid molecule library comprises the engineered DNA molecule according to the first aspect, the poly (A) tail according to the third aspect, the DNA fragment or the hybrid molecule fragment of DNA and RNA for encoding the poly (A) tail according to the third aspect of the present application, or the RNA molecule according to the fourth aspect.

In addition, the present application also provides a method for regulating protein expression, which comprises introducing a plurality of nucleic acid molecules in the aforementioned nucleic acid molecule library into cells of interest at different times and/or in different quantity ratios. In some embodiments, the nucleic acid molecule is the engineered DNA molecule according to the aforementioned first aspect. In some embodiments, the nucleic acid molecule is the RNA molecule according to the aforementioned fourth aspect.

It should be understood that aspects and embodiments according to the present application described herein include the aspects and embodiments that "comprise/comprising", "consist/consisting of", and "essentially consist/consisting of" the same. The preferred embodiments of the present application are described in detail above; however, the present application is not limited thereto. Within the technical concept of the present application, a variety of simple modifications can be made to the technical solutions of the present application, including combinations of various technical features in any other suitable manner. These simple modifications and combinations should also be regarded as the contents disclosed in the present application, and belong to the protection scope of the present application.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the replication stability of 10 poly (A)s in the present application in *E. coli* DH5 α in Example 2;

Fig. 2 shows the base deletion statistics of 10 poly (A)s in the present application in *E. coli* DH5 α in Example 2;

Fig. 3 shows the replication stability of poly (A) P1, P2, P3, P4 and poly (A) controls C1 and

C2 in *E. coli* DH5 α under a plasmid system comprising an HPV antigen sequences in Example 3;

Fig. 4 shows the base deletion statistics of poly (A) P3, P4 and poly (A) controls C1 and C2 in *E. coli* DH5 α under a plasmid system comprising an HPV antigen sequence in Example 3;

Fig. 5 shows the comparison of the replication stability between poly (A) P1, P2, P3 and poly (A) control C1 under two temperature conditions of 30°C and 37°C in *E. coli* DH5 α in a plasmid system comprising an HPV antigen sequence in Example 3;

Fig. 6 shows an example of a universal vector plasmid DNA profile in the Examples.

Fig. 7 shows the animal imaging results of the expression levels of luciferase with poly (A) P3, P4, P5, P8, P9 and control C2 in mice in Example 4;

Fig. 8 shows the quantitative results of fluorescence intensity after animal imaging of the expression levels of luciferase with poly (A) P3, P4, P5, P8, P9 and control C2 in mice in Example 4 (ns, no significant difference; **, significant difference, $p < 0.01$).

DETAILED DESCRIPTION

The present application firstly provides a method for stably amplifying a poly (A) tail transcription template DNA in vitro, so as to reduce the mutation frequency of the poly (A) tail transcription template sequence when the DNA is replicated in large quantities in a cell. Thus, a large amount of RNA comprising a poly (A) tail with a defined sequence is obtained based on the DNA. On this basis, RNA with a poly (A) tail that is engineered to have a specific function, such as a mRNA, can be produced on a large scale through in vitro fermentation.

In addition, the present application also provides a DNA comprising a poly (A) tail transcription template that can be stably amplified in vitro, and an RNA transcribed from the DNA. Furthermore, under the premise of satisfying in vitro stable amplification, the present application further provides a group of poly (A) tails having different regulatory effects on RNA stability and/or expression efficiency, an RNA comprising the poly (A) tail, a DNA comprising the poly (A) tail coding sequence, and a library consisting of the poly (A) tail, the RNA or the DNA.

Furthermore, the present application also provides uses of the aforementioned poly (A) tail, RNA, DNA, and library.

Terms

As used herein, “element **a**”, “element **b**”, “element **c**”, “element **d**”, and “element **e**” are types of elements comprised in poly (A). The element **a** consists of a plurality of consecutive adenine (A) nucleotides, and the length range of element **a** is ≥ 20 nt; the element **b** consists of a plurality of consecutive A nucleotides, and the length range of element **b** is $3 \text{ nt} \leq b < 20 \text{ nt}$; the element **c** consists of a non-A nucleotide, and the nucleotide is selected from T, C and G nucleotides; the element **d** consists of any two or more consecutive nucleotides, and the nucleotides are selected from A, T, C and G nucleotides, wherein the nucleotides at the 5' terminus and 3' terminus of element **d** are not A nucleotides, and the element **d** does not comprise 3 or more consecutive A

nucleotides, and the length range of element **d** is $2 \text{ nt} \leq d \leq 20 \text{ nt}$. The element **e** consists of one or two consecutive A, and the element **e** is located at the 3' end of the poly (A) tail coding sequence and is adjacent to the element **d** or the element **c** when it exists. When the poly (A) comprises two or more "element **b**", "element **c**", and "element **d**", the sequences of each two elements **b** may be the same or different, the sequences of each two elements **c** may be the same or different, and the sequences of each two elements **d** may be the same or different, as long as they each meet the above definitions of elements **a**, **b**, **c** and **d**. In the present application, the "element **a**", "element **b**", "element **c**", "element **d**", "element **e**", etc. in the poly (A) tail may be indicated by the term "element".

As used herein, when describing the positional relationship of two or more elements as "non-adjacent", it means that the two or more elements are not adjacent to each other. In other words, the two or more elements at least comprise one or more other nucleotides or bases other than the nucleotides of the two elements between each two elements.

As used herein, "encoding\coding" means i) a DNA sequence comprising genetic information that can be transcribed into an RNA molecule, and/or ii) an RNA molecule comprising genetic information that can be translated into an amino acid sequence. As used herein, therefore, the "coding sequence" refers to a ribonucleotide (RNA) sequence or a fragment thereof in a mRNA precursor or a mature mRNA that can be translated into a protein, and also refers to a complementary sequence or a fragment thereof of a deoxyribonucleotide (DNA) sequence that serves as a template for transcribing the mRNA precursor or mature mRNA. In addition, the "coding sequence" of the present application may further comprise polynucleotide sequences encoding proteins, functional nucleic acids, or fragments thereof, such as miRNA, shRNA, dsRNA, guide RNA, poly (A) tail, 5' UTR, 3' UTR, etc. Among them, a DNA molecule comprising genetic information that can be transcribed into an RNA molecule is called the "coding nucleic acid" of the RNA molecule; and an RNA molecule comprising genetic information that can be translated into an amino acid sequence is called the "coding nucleic acid" of the amino acid sequence.

In the present application, the nucleotides in all polynucleotide sequences are numbered from the 5' end to the 3' end, that is, the nucleotide at the 5' terminus is the first nucleotide, and the nucleotide at the 3' terminus is the last nucleotide. Unless otherwise specified, "5' end" and "5' terminus" can be used interchangeably; "3' end" and "3' terminus" can be used interchangeably. "5' end" and "3' end" focus on describing the relative position relationship between nucleotides, between nucleotide sequence segments, or between nucleotides and nucleotide sequence segments in the same nucleic acid sequence; "5' terminus" and "3' terminus" are used to describe the positions of the first and last nucleotides of a nucleic acid sequence or a segment of a nucleic acid sequence, respectively. "5' end side" is used to describe the relative position relationship between two non-overlapping sequences in the same polynucleotide sequence; when describing a sequence being located on the 5' end side of another sequence, it means that the sequence is more closer to

the "5' end" of the polynucleotide sequence compared to another sequence. Similarly, when describing that a sequence is located on the 3' end side of another sequence, it means that the sequence is closer to the "3' end" of the polynucleotide sequence compared to another sequence, moreover, the sequence and another sequence do not comprise any overlapping parts each other. Specifically, for example, "the DNA coding sequence of the poly (A) tail is located at the 3' end of the RNA coding sequence" means that the DNA coding sequence of the poly (A) tail, as the component of the RNA coding sequence, comprises the nucleotides at the 3' terminus of the RNA coding sequence. In addition, as used herein, the "5' portion" refers to the one that near the half of the 5' terminus of the polynucleotide sequence, bounded by the "center position" of the polynucleotide sequence. The "3' portion" refers to the one that near the half of the 3' terminus of the polynucleotide sequence, bounded by the center position of the polynucleotide sequence. The number of nucleotides from the "central position" as described in this application to the 5' end is equal to the number of nucleotides from the "central position" as described in this application to the 3' end.

As used herein, when referring to the replication of a nucleic acid molecule, the term "conservative" means a low probability of mutation during the replication process. In this context, the "conservative" is a relative concept. For example, when describing that "the DNA coding sequence of the poly (A) tail is used to make the replication of the DNA molecule coding RNA more conservative in the host cell", it refers to the probability that the parent DNA molecule encoding RNA replicates into an offspring DNA molecule. If the RNA molecule comprises the Poly (A) tail, the offspring DNA molecule has a higher probability of 100% sequence identity with the parent DNA molecule compared to the coding DNA of an RNA molecule that does not comprise the Poly (A) tail (such as an RNA molecule comprising some other Poly (A) tail); the average sequence identity between multiple offspring DNA molecules obtained by replication of the parent DNA molecule and the parent DNA molecule is higher.

In the present application, when describing "regulating" the expression of an RNA molecule, the "regulating" means increasing or decreasing the total amount of the protein or functional RNA expressed by the RNA molecule within the same time period; or enabling the RNA to express the protein or functional RNA within a longer or shorter time period, and the increased or decreased, or longer or shorter time period is compared to another RNA molecule expressing the same protein or functional RNA. When describing "regulating" protein expression, it means regulating the expression of RNA molecules comprising the protein coding sequence. The regulatory effect described herein can be achieved by connecting the poly (A) tail of the present application to the 3' terminus of an RNA molecule that does not comprise a poly (A) tail, or by substituting the original poly (A) tail of the RNA with the poly (A) tail of the present application.

As used herein, the percentage of "identity", for example 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 98.5%, 99%, or 99.5% identity, refers to the degree of similarity between

amino acid sequences or between nucleotide sequences, as determined by sequence alignment, which is 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 98.5%, 99%, or 99.5%. For example, after two sequences have identical residues at as many positions as possible by introducing gaps, the ratio of the number of positions occupied by the identical bases or amino acid residues to the total number of positions is determined. The percentage of "identity" can be determined by using software programs known in the art. Preferably the alignment is performed by using default parameters. A preferred alignment program is BLAST. Preferred programs are BLASTN and BLASTP. Details of these programs are available at the following: ncbi.nlm.nih.gov/cgi-bin/BLAST.

As used herein, nucleic acid "complementarity" refers to the ability of one nucleic acid to form hydrogen bonds with another nucleic acid through traditional Watson-Crick base-pairing. Percent complementarity refers to the percentage of residues in a nucleic acid molecule that can form hydrogen bonds (i.e., Watson-Crick base-pairing) with another nucleic acid molecule (e.g., about 5, 6, 7, 8, 9, 10 out of 10 correspond to about 50%, 60%, 70%, 80%, 90%, and 100% complementarity, respectively). "Completely complementary" means that all consecutive residues of a nucleic acid sequence will form hydrogen bond with the same number of consecutive residues in a second nucleic acid sequence. As used herein, "substantially complementary" refers to a degree of complementarity of any one of at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% over a region of about 40, 50, 60, 70, 80, 100, 150, 200, 250 or more nucleotides, or to two nucleic acids that hybridize under stringent conditions. For a single base or a single nucleotide, according to the Watson-Crick base-pairing principle, when A is paired with T or U, and C is paired with G or I, it is called complementary or matching, and vice versa; any other base-pairing is called non-complementary. In the present application, the "complementary polynucleotide sequence" of a certain polynucleotide sequence refers to a polynucleotide sequence that is completely complementary to the certain polynucleotide sequence.

As used herein, a "conservative substitution variant" of a protein, polypeptide or amino acid sequence refers to one in which one or more amino acid residues are subjected to amino acid substitution without altering the overall conformation and function of the protein or enzyme, including but not limited to substituting the amino acids in the amino acid sequence of the parent protein in the manner described by the aforementioned "conservative substitution". Therefore, the similarity between two proteins or amino acid sequences with similar functions may be different. For example, 70-99% similarity (identity) based on the MEGALIGN algorithm. The "conservative substitution variants" also comprise polypeptides or enzymes with 60% or more amino acid identity as determined by BLAST or FASTA algorithms, preferably 75% or more, more preferably 85% or more, and most preferably even 90% or more, and have the same or substantially similar properties or functions as the native or parent protein or enzyme.

In the context of this application, the terms "DNA" and "RNA" refer to single-stranded or

double-stranded DNA or RNA molecules. Unless otherwise indicated, the terms "DNA" and "DNA molecule" refer to double-stranded DNA molecules composed of A, C, G and/or T nucleotides, and the terms "RNA" and "RNA molecule" refer to single-stranded RNA molecules composed of A, C, G and/or U nucleotides. Herein, the A, C, G, T and U nucleotides refer to nucleotides comprising adenine, guanine, cytosine, thymine and uracil as respective nitrogenous bases.

RNA molecules comprise coding RNA or non-coding RNA (ncRNA), such as pre-mRNA, mature mRNA, or long non-coding RNA (lncRNA).

As used herein, the "hybrid molecule of DNA and RNA" is a molecule comprising a polynucleotide sequence consisting of deoxyribonucleotides and ribonucleotides. The hybrid molecule of DNA and RNA can be obtained by the following method:

substituting one or more deoxyribonucleotides in the DNA with ribonucleotides;

substituting one or more ribonucleotides in the RNA with deoxyribonucleotides; or

de novo synthesizing by using deoxyribonucleotides and ribonucleotides as raw materials through biological or chemical synthesis. It should be noted that the method of obtaining a hybrid molecule of DNA and RNA is not limited to the above method, and a hybrid molecule of DNA and RNA obtained by any method belong to the category of "a hybrid molecule of DNA and RNA" defined in this application.

As used herein, when describing two nucleic acid molecules as having "the same genetic information", it means that the two nucleic acid molecules are complementary, or comprise exactly the same base sequence, or one nucleic acid molecule having exactly the same base sequence as another nucleic acid molecule can be obtained by converting one or more thymines in the base sequence of one nucleic acid molecule into uracil. Therefore, any two of the DNA, RNA, and a hybrid molecule of DNA and RNA can have the same genetic information. Herein, the term "base sequence" refers to the order of arrangement of bases in a polynucleotide molecule. Unless otherwise specified, those skilled in the art should be understood that the thymine may be indicated by "T" when the base sequence or polynucleotide sequence described in the present application are used to describe a DNA sequence, while the "T" will be substituted by "U" (uracil) when the base sequence or polynucleotide sequence are used to describe RNA (such as mRNA). Therefore, whenever a DNA is disclosed by a specific sequence number (SEQ ID NO) herein, an RNA (e.g. mRNA or poly(A) tail) sequence complementary or corresponding to the DNA is also disclosed, wherein each "T" in the DNA sequence is substituted by a "U".

Poly (A) Tail and Uses thereof

As used herein, the term "Poly A tail" or "Poly (A) sequence" refers to an uninterrupted or non-interrupted sequence of adenylate residues typically located at the 3'-terminus of an RNA molecule. In RNA, in the presence of a 3'-UTR, the Poly-A sequence is linked to the 3' end of the 3'-UTR. An uninterrupted Poly-A tail is characterized by consecutive adenylate residues. The Poly-A tail can be of any length. In some embodiments, the Poly-A tail comprises, or consists of, at

least 20, at least 30, at least 40, at least 80 or at least 100, and at most 500, at most 400, at most 300, at most 200 or at most 150 adenylate nucleotides (A), in particular about 120 A. Typically, the vast majority of nucleotides in the Poly-A tail are adenosines, wherein the vast majority refers to at least 75%, at least 80%, at least 85%, at least 90% of the nucleotides, etc., but the remaining nucleotides are allowed to be nucleotides other than A (non-A nucleotides), such as U (uridylic acid), G (guanylic acid) or C (cytidylic acid).

In some embodiments, the in vitro preparation process of the RNA is a prokaryotic fermentation process, that is, the coding nucleic acid of the RNA molecule comprising the poly (A) tail is introduced into prokaryotic cells, and the prokaryotic cells are amplified to achieve the purpose of amplifying the coding nucleic acid, and then the amplified coding nucleic acid is transcribed into the RNA. In some embodiments, the in vitro preparation process of the RNA is to connect an RNA fragment comprising a protein coding sequence to a poly (A) tail by homologous recombination, enzyme digestion and ligation, or other non-homologous recombination methods, and the poly (A) tail is prepared by a prokaryotic fermentation process. During the prokaryotic fermentation process, the coding nucleic acid comprising the poly (A) tail is introduced into prokaryotic cells, and the prokaryotic cells are amplified to achieve the purpose of amplifying the coding nucleic acid; subsequently, the amplified coding nucleic acid is transcribed into an RNA comprising the poly (A) tail. In some embodiments, the aforementioned coding nucleic acid is linear. In some embodiments, the aforementioned coding nucleic acid is circular. In some embodiments, the aforementioned coding nucleic acid is a plasmid. In some embodiments, the aforementioned coding nucleic acid is single-stranded or double-stranded. In some embodiments, the aforementioned coding nucleic acid is chemically modified before being introduced into a prokaryotic cell. In some embodiments, the aforementioned coding nucleic acid is chemically synthesized before being introduced into a prokaryotic cell. In some embodiments, the coding nucleic acid is inserted into the nucleoid / karyoid genomic DNA of the prokaryotic cell. In some embodiments, the coding nucleic acid is free in the cytoplasm or outside the nucleoid / karyoid of the prokaryotic cell. In some embodiments, the prokaryotic cell is *E. coli*.

Based on this, the present application provides a series of poly (A) tails, which are highly conserved during the in vitro preparation of RNA. The poly (A) tail comprises one or more non-A nucleotides at one or more positions.

In some embodiments, the poly (A) tail coding sequence comprises:

a single element **a** and at least one element **b**, and at least one element **c**;

a single element **a** and at least one element **b**, and at least one element **d**; or

a single element **a** and at least one element **b**, and at least one element **c** and at least one element **d**,

wherein the element **a** and element **b** are not adjacent, the element **c** and element **d** are not adjacent, and

the poly (A) tail coding sequence does not comprise any two elements **b** adjacent to each other, does not comprise any two elements **c** adjacent to each other, and does not comprise any two elements **d** adjacent to each other.

In some embodiments, the poly (A) tail coding sequence further comprises a single element **e**, wherein the element **e** consists of one or two consecutive A, which is located at the 3' terminus of the poly (A) tail coding sequence, and is adjacent to element **d** or element **c**.

The poly (A) tail herein can be a segment of RNA, or a hybrid molecule of DNA and RNA.

The present application also provides a poly (A) tail that can regulate protein expression level. In addition, the present application also provides a poly (A) tail that can regulate protein expression level while maintaining highly conservative during the in vitro preparation process. In some embodiments, the poly (A) tail that can regulate protein expression level; or the poly (A) tail that can regulate protein expression level while maintaining highly conservative during the in vitro preparation process is selected from the group consisting of:

element **a**-element **c**-element **b**-element **c**-element **b**-element **c**-element **b**-element **c**-element **b**;

element **b**-element **c**-element **b**-element **c**-element **a**-element **d**-element **b**-element **c**-element **b**-element **c**-element **b**;

element **b**-element **c**-element **b**-element **c**-element **b**-element **d**-element **a**-element **c**;

element **a**-element **d**-element **b**-element **c**-element **b**-element **c**-element **b**; and

element **b**-element **c**-element **b**-element **c**-element **b**-element **d**-element **a**.

In some embodiments, the poly (A) tail that can regulate protein expression level; or the poly (A) tail that can regulate protein expression level while maintaining highly conservative during the in vitro preparation process is selected from the group consisting of:

60A-G-19A-G-19A-G-19A-G-3A;

7A-C-18A-G-60A-element **d**-7A-C-18A-G-14A;

60A-element **d**-19A-G-19A-G-17A;

19A-G-19A-G-19A-element **d**-60A;

19A-G-19A-G-19A-element **d**-60A-G;

19A-G-19A-G-19A-element **d**-60A;

19A-C-19A-C-19A-element **d**-60A; and

19A-T-19A-T-19A-element **d**-60A.

In some embodiments, the poly (A) tail that can regulate protein expression level; or the poly (A) tail that can regulate protein expression level while maintaining highly conservative during the in vitro preparation process is: 60A-element **d**-19A-G-19A-G-17A or 19A-G-19A-G-19A-element **d**-60A.

Particularly, the two elements connected by "-" are directly connected, and there is no nucleotide between the two elements.

In the above-mentioned structure of the poly (A) tail, "yA" represents the number of consecutive A in element **a** or element **b**, wherein y is a natural number, for example, 19A means that it comprises 19 consecutive A; 60A means that it comprises 60 consecutive A.

In some embodiments, the poly (A) tail that can regulate protein expression level; or the poly (A) tail that can regulate protein expression level while maintaining highly conservative during the in vitro preparation process is selected from any one of the polynucleotide sequences represented by SEQ ID NOs: 1-10.

In addition, the present application also provides use of the above-mentioned poly (A) tail for regulating protein expression, wherein in said use, the poly (A) tail is located at the 3' end of the mRNA, such as the 3' end of the 3' UTR. In some embodiments, the regulating protein expression uses the method for regulating protein expression described below.

Engineered DNA Molecules and Libraries

The present application also provides an engineered DNA molecule that can replicate in a cell, which comprises the coding sequence of the aforementioned Poly A tail or a complementary sequence thereof. Those skilled in the art should be understood that, in addition to the coding sequence of the Poly A tail, the engineered DNA molecule should also comprise structural elements necessary for the DNA molecule to replicate or to replicate efficiently in a cell. The structural elements necessary for the DNA molecule to replicate or to replicate efficiently in a cell are known in the art and comprise, for example, an origin of replication (ORI). In some embodiments, the engineered DNA molecule further comprises a label gene or a fragment thereof, and/or a reporter gene or a fragment thereof, and a unique restriction endonuclease site allowing insertion of DNA elements, preferably a restriction endonuclease site that functions as a multiple cloning site (MCS). The label gene facilitates the identification of a cell containing a plasmid comprising the label gene, which may be selected from, for example, an antibiotic resistance gene. Each restriction endonuclease site in the MCS can be specifically recognized by a different restriction endonuclease.

In some embodiments, the DNA molecule is a DNA plasmid. As used herein, the term "DNA plasmid" refers to a plasmid consisting of a double-stranded DNA molecule. In some embodiments, the "plasmid" is a circular DNA molecule. In some embodiments, the "plasmid" may also encompass a linear DNA molecule. Specifically, the term "plasmid" also encompasses molecules obtained by linearizing a circular plasmid, for example by cleaving a circular plasmid with a restriction endonuclease, thereby converting the circular plasmid molecule into a linear molecule, as well as a linear molecule replicable in a prokaryote. A plasmid can be replicates, i.e., can be amplified in a cell independently of the genomic genetic information stored in a nucleoid or karyoid of a prokaryotic cell, and can be used for cloning, i.e., for amplifying genetic information in a bacterial cell. Preferably, the DNA plasmid according to the present application is a medium copy or high copy plasmid, more preferably a high copy plasmid. Examples of such high copy plasmids are vectors based on pUC, pTZ plasmids or any other plasmid comprising an ORI supporting high

copy of the plasmid (e.g., pMB1, pCoIE1 etc.).

In some embodiments, the DNA molecule is a DNA molecule or a fragment thereof constituting a nucleoid or karyoid of a prokaryotic organism, that is, the coding sequence comprising the aforementioned poly (A) tail or a complementary sequence thereof can be replicated along with the prokaryotic genome.

In some embodiments, the DNA molecule is further connected to a gene of interest fragment on the 5' end side of the poly (A) tail coding sequence, and the gene of interest fragment and the poly (A) tail coding sequence co-encode RNA. In some embodiments, the gene of interest fragment and the poly (A) tail coding sequence co-encode mRNA. The gene of interest fragment comprises a coding sequence of a protein, a polypeptide or a fragment thereof. In some embodiments, the gene of interest fragment also comprises a coding sequence for an element that can be used to initiate or regulate the expression of the protein, polypeptide or a fragment thereof after transcription, and the elements include, but are not limited to, 5' UTR, 3' UTR, etc. In some embodiments, the gene of interest fragment comprises a coding sequence of at least one untranslated region (UTR). In some embodiments, the gene of interest fragment comprises at least the coding sequence of the 5' UTR and the coding sequence of the protein, polypeptide or a fragment thereof. In some embodiments, the gene of interest fragment comprises from 5' to 3' sequentially at least: the coding sequence of 5' UTR, the coding sequence of protein, polypeptide or a fragment thereof, and the coding sequence of 3' UTR. The coding sequence of the protein, polypeptide or a fragment thereof can be ultimately translated into one or more proteins, or one or more polypeptides, for example, short peptides, oligopeptides, polypeptides, fusion proteins, proteins and fragments thereof, such as parts of known proteins, such as functional parts. The functional portion can be, for example, a biologically active portion of a protein, or an antigenic portion that can effectively generate antibodies, such as an antigenic epitope. The two ends of the coding sequence of the protein, polypeptide or a fragment thereof respectively comprise a start codon (5' end) and a stop codon (3' end), which are respectively the first three nucleotides and the last three nucleotides of the mRNA molecule that can be translated. The 5' UTR usually comprises at least one ribosome binding site (RBS), such as the Shine-Dalgarno sequence in a prokaryote, or at least one translation start site, such as the Kozak sequence in a eukaryote. RBS promotes efficient and accurate translation of mRNA molecules by recruiting ribosomes upon the translation initiation. The activity of which can be optimized by varying the length and sequence of a given RBS or translation-initiating site, as well as the distance from the given RBS or translation-initiating site to the start codon. Alternatively or optionally, the 5' UTR comprises an internal ribosome entry site or IRES. The 3' UTR may comprise one or more regulatory sequences, for example binding sites for amino acid sequences that enhance the stability of the mRNA molecule, binding sites for regulatory RNA molecules (such as miRNA molecules), and/or signal sequences involved in the intracellular transport of mRNA molecules.

Based on the foregoing embodiments, in some embodiments, the gene of interest fragment

further comprises one or more additional regulatory sequences, such as binding sites for amino acid sequences that enhance the stability of mRNA molecules, binding sites for amino acid sequences that enhance the translation of mRNA molecules, regulatory elements (such as riboswitches), binding sites for regulatory RNA molecules (such as miRNA molecules), and/or nucleotide sequences that positively affect translation initiation. Furthermore, within the 5' UTR, preferably there is no functional upstream open reading frame, out-of-frame upstream translation start site, out-of-frame upstream start codon, and/or nucleotide sequences that produce secondary structures that reduce or prevent translation. The presence of such nucleotide sequences in the 5' UTR may negatively affect translation.

The coding sequence of the protein, polypeptide or a fragment thereof comprises codons that can be translated into an amino acid sequence. All the codons comprised in the coding sequence may be naturally occurring codons coding amino acids or may be partially or entirely composed of artificially synthesized codons. In some embodiments, some or all of the codons are codon optimized. In some embodiments, some or all of the codons encode unnatural amino acids.

In some embodiments, the DNA molecule further comprises structural elements necessary for initiating or regulating transcription of the RNA on the 5' end side of the gene of interest fragment, and the structural elements are known in the art. In some embodiments, the structural element comprises at least a promoter. Promoters and their sequences are known in the art, including weak promoters, medium strength promoters, strong promoters, mini-promoters or core promoters, etc. In some specific embodiments, the promoter is a strong promoter. In some embodiments, the promoter can initiate transcription of the gene of interest fragment and/or poly (A) tail in a prokaryotic cell. In some embodiments, the promoter can initiate transcription of the gene of interest fragment and/or poly (A) tail in a eukaryotic cell. The "promoter" comprises at least one transcription recognition site followed by a transcription factor binding site. The recognition and binding sites may interact with an amino acid sequence that mediates or regulates transcription. Compared with the recognition site, the binding site is more closer to the aforementioned gene of interest fragment. The binding site may be, for example, a Pribnow box in a prokaryote or a TATA box in a eukaryote. For example, in some embodiments, when the Pribnow box is used, the transcription recognition site can be located about 35 bp upstream of the transcription start site, while the transcription factor binding site can be located about 10 bp upstream of the transcription start site. In some embodiments, the promoter comprises at least one additional regulatory element, for example an AT-rich upstream element located approximately 40 and/or 60 nucleotides before the transcription start site, and/or an additional regulatory element for enhancing promoter activity located between the recognition site and the binding site. In some embodiments, the promoter is a strong promoter, i.e., the promoter comprises a sequence for promoting the transcription of the aforementioned RNA coding sequence. Strong promoters are known to those skilled in the art, such as the OXB18, OXB19 and OXB20 promoters derived from the RecA promoter of *E. coli* or can be identified or

synthesized by routine laboratory procedures. In some embodiments, the promoter is a T7 promoter. In some embodiments, the promoter further comprises additional regulatory elements, such as an enhancer comprised in the DNA plasmid that can promote the transcription of the aforementioned RNA coding sequence.

The present application also provides a library comprising the aforementioned engineered DNA molecules. In some embodiments, the library comprises at least two DNA molecules having different poly (A) tail coding sequences.

In addition, the present application also provides use of the above-mentioned engineered DNA molecule for stably amplifying the poly (A) tail coding sequence or the coding sequence of RNA with a poly (A) tail. In some embodiments, the method for amplifying the coding sequence of the poly (A) tail or the coding sequence of the RNA with the poly (A) tail is as described below in the method for stably amplifying the poly (A) tail transcription template DNA in vitro.

Engineered RNAs and libraries

The present application provides an RNA comprising the aforementioned poly (A) tail and a gene of interest fragment on the 5' end side of the poly (A) tail coding sequence. In some embodiments, the RNA further comprises a 5'-cap structure. In some embodiments, the RNA is mRNA.

As used herein, "mRNA" (messenger RNA) is any RNA, naturally occurring, non-naturally occurring or modified, that encodes at least one protein, polypeptide, or a fragment thereof, which is enabled to be translated to produce the encoded protein, polypeptide, or a fragment thereof in vitro, in vivo, in situ, or ex vivo. Therefore, the mRNA may be a mature mRNA or a pre-mature mRNA, and the elements or structures that the mRNA must comprise or optionally comprise are known in the art. In some embodiments, the mRNA comprises coding sequences for multiple necessary functional elements to express, regulate, or enhance the expression level of the protein, polypeptide, or a fragment thereof. The functional elements include, but are not limited to, 5' cap, 5' UTR, 3' UTR, etc. Both the 5' UTR and the 3' UTR are usually transcribed from genomic DNA, which are the elements present in the pre-mature mRNA. As mature mRNA,

The term "5' cap" is located at the 5' terminus of the mRNA, which comprises a methylated guanylate that is linked to the 5' terminus of the mRNA via pyrophosphate so as to form a 5',5'-triphosphate linkage with its adjacent nucleotide. There are usually three types of 5' cap structures (m⁷G5'ppp5'Np, m⁷G5'ppp5'NmpNp, and m⁷G5'ppp5'NmpNmpNp), which are called type O, type I and type II, respectively. The type O means that the ribose of the terminal nucleotide is not methylated, the type I means that the ribose of one terminal nucleotide is methylated, and the type II means that the ribose of both terminal nucleotides is methylated. In some embodiments, as for the 5' cap, according to the manufacturer's protocol, a 5'-guanosine cap structure can be produced by completion of capping the 5' 'of a polynucleotide during an in vitro transcription reaction through utilizing the following chemical RNA cap analogs: 3'-O-Me-m⁷G(5')ppp(5')G

[ARCA cap], G(5')ppp(5')A, G(5')ppp(5')G, m7G(5')ppp(5')A, m7G(5')ppp(5')G (New England BioLabs, Ipswich, MA), or m7G(5')ppp(5')(2'-OMeA)pG (CleanCapAG). For example, in some embodiments, 5' capping of the modified RNA can be accomplished after transcription by using vaccinia virus capping enzyme to produce an O-type cap structure: m7G(5')ppp(5')G (New England BioLabs, Ipswich, MA). The type I cap structures can be generated by using both vaccinia virus capping enzyme and 2'-O methyl-transferase to produce m7G(5')ppp(5')(2'-OMeA)pG. The type II cap structures can be generated from the type I cap structures by subsequent 2'-O-methylation of the third nucleotide from 5'-terminus by using a 2'-O-methyl-transferase. The type III cap structures can be generated from the type II cap structures by subsequent 2'-O-methylation of the fourth nucleotide from 5'-terminus by using a 2'-O-methyl-transferase.

In some embodiments, some or all of the uridines in the mRNA are chemically modified uridines.

In some embodiments, some or all of the uridines in the mRNA are pseudouridines or 1-methyl-pseudouridines.

In some embodiments, some or all of the uracil nucleotides in the mRNA are substituted with pseudouridine (ψ) nucleotides or N1-methyl-pseudouridine (m1 ψ) nucleotides.

In some embodiments, the mRNA further comprises a stabilization element. The stabilization elements may comprise, for example, histone stem-loops. In some embodiments, the mRNA comprises a coding region, at least one histone stem-loop, and optionally a poly (A) sequence or a polyadenylation signal. The poly (A) sequence or polyadenylation signal should generally enhance the expression level of the encoded protein. In some embodiments, the mRNA comprises a combination of a poly (A) sequence or polyadenylation signal and at least one histone stem-loop, although both of them have alternative mechanisms in nature, they act synergistically to increase the protein expression to a level beyond those observed with either element alone. The synergistic effect of a combination of the poly (A) and at least one histone stem-loop is independent of the order of the elements or the length of the poly (A) sequence. In some embodiments, the histone stem-loop is generally derived from a histone gene, and comprises an intramolecular base-pairing formed loop by two adjacent partial or complete reverse complementary sequences separated by a spacer region (composed of a short sequence). The unpaired loop region is typically unable to base pair with either of the stem-loop elements. The stability of the stem-loop structure generally depends on the length, the number of mismatches or bulges, and the base composition of the paired region. In some embodiments, wobble base-pairing (non-Watson-Crick base-pairing) can be produced. In some embodiments, said at least one histone stem-loop sequence comprises 15-45 nucleotides in length.

In some embodiments, one or more AU-rich sequences of the mRNA may be removed. Such sequences are sometimes called AURES, which are destabilizing sequences found in the 3' UTR. The AURES may be removed from the mRNA. Alternatively, The AURES may be retained in the

mRNA.

In some embodiments, the mRNA is formulated within a lipid nanoparticle (LNP). In some embodiments, lipids are mixed with the mRNA to form lipid nanoparticles. In some embodiments, the RNA is formulated in lipid nanoparticles. In some embodiments, the lipid nanoparticles are first formed as empty lipid nanoparticles, and then are combined or encapsulated with the mRNA of the vaccine just prior to administration (e.g., within a few minutes to an hour).

The lipid nanoparticles generally comprise ionizable lipids, non-cationic lipids, sterols and PEG lipid components and a target nucleic acid, such as the above-mentioned mRNA. The lipid nanoparticles according to the present disclosure can be produced by using components, compositions and methods generally known in the art, see, e.g., PCT/US2016/052352, PCT/US2016/068300, PCT/US2017/037551, PCT/US2015/027400, PCT/US2016/047406, PCT/US2016000129, PCT/US2016/014280, PCT/US2016/014280, PCT/US2017/038426, PCT/US2014/027077, PCT/US2014/055394, PCT/US2016/52117, PCT/US2012/069610, PCT/US2017/027492, PCT/US2016/059575 and PCT/US2016/069491, all of which are incorporated herein by reference in their entireties.

The present application also provides a library comprising the aforementioned mRNA molecules. The library comprises at least two mRNA molecules with different poly (A) tails.

The present application also provides uses of the mRNA and the mRNA library. At least two or more mRNA molecules having poly (A) tails with different influence gradients on the expression level of mRNA can be used to regulate the expression level of the coding sequence of the aforementioned protein, polypeptide or a fragment thereof, for example, by adjusting the ratio of different mRNA molecules in the library comprising said two or more mRNA molecules, or by introducing one or more of said two or more mRNA molecules with the same or different contents at different times.

Cells

The present application also provides cells comprising the aforementioned engineered DNA molecules, wherein the DNA molecules can be stored and/or amplified in the cells. In some embodiments, the cells are prokaryotic cells in which the DNA molecules can be replicated. In some embodiments, the cells are prokaryotic cells in which the DNA molecules can be replicated and/or transcribed. In some embodiments, the DNA molecules are eukaryotic cells in which the DNA molecules can be replicated. In some embodiments, the DNA molecules can be transcribed and/or replicated in the DNA-containing cells.

In some embodiments, the cell is a prokaryotic cell. In some embodiments, the cell is a bacterium, an actinomycete, a cyanobacterium, a mycoplasma, a rickettsia, and a chlamydia. In some embodiments, the cell is selected from the group consisting of *Bacillus subtilis*, *Lactobacillus*, *Acetobacter*, *Corynebacterium*, *Brevibacterium*, *Arthrobacter*, *Pseudomonas*, and *Pediococcus*. In some embodiments, the cell is a recA-bacterium. In some embodiments, the cell is *E. coli*. In some

embodiments, the cell is *E. coli*, which is selected from the group consisting of K-12 and derivatives thereof, and B strain and derivatives thereof. In some embodiments, the *E. coli* is selected from the group consisting of MG1655, DH5 or DH5 α , DH10B, BL21, DB3.1, HB101, JM109, JM110, MC1061, MG1655, Pir1, Stbl2, Stbl3, Top10, XL1Blue, XL10Gold, BLR, HMS174, Tuner, Rostetta2, Lemo21, T7Express, and Origami2, etc. In some embodiments, the cell is selected from the group consisting of *Streptomyces*, *Micromonospora*, and *Nocardia*. In some embodiments, the cell is a fungus. In some embodiments, the cell is selected from a yeast or a mold.

Methods

The present application provides a method for stably amplifying a poly (A) tail transcription template DNA in vitro, so as to reduce the mutation frequency of the poly (A) tail transcription template sequence when the DNA is replicated in large quantities in a cell. The method comprises: expanding the cells comprising the engineered DNA molecules.

In some embodiments, prior to expanding the cells, the method further comprises introducing the engineered DNA molecule into the cells. In some embodiments, the introducing may comprise chemical transformation or electro-transformation. In some embodiments, the introducing is a natural endocytic process of the engineered DNA molecule performed by the cell.

In some embodiments, after expanding the cells, the method further comprises extracting cellular DNA and synthesizing the RNA by in vitro transcription. In some embodiments, after expanding the cells, the method further comprises inducing transcription of the RNA in the cells, and then extracting and isolating the RNA therein. In some embodiments, the method further comprises extracting the cellular DNA and transducing it into a second cell that can transcribe the RNA. In some embodiments, the transducing comprises administering to a human, wherein the administering is selected from the group consisting of intravenous, intraperitoneal, subcutaneous, intracranial, intrathecal, intraarterial (e.g., via the carotid artery), intramuscular, and intratumoral injection or perfusion.

In addition, the present application also provides a method for regulating protein expression, wherein the method comprises:

introducing two or more of the aforementioned engineered DNA molecules into cells of interest at different times and/or in different quantitative ratios; or

introducing two or more of the aforementioned RNA molecules into cells of interest at different times and/or in different quantitative ratios;

wherein the two or more of the aforementioned engineered DNA molecules and the two or more of the aforementioned RNA molecules have different poly (A) tails, and the poly (A) tails have different influence gradients on the expression level of RNA.

In some embodiments, the present application also provides a method for regulating protein expression, wherein the method comprises introducing the aforementioned engineered DNA molecule or the aforementioned RNA molecule into a cell of interest. In some embodiments, the

coding sequence of the poly (A) tail encoded by the DNA and the poly (A) tail comprised in the RNA comprises a structure selected from the group consisting of:

element **a**-element **c**-element **b**-element **c**-element **b**-element **c**-element **b**-element **c**-element **b**;

element **b**-element **c**-element **b**-element **c**-element **a**-element **d**-element **b**-element **c**-element **b**-element **c**-element **b**;

element **b**-element **c**-element **b**-element **c**-element **b**-element **d**-element **a**-element **c**;

element **a**-element **d**-element **b**-element **c**-element **b**-element **c**-element **b**; or

element **b**-element **c**-element **b**-element **c**-element **b**-element **d**-element **a**.

In some embodiments, the coding sequence of the poly (A) tail encoded by the DNA and the poly (A) tail comprised in the RNA comprises a structure selected from the group consisting of:

60A-G-19A-G-19A-G-19A-G-3A;

7A-C-18A-G-60A-element **d**-7A-C-18A-G-14A;

60A-element **d**-19A-G-19A-G-17A;

19A-G-19A-G-19A-element **d**-60A;

19A-G-19A-G-19A-element **d**-60A-G;

19A-G-19A-G-19A-element **d**-60A;

19A-C-19A-C-19A-element **d**-60A; and

19A-T-19A-T-19A-element **d**-60A.

In some embodiments, the coding sequences of the poly (A) tail encoded by the DNA and the poly (A) tail comprised in the RNA comprise the following structure: 60A-element **d**-19A-G-19A-G-17A, or 19A-G-19A-G-19A-element **d**-60A.

Particularly, the two elements connected by the "-" are directly connected, and there is no nucleotide between the two elements.

In the above-mentioned structure of the poly (A) tail, "yA" represents the number of consecutive A in element **a** or element **b**, and y is a natural number, for example, 19A means that it comprises 19 consecutive A; 60A means that it comprises 60 consecutive A.

In some embodiments, the coding sequences of the poly (A) tail encoded by the DNA and the poly (A) tail comprised in the RNA comprise or consist of any one polynucleotide sequence selected from the polynucleotide sequences represented by SEQ ID NOs: 1-10.

It should be understood that the present application encompasses various aspects, embodiments and combinations of the aspects and/or embodiments described herein. The above description and the following Examples are intended to illustrate rather than limit the scope of the present application. Other aspects, improvements and modifications within the scope of the present application will be apparent to those skilled in the art. Therefore, those skilled in the art should appreciate that the scope of the present application also comprises such improvements and modifications to these aspects and embodiments.

EXAMPLES

Example 1: Construction of poly (A) Tail

The poly (A) tails and the DNA sequences coding the poly (A) tails shown in Table 1 below were constructed by conventional genetic engineering methods.

Table 1.

Design	Specific structure	Sequence	Length
P1	60A-G-19A-G-19A-G-19A-G-AAA	SEQ ID NO: 1	124 nt
P2	7A-C-18A-G-60A-element d -7A-C-18A-G-14A	SEQ ID NO: 2	130 nt
P3	60A-element d -19A-G-19A-G-17A	SEQ ID NO: 3	123 nt
P4	19A-G-19A-G-19A- element d -60A	SEQ ID NO: 4	125 nt
P5	19A-G-19A-G-19A- element d -60A-G	SEQ ID NO: 5	126 nt
P6	19A-G-19A-G-19A-element d -60A	SEQ ID NO: 6	125 nt
P7	19A-G-19A-G-19A-element d -60A	SEQ ID NO: 7	131 nt
P8	19A-C-19A-C-19A-element d -60A	SEQ ID NO: 8	125 nt
P9	19A-T-19A-T-19A-element d -60A	SEQ ID NO: 9	125 nt
P10	19A-G-19A-G-19A-element d -60A	SEQ ID NO: 10	125 nt
C1	A60-element d -A60	SEQ ID NO: 11	130 nt
C2	A30-element d -A70	SEQ ID NO: 12	110 nt
C3	A60-element c -A60	SEQ ID NO: 13	121 nt
C4	A60-element d -A60	SEQ ID NO: 14	126 nt

Example 2: Testing poly (A) Function by Using Luciferase Coding Sequence as an Example

2.1 Testing the replication stability of mRNA-encoding DNA molecules in prokaryotic cells

Using luciferase as the protein coding region, the stability of different poly (A) variants in *E. coli* and their effects on luciferase expression in cells were investigated.

1) Construction of a universal vector comprising the luciferase protein coding region

The *E. coli* cloning vector pUC57 was used as the vector backbone in this universal vector, and a T7 promoter sequence (5'-TAATACGACTCACTATAAGG-3'), a 5' UTR, a luciferase protein, a 3' UTR and a polyadenylic acid string poly (dA:dT) were sequentially arranged between the multiple cloning sites: Xba I restriction site and the EcoR I restriction site.

2) The polyadenylic acid string poly (dA: dT) in the universal vector was substituted with P1-P10 of the present application, and controls C1-C4 (A60-10 nt spacer-A60 (control C1), A30-10 nt spacer-A70 (control C2), A60-1 nt spacer-A60 (control C3) or A60-6 nt spacer-A60 (control C4). Among them, the C1 and C2 are derived from a patent of the literature (U.S. Pat. No. US10717982B2).

All primers required for constructing P1-P10 and control C1-C4 were synthesized, and double digestion was conducted by two restriction endonucleases to remove the poly (dA:dT) from the

universal vector constructed in step 1); and then P1-P10 and C1-C4 were ligated to the vector from which the poly (dA:dT) was removed by T4 DNA ligase 1, thereby completing the substitution of the poly (dA:dT) in the universal vector.

3) Detection of the replication stability of different poly (A) variants in *E. coli*

The vector plasmid constructed in step 2) was confirmed to be correct by sequencing, and then transformed into *E. coli* DH5 α . The transformed plate was grown at 30°C, then completing plasmid extraction and sequencing. After the sequencing was completed, the stability and base deletion of different poly (A) variants were analyzed and calculated based on the sequencing results. The replication stability is expressed by the percentage of clones without any base changes, and the higher the percentage, the higher the replication stability of the plasmid in *E. coli*.

The results are shown in Figs. 1-2, and the specific experimental results are as follows:

A total of 100 clones were tested for control C1, in which 15 clones had base deletions, accounting for 15%; and the number of correct clones was 85, accounting for 85%.

A total of 50 clones were tested for control C2, in which all clones were correct without any base changes or deletions, with a correct clone percentage of 100%.

A total of 50 clones were tested for control C3, in which 9 clones had base deletions, accounting for 18%; and the number of correct clones was 41, accounting for 82%.

A total of 50 clones were tested for control C4, in which 14 clones had base deletions, accounting for 28%; and the number of correct clones was 36, accounting for 72%.

A total of 100 clones were tested for P1, in which 9 clones had base deletions, accounting for 9%; and the number of correct clones was 91, accounting for 91%.

A total of 100 clones were tested for P2, in which 12 clones had base deletions, accounting for 12%; the number of correct clones was 88, accounting for 88%.

A total of 62 clones were tested for P3, in which 5 clones had base deletions, accounting for 8%; and the number of correct clones was 57, accounting for 92%.

A total of 50 clones were tested for P4, in which 3 clones had base deletions, accounting for 6%; and the number of correct clones was 47, accounting for 94%.

A total of 50 clones were tested for P5, in which 4 clones had base deletions, accounting for 8%; and the number of correct clones was 46, accounting for 92%.

A total of 50 clones were tested for P6, in which 5 clones had base deletions, accounting for 10%; and the number of correct clones was 45, accounting for 90%.

A total of 50 clones were tested for P7, in which 7 clones had base deletions, accounting for 14%, and the number of correct clones was 43, accounting for 86%.

A total of 50 clones were tested for P8, in which 3 clones had base deletions, accounting for 6%, and the number of correct clones was 47, accounting for 94%.

A total of 50 clones were tested for P9, in which 4 clones had base deletions, accounting for 8%, and the number of correct clones was 46, accounting for 92%.

A total of 50 clones were tested for P10, in which 5 clones had base deletions, accounting for 10%, and the number of correct clones was 45, accounting for 90%.

Taking into account the results of the correct clone ratio and the number of deleted bases, the poly (A) variant designed in the present application is superior to or comparable to the prior art in terms of replication stability in *E. coli* cells. Particularly, the replication stability of poly (A) variants P3, P4, and P8 is the highest. The replication stability of P3, P4, and P8 is equivalent to that of C2, with no statistically significant difference ($p>0.05$, χ^2 test). In addition, the replication stability of P3, P4, and P8 is better than that of the controls C1, C3 and C4, and the difference is statistically significant ($p<0.05$, χ^2 test).

Example 3: Testing the Poly (A) Function by Using HPV Antigen Protein Coding Sequence as an Example

1) Construction of plasmid with HPV as protein coding region

As described above, in Example 1, vectors comprising luciferase coding genes combined with different poly (A)s were constructed. Based on these vectors, the luciferase coding gene was substituted with the HPV coding gene by conventional molecular cloning methods; and the main elements were arranged in order of T7 promoter sequence (5'-TAATACGACTCACTATAAGG-3'), 5' UTR, HPV antigen protein coding sequence, 3' UTR and poly (A) coding sequence.

2) Performing small-scale bacterial culture to test the stability of four HPV poly (A) variants in *E. coli*

The HPV vector plasmid comprising P1, P2, P3 and P4 constructed in step 1) was confirmed to be correct by sequencing, and then transformed into *E. coli* DH5 α . The transformed plates were grown at 30°C, then completing plasmid extraction and sequencing. After completion of sequencing, the stability and base deletion of different poly (A) variants were analyzed and calculated based on the sequencing results. The stability is expressed by the percentage of clones without any base changes, and the higher the percentage, the more stable it is.

The results showed (Figs. 3-4) that, when the gene of interest was substituted with the HPV antigen coding sequence,

a total of 88 clones were tested for control C1, in which 47 clones had base deletions, accounting for 53%, and the number of correct clones was 41, accounting for 47%.

a total of 50 clones were tested for control C2, in which 1 clone had a base deletion, accounting for 2%, and the number of correct clones was 49, accounting for 98%.

a total of 101 clones were tested for P1, in which 14 clones had base deletions, accounting for 14%, and the number of correct clones was 87, accounting for 86%.

a total of 100 clones were tested for P2, in which 12 clones had base deletions, accounting for 12%, and the number of correct clones was 88, accounting for 88%.

a total of 70 clones were tested for P3, in which 4 clones had base deletions, accounting for 6%,

and the number of correct clones was 66, accounting for 94%.

a total of 50 clones were tested for P4, in which 9 clones had a plurality of base deletions, accounting for 18%, and the number of correct clones was 41, accounting for 82%.

Taking into account the ratio of mutations and the average number of base deletions, when the protein-coding gene of the luciferase in Example 1 was substituted with the HPV antigen protein, the different poly (A) variants of the present application still maintain high replication stability; particularly, P3 and C2 have comparable stability, with no statistically significant difference ($p>0.05$, χ^2 test), and compared with new variants in other groups of the present application, the cloning stability is optimal. The probability of large fragment deletion in C2 is $1/50=2\%$, while the probability of large fragment deletion in P3 is $1/70=1.4\%$. Since large poly (A) fragment deletion will affect the in vivo expression and efficacy of mRNA products, P3 is more in line with product requirements. The above results indicate that the poly (A) variants designed in the present application are universally applicable in examples involving different protein coding regions.

As described above, in Example 1 and Example 2, the plates transformed with *E. coli* were cultured in a biochemical incubator at 30°C overnight, and the obtained clones were sequenced to evaluate the replication stability. In addition, since the culture temperature of *E. coli* affects the DNA replication rate, which further affects the replication stability. With respect to Example 3, after the plates transformed with *E. coli* were cultured in a 37°C biochemical incubator, the sequencing detections were also compared in the present application. The results show (Fig. 5) that, culturing at 37°C significantly increases the base deletion ratio of the control C1, from 53% to 98% when culturing at 30°C. Unlike the control, there is no significant difference in the mutation rates of P1, P2 and P3 under the two temperature conditions, indicating that the poly (A) variants designed in the present application still have high replication stability in the examples with respect to different temperature culture conditions of *E. coli*, resulting in a universal application per se.

3) Detection of the stability of the three poly (A) variants in large-scale fermentation and at different generations

The production of mRNA drugs must rely on large-scale fermentation to prepare sufficient template plasmids, and the stability of the plasmid during fermentation (in this application, the stability of the plasmid particularly refers to the stability of poly dA:dT) is crucial to the production of mRNA drugs with uniform quality. On the other hand, in order to meet the stability requirements of different batches of production, it is necessary to establish a strain library comprising the plasmid of interest, including strain libraries of different generations such as primary library and secondary library. Therefore, it is necessary to evaluate the plasmid stability in *E. coli* for different generations. In response to the above two issues, the present application detected the stability of P1 and P3 between different generations of the fermentation process in Example 3. According to the sequencing results, 4 correct *E. coli* clones were selected for each poly (A) variant to perform passage culture through fermentation, respectively. The results show that, at the 3rd, 5th, 7th, and

9th seed passages, the plasmids of the four clones of P1 and P3 remain stable, without any base changes.

Example 4: Detection of Luciferase mRNA Expression Level in Mice

In eukaryotic cells, a poly (A) tail with a certain length is essential for protecting the 3' end of mRNA, maintaining mRNA stability and promoting protein expression. Affected by the physiological or environmental factors in the body, poly (A) gradually becomes shorter, thus priming mRNA degradation. This application investigated the effects of different poly (A) variants on protein expression levels. We performed in vivo expression assays in mice to evaluate the effects of different poly A variants on luciferase activity. During the implementation of this application, luciferase mRNA-LNP comprising control C2, as well as P3, P4, P5, P8 and P9 were injected intramuscularly into mice. Animal imaging was performed 6 hours after injection, and the fluorescence intensity was quantitatively to compare the effects of different poly (A)s on the activity of luciferase in vivo. The specific experimental process is as follows:

Luciferase mRNA was synthesized by in vitro transcription, and linearized DNA was obtained by digestion with type II restriction endonuclease BspQ I. The 3' end of the linearized DNA was exhibited as different poly (A): control C2, P3, P4, P5, P8 or P9. The linearized DNA was used as a template for in vitro transcription. A 100 µl reaction system in which 1X reaction buffer; 5 mM (final concentration) of ATP, CTP, N1M-UTP and GTP, respectively; 4 mM (final concentration) of CleanCap AG; and 5 µl of in vitro transcriptase were included. After mixing the reaction mixture thoroughly, the reaction was conducted at 37°C for 3 h. The in vitro transcribed mRNA was collected by LiCl precipitation, and finally dissolved in enzyme-free water.

Animal experiment: the in vitro synthesized luciferase mRNA was encapsulated into LNPs, and the obtained mRNA stock solution was dispersed in 20 mM acetic acid solution (pH 5.0) respectively to obtain an RNA solution with an mRNA concentration of 200 µg/mL. The mixed fat was prepared by mixing ionizable fat, cholesterol, DSPC, and DMG-PEG2000 with the molar ratio of ionizable fat: cholesterol: DSPC: DMG-PEG2000 = 50: 38.5: 10: 1.5. The mRNA and the lipid mixture were mixed by controlling the flow rates of the aqueous phase and the oil phase through T mixing, and the injection pump was started to mix the mRNA solution with the lipid mixture to form LNP. Then, the solution was diluted 10 times with diluent, then concentrated by centrifugation in an ultrafiltration tube, followed by three times of replacement. The solution obtained above was added with Tris aqueous solution to adjust the pH to 7.0-8.0 to obtain an LNP-encapsulated mRNA solution, and the LNP means lipid nanoparticle. The concentration and particle size of the LNP-encapsulated mRNA were determined by using Ribogreen RNA quantification kit (Invitrogen, R11490) and Darwin ZetaSizer particle size analyzer, respectively. In the 4-component LNP, the molar ratio of each component is given as SM102: DSPC: cholesterol: DMG-PEG2000 = 50:10:38.5:1.5. After encapsulation, quality control was conducted for the LNP by measuring the

particle size, encapsulation efficiency, PDI and other indicators. The quality control results show that the prepared LNPs meet the standards of particle size range of 50nm-150nm, PDI<0.3, and encapsulation efficiency>90%, which could be used for subsequent experiments. The mRNA content in LNP was determined by the ribogreen method, and then LNP was diluted to an mRNA content of 100 ng/μL.

BALB/c mice were randomly divided into groups according to body weight, and administered after 2-3 days of adaptive feeding. Each of control C2, P3, P4, P5, P8 and P9 was respectively injected into five mice, and each mouse was administered 100 μl (10 μg mRNA) by intramuscular injection. In addition, five mice injected with PBS were used as a control group. Animal imaging was performed 6 h after the injection, and then fluorescence values were calculated. The results show (Figs. 7-8) that, compared with the control C2, the expression activity of P3 is significantly increased by 1.8 times, with a statistically significant difference ($p<0.01$, student's t-test). The expression levels of P4, P5, P8 and P9 are comparable to that of C2, with no significant difference ($p>0.05$, student's t-test).

The sequences used in the above examples of this application are shown in the following sequence listing. It should be understood that, the following sequences are merely exemplary sequences of the embodiments of the present application, and do not impose any limitation on the embodiments of the present application. The nucleic acid sequences in the following sequence listing may represent DNA sequences or RNA sequences, and "T" represents a uridine when they represent RNA sequences.

Sequence Listing:

SEQ ID NO.	Name	Sequence
Nucleic acid sequence		
1	P1	AA AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAAAA AAAAAAAAAGAAAAAAAAAAAAAAAAAGAAA
2	P2	AAAAAACAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAA AAAG GAAAAAACAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAA
3	P3	AA AAAAAAAAAAAAAAAAAGATATCAAAAAAAAAAAAAAAAAAGAAA AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAA
4	P4	AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAAA AAAAAAAAAAAAAAAAAGATATCAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
5	P5	AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAAA AAAAAAAAAAAAAAAAAGATATCAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG

6	P6	AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAA AAAAAAAAAAAAAAAAAGTATACAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
7	P7	AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAA AAAAAAAAAAAAAAAAAGATATCGTATACAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
8	P8	AAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAACAAA AAAAAAAAAAAAAAAAAGATATCAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
9	P9	AAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAATAAAA AAAAAAAAAAAAAAAAAGATATCAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
10	P10	AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAA AAAAAAAAAAAAAAAAGAATCTAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
11	C1	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAGCATATGACTAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
12	C2	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCATATGACTAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAA
13	C3	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
14	C4	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAATGCATAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
15	Element d, example 1	GATATC
16	Element d, example 2	GTATAC
17	Element d, example 3	GAATCT
18	Element d, example 4	GCATATGACT
19	Element d, example 5	GATATCGTATAC
20	3' UTR	CTCGCTTCTTGCTGTCCAATTTCTATTAAAGGTTTCCTTGTTCCTAA GTCCAATACTAACTGGGGGATATTATGAAGGGCCTTGAGCATCTG GATTCTGCCTGCTCGCTTCTTGCTGTCCAATTTCTATTAAAGGTTTCCT TTGTTCCCTAAGTCCAATACTAACTGGGGGATATTATGAAGGGCCT TGAGCATCTGGATTCTGCCT
21	5' UTR	ACATTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGAC ACC

22	Sequence of luciferase encoding gene	ATGGAAGACGCCAAAAACATTAAGAAGGGGCCAGCGCCATTCTACCC ACTCGAAGACGGGACCGCCGGCGAGCAGCTGCACAAAGCCATGAAG CGCTACGCCCTGGTGCCCGGCACCATCGCCTTTACCGACGCACATATC GAGGTGGACATTACCTACGCCGAGTACTTCGAGATGAGCGTTCGGCT GGCAGAAGCTATGAAGCGCTATGGGCTGAATACAAACCATCGGATCG TGGTGTGCAGCGAGAATAGCTTGCA GTTCTTCATGCCCCGTGTTGGGT GCCCTGTTCATCGGTGTGGCTGTGGCCCCAGCTAACGACATCTACAA CGAGCGCGAGCTGCTGAACAGCATGGGCATCAGCCAGCCCACCGTC GTATTCGTGAGCAAGAAAGGGCTGCAAAAGATCCTCAACGTGCAAA AGAAGCTACCGATCATACAAAAGATCATCATCATGGATAGCAAGACC GACTACCAGGGCTTCCAAAGCATGTACACCTTCGTGACTTCCCATTG CCACCCGGCTTCAACGAGTACGACTTCGTGCCCCGAGAGCTTCGACCG GGACAAAACCATCGCCCTGATCATGAACAGTAGTGGCAGTACCGGAT TGCCCAAGGGCGTAGCCCTACCGCACCGCACCGCTTGTGTCCGATTC AGTCATGCCCGCGACCCCATCTTCGGCAACCAGATCATCCCCGACAC CGCTATCCTCAGCGTGGTGCCATTTCACCACGGCTTCGGCATGTTTAC CACGCTGGGCTACTTGATCTGCGGCTTTCGGGTCGTGCTCATGTACCG CTTCGAGGAGGAGCTATTCTTGCGCAGCTTGCAAGACTATAAGATT AATCTGCCCTGCTGGTGCCCACTATTTAGCTTCTTCGCTAAGAGCA CTCTCATCGACAAGTACGACCTAAGCAACTTGCACGAGATCGCCAGC GGCGGGGCGCCGCTCAGCAAGGAGGTAGGTGAGGCCGTGGCCAAA CGCTTCCACCTACCAGGCATCCGCCAGGGCTACGGCCTGACAGAAAC AACCAGCGCCATTCTGATCACCCCCGAAGGGGACGACAAGCCTGGC GCAGTAGGCAAGGTGGTGCCCTTCTTCGAGGCTAAGGTGGTGGACTT GGACACCGGTAAGACACTGGGTGTGAACCAGCGCGGCGAGCTGTGC GTCCGTGGCCCCATGATCATGAGCGGCTACGTTAACAACCCCGAGGC TACAAACGCTCTCATCGACAAGGACGGCTGGCTGCACAGCGGCGAC ATCGCCTACTGGGACGAGGACGAGCACTTCTTCATCGTGGACCGGCT CAAAAGCCTGATCAAATACAAGGGCTACCAGGTAGCCCCAGCCGAA CTGGAGAGCATCCTGCTGCAACACCCCAACATCTTCGACGCCGGGGT CGCCGGCCTGCCCCGACGACGATGCCGGCGAGCTGCCCGCCGAGTC GTCGTGCTGGAACACGGTAAAACCATGACCGAGAAGGAGATCGTGG ACTATGTGGCCAGCCAGGTTACAACCGCCAAGAAGCTGCGCGGTGG TGTTGTGTTTCGTGGACGAGGTGCCTAAAGGACTGACCGGCAAGTTG GACGCCCCGAAGATCCGCGAGATTCTCATTAAGGCCAAGAAGGGCG GAAAGATCGCCGTGTAA
23	Sequence of HPV protein encoding gene	ATGGATGCTATGAAACGGGGCCTGTGCTGCGTGCTGCTCCTGTGCGG CGCTGTGTTTGTGAGCCCTAGCATACCCAGGACTGCTCCTTCCAAC ACAGCCCCATCTCCTCCGACTTCGCTGTCAAAATCCGTGAGCTGTCT GACTACCTGCTTCAAGATTACCCAGTCACCGTGGCCTCCAACCTGCA GGACGAGGAGCTCTGCGGGGGCCTCTGGCGGCTGGTCCTGGCACAG CGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAAG GCTTGCTGGAGCGCGTGAACACGGAGATACACTTTGTACCAAATGT GCCTTTCAGCCCCCCCCAGCTGTCTTCGCTTCGTCCAGACCAACAT

		CTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGAAG CCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGGAGCTGCAGTG TCAGCCCGACTCCTCAACCCTGCCACCCCATGGAGTCCCCGGCCCC TGGAGGCCACAGCCCCGACAGCCCCGGGCGGGCAGCGGCGATAT GCACCAGAAGAGAACCGCCATGTTCCAGGACCCTCAGGAGAGACCT AGGAAGCTGCCTCACCTGTGTACAGAGCTCCAGACAACCATCCACG ACATCATCCTGGAGTGCCTGTACTGTAAGCAGCAGCTGCTGAGAAGA GAGGTGTACGACTTCGCCTTCAGAGACCTGTGCATCGTGTACAGAGA CGGCAACCCTTACGCCGTGTGCGATAAGTGTCTGAAGTTCTATTCCA AAATCTCCGAATATAGGTACATGCACGGCGACACCCCTACCCTGCAC GAGTACATGCTGGACCTCCAGCCTGAGACCACAGACCTGTACTGCTA CGAGCAGCTGAACGACAGCTCTGAGGAAGAGGACGAGATTGACGG ACCTGCTGGCCAGGCCGAGCCTGACAGAGCCCACTACAATATCGTGA CATTCTGTTGCAAATGCGACTCCACACTGGACAAGTGCCTGAAGTTC TACAGCAAGATCTCTGAGTACAGATACTACTGCTACTCTGTGTACGGC ACCACACTGGAGCAGCAGTACAACAAGCCTCTGTGCGACCTCCTGAT CCGCTGCATCAACTGCCAGAAGCCTCTGTGCCCTGAGGAGAAGCAG AGACACCTGGACAAGAAGCAGCGGTTCCACAACATCAGAGGCAGAT GGACCGGCAGGTGCATGTCCTGCTGTAGATCCTCCAGAACCAGACGG GAGACCCAGCTGCACTACAACATCGTGACCTTCTGCTGCAAGTGCGA CTCTACCCTGAGACTGTGCGTGCAGTCTACCCACGTGGACATCAGAA CCCTGGAGGACCTGCTGATGGGCACCCTGGGCATCGTGTGCCCTATC TGCTCTCAGAAGCCTATGGCCAGGTTTCGAGGACCCTACCAGAAGACC CTACAAGCTGCCTGACCTGTGCACCGAGCTGAACACCTCTCTGCAAG ACATCGAGATCACCTGCGTGTACTGCAAGACCGTGCTGGAGCTGACC GAGGTGTTTCGAGTTCGCCTTCAAGGACCTGTTTCGTGGTGTACAGAGA CAGCATCCCTCACGCTGCCTGCCACAAGTGCATCGACTTCTATTCCAG GATCAGGGAGCTGCGCTATTACTCCGACTCTGTGATGTACGGCCCCA AGGCCACCCTCCAGGACATCGTGCTGCACCTGGAGCCTCAGAACGA GATCCCCGTGGACCTGCTGTGCCACGAGCAGCTGTCTGACTCTGAAG AGGAGAACGACGAGATCGACGGCGTGAACCACCAGCACCTGCCTGC CAGGAGAGCTGAACCCAGCGGCATACCATGCTGTGTATGTGCTTCT ACTCTAGGATCAGAGAGCTGAGGTACTACTCTGACTCTGTGTACGGC GACACCCTGGAGAAGCTGACCAACACCGGCCTGTACAACCTGCTGA TCCGGTGCCTGAGGTGCCAGAAGCCTCTGAACCCTGCCGAGAAGCT GAGACACCTGAACGAGAAGAGAAGATTCCACAAGATCGCTGGCCAC TACAGAGGCCAGTGCCACTCTTGCTGCAACAGAGCCAGACAGGAGA GACTCCAGCGGAGAAGGGAGACCCAGGTGGCCAGGAGAGCCGAGC CTCAGAGACACACCATGCTGTGCATGTGCTGCAAGTGCGAGGCCAG AATCGAGCTGGTGGTGGAGAGCTCTGCCGACGACCTGAGAGCCTTC CAGCAGCTGTTCTGTCTACCCTGAGCTTCGTGTGCCCTTGGTGCGC CTCTCAGCAGTAA
24	5' UTR-HPV protein	ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGAC ACCGGATCCGCCACCATGGATGCTATGAAACGGGGCCTGTGCTGCGT

encoding gene-3' UTR-poly (A) P3	GCTGCTCCTGTGCGGCGCTGTGTTTGTGAGCCCTAGCATCACCCAGG ACTGCTCCTTCCAACACAGCCCCATCTCCTCCGACTTCGCTGTCAAA ATCCGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCAGTCACCGTG GCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGCCTCTGGCGGC TGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTGCTGG GTCCAAGATGCAAGGCTTGCTGGAGCGCGTGAACACGGAGATACAC TTTGTACACAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTTCGCTTC GTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCT GGTGGCGCTGAAGCCCTGGATCACTCGCCAGAACTTCTCCCGGTGCC TGGAGCTGCAGTGTGAGCCCGACTCCTCAACCCTGCCACCCCCATGG AGTCCCCGGCCCCCTGGAGGCCACAGCCCCGACAGCCCCGGGCGGCG GCAGCGGCGATATGCACCAGAAGAGAACCGCCATGTTCCAGGACCC TCAGGAGAGACCTAGGAAGCTGCCTCACCTGTGTACAGAGCTCCAG ACAACCATCCACGACATCATCCTGGAGTGCGTGTACTGTAAGCAGCA GCTGCTGAGAAGAGAGGTGTACGACTTCGCCTTCAGAGACCTGTGC ATCGTGTACAGAGACGGCAACCCTTACGCCGTGTGCGATAAGTGTCT GAAGTTCTATTCCAAAATCTCCGAATATAGGTACATGCACGGCGACAC CCCTACCCTGCACGAGTACATGCTGGACCTCCAGCCTGAGACCACAG ACCTGTACTGCTACGAGCAGCTGAACGACAGCTCTGAGGAAGAGGA CGAGATTGACGGACCTGCTGGCCAGGCCGAGCCTGACAGAGCCCAC TACAATATCGTGACATTCTGTTGCAAATGCGACTCCACACTGGACAA GTGCCTGAAGTTCTACAGCAAGATCTCTGAGTACAGATACTACTGCTA CTCTGTGTACGGCACCACTGGAGCAGCAGTACAACAAGCCTCTGT GCGACCTCCTGATCCGCTGCATCAACTGCCAGAAGCCTCTGTGCCCT GAGGAGAAGCAGAGACACCTGGACAAGAAGCAGCGGTTCCACAAC ATCAGAGGCAGATGGACCGGCAGGTGCATGTCCTGCTGTAGATCCTC CAGAACCAGACGGGAGACCCAGCTGCACTACAACATCGTGACCTTC TGCTGCAAGTGCGACTCTACCCTGAGACTGTGCGTGCAGTCTACCCA CGTGGACATCAGAACCCTGGAGGACCTGCTGATGGGCACCCTGGGC ATCGTGTGCCCTATCTGCTCTCAGAAGCCTATGGCCAGGTTTCAGGA CCCTACCAGAAGACCCTACAAGCTGCCTGACCTGTGCACCGAGCTGA ACACCTCTCTGCAAGACATCGAGATCACCTGCGTGTACTGCAAGACC GTGCTGGAGCTGACCGAGGTGTTTCGAGTTCGCCTTCAAGGACCTGTT CGTGGTGTACAGAGACAGCATCCCTCACGCTGCCTGCCACAAGTGCA TCGACTTCTATTCCAGGATCAGGGAGCTGCGCTATTACTCCGACTCTG TGATGTACGGCCCCAAGGCCACCCTCCAGGACATCGTGCTGCACCTG GAGCCTCAGAACGAGATCCCCGTGGACCTGCTGTGCCACGAGCAGC TGTCTGACTCTGAAGAGGAGAACGACGAGATCGACGGCGTGAACCA CCAGCACCTGCCTGCCAGGAGAGCTGAACCCCAGCGGCATACCATG CTGTGTATGTGCTTCTACTCTAGGATCAGAGAGCTGAGGTACTACTCT GACTCTGTGTACGGCGACACCCTGGAGAAGCTGACCAACACCGGCC TGTACAACCTGCTGATCCGGTGCCTGAGGTGCCAGAAGCCTCTGAAC CCTGCCGAGAAGCTGAGACACCTGAACGAGAAGAGAAGATTCCACA AGATCGCTGGCCACTACAGAGGCCAGTGCCACTCTTGCTGCAACAG
---	--

		AGCCAGACAGGAGAGACTCCAGCGGAGAAGGGAGACCCAGGTGGC CAGGAGAGCCGAGCCTCAGAGACACACCATGCTGTGCATGTGCTGC AAGTGCGAGGCCAGAATCGAGCTGGTGGTGGAGAGCTCTGCCGACG ACCTGAGAGCCTTCCAGCAGCTGTTCTGTCTACCCTGAGCTTCGTG TGCCCTTGGTGCGCCTCTCAGCAGTAAGGCGCGCCGCTCGCTTTCTT GCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCTAAGTCCAACACT AAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTGC TCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCTAAG TCCAACACTAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGA TTCTGCCTGAATTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGATATCAAAAAAAAAA AAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAA AAA
25	5' UTR-HPV protein encoding gene-3' UTR-poly (A) P4	ACATTGCTTCTGACACAACCTGTGTTCACTAGCAACCTCAAACAGAC ACCGGATCCGCCACCATGGATGCTATGAAACGGGGCCTGTGCTGCGT GCTGCTCCTGTGCGGCGCTGTGTTTGTGAGCCCTAGCATCACCCAGG ACTGCTCCTTCCAACACAGCCCCATCTCCTCCGACTTCGCTGTCAA ATCCGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCAGTCACCGTG GCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGCCTCTGGCGGC TGGTCCTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTCGCTGG GTCCAAGATGCAAGGCTTGCTGGAGCGCGTGAACACGGAGATACAC TTTGTACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTTCGCTTC GTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCT GGTGGCGCTGAAGCCCTGGATCACTCGCCAGAACTTCTCCCGGTGCC TGGAGCTGCAGTGTGAGCCCGACTCCTCAACCCTGCCACCCCATGG AGTCCCCGGCCCCCTGGAGGCCACAGCCCCGACAGCCCCGGGCGGCG GCAGCGGCGATATGCACCAGAAGAGAACCGCCATGTTCCAGGACCC TCAGGAGAGACCTAGGAAGCTGCCTCACCTGTGTACAGAGCTCCAG ACAACCATCCACGACATCATCTGGAGTGCCTGTACTGTAAGCAGCA GCTGCTGAGAAGAGAGGTGTACGACTTCGCCTTCAGAGACCTGTGC ATCGTGTACAGAGACGGCAACCCTTACGCCGTGTGCGATAAGTGTCT GAAGTTCTATTCCAAAATCTCCGAATATAGGTACATGCACGGCGACAC CCCTACCCTGCACGAGTACATGCTGGACCTCCAGCCTGAGACCACAG ACCTGTACTGCTACGAGCAGCTGAACGACAGCTCTGAGGAAGAGGA CGAGATTGACGGACCTGCTGGCCAGGCCGAGCCTGACAGAGCCCAC TACAATATCGTGACATTCTGTTGCAAATGCGACTCCACACTGGACAA GTGCCTGAAGTTCTACAGCAAGATCTCTGAGTACAGATACTACTGCTA CTCTGTGTACGGCACCACACTGGAGCAGCAGTACAACAAGCCTCTGT GCGACCTCCTGATCCGCTGCATCAACTGCCAGAAGCCTCTGTGCCCT GAGGAGAAGCAGAGACACCTGGACAAGAAGCAGCGGTTCCACAAC ATCAGAGGCAGATGGACCGGCAGGTGCATGTCCTGCTGTAGATCCTC CAGAACCAGACGGGAGACCCAGCTGCACTACAACATCGTGACCTTC TGCTGCAAGTGCGACTCTACCCTGAGACTGTGCGTGCAGTCTACCCA

		CGTGGACATCAGAACCCTGGAGGACCTGCTGATGGGCACCCTGGGC ATCGTGTGCCCTATCTGCTCTCAGAAGCCTATGGCCAGGTTTCGAGGA CCCTACCAGAAGACCCTACAAGCTGCCTGACCTGTGCACCGAGCTGA ACACCTCTCTGCAAGACATCGAGATCACCTGCGTGTACTGCAAGACC GTGCTGGAGCTGACCGAGGTGTTTCGAGTTCGCCTTCAAGGACCTGTT CGTGGTGTACAGAGACAGCATCCCTCACGCTGCCTGCCACAAGTGCA TCGACTTCTATTCCAGGATCAGGGAGCTGCGCTATTACTCCGACTCTG TGATGTACGGCCCCAAGGCCACCCTCCAGGACATCGTGCTGCACCTG GAGCCTCAGAACGAGATCCCCGTGGACCTGCTGTGCCACGAGCAGC TGTCTGACTCTGAAGAGGAGAACGACGAGATCGACGGCGTGAACCA CCAGCACCTGCCTGCCAGGAGAGCTGAACCCAGCGGCATACCATG CTGTGTATGTGCTTCTACTCTAGGATCAGAGAGCTGAGGTACTACTCT GACTCTGTGTACGGCGACACCCTGGAGAAGCTGACCAACACCGGCC TGTACAACCTGCTGATCCGGTGCCTGAGGTGCCAGAAGCCTCTGAAC CCTGCCGAGAAGCTGAGACACCTGAACGAGAAGAGAAGATTCCACA AGATCGCTGGCCACTACAGAGGCCAGTGCCACTCTTGCTGCAACAG AGCCAGACAGGAGAGACTCCAGCGGAGAAGGGAGACCCAGGTGGC CAGGAGAGCCGAGCCTCAGAGACACACCATGCTGTGCATGTGCTGC AAGTGCGAGGCCAGAATCGAGCTGGTGGTGGAGAGCTCTGCCGACG ACCTGAGAGCCTTCCAGCAGCTGTTCCCTGTCTACCCTGAGCTTCGTG TGCCCTTGGTGCCTCTCAGCAGTAAGGCGCGCCGCTCGCTTTCTT GCTGTCCAATTTCTATTAAAGGTTCCCTTGTTCCTAAGTCCAATACT AACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTGC TCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCCTTGTTCCTAAG TCCAATACTAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGA TTCTGCCTGAATTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAGATATCAAAAAAAAAA AAA AAAAA
26	Sequence of HPV protein	MDAMKRGLCCVLLLCGAVFVSPSITQDCSFQHSPISDFAVKIRELSDYL LQDYPVTVASNLQDEELCGGLWRLVLAQRWMERLKTVAGSKMQGLLE RVNTEIHFVTKCAFQPPPSCLRFVQTNISRLQETSEQLVALKPWITRQNF SRCLELQCQPDSSLPPPWSRPLEATAPTAPGGGSGDMHQKRTAMFQD PQERPRKLPHLCTELQTTIHDIILECVYCKQQLLRREVYDFAFRDLCIVY RDGNPYAVCDKCLKFYISKISEYRYMHGDTPTLHEYMLDLQPETTDLYC YEQLNDSSEEEDEIDGPAGQAEPDRAHYNIVTFCKCDSTLDKCLKFY KISEYRYCYSVYGTITLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLD KKQRFHNIRGRWTGRCMSCRSSRTRRETQLHYNIVTFCKCDSTLRCL VQSTHVDIRTLEDLLMGTLGIVCPICSQKPMARFEDPTRRPYKLPDLCTE LNTSLQDIEITCVYCKTVLELTFEFAFKDLFVVYRDSIPHAACHKCID FYSRIRELRYYSDSVMYGPKATLQDIVLHLEPQNEIPVDLLCHEQLSDSE EENDEIDGVNHQHLPARRAEPQRHTMLCMCFYSRIRELRYYSDSVYGD TLEKLTNTGLYNLLIRCLRCQKPLNPAEKLRLHLEKRRFHKIAGHYRGQ

		CHSCCNRARQERLQRRRETQVARRAEPQRHTMLCMCCKCEARIELVVE SSADDLRAFQQLFLSTLSFVCPWCASQQ
--	--	--

CLAIMS

1. An engineered DNA molecule capable of being replicated in a cell, comprising a polyadenosine tail (Poly A tail) coding sequence, wherein the poly (A) tail coding sequence comprises: a single element **a** and at least one element **b**, and at least one element **c** and/or at least one element **d**:

the element **a** consists of a plurality of consecutive adenine (A) nucleotides, and the length range of the element **a** is ≥ 20 nt;

the element **b** consists of a plurality of consecutive A nucleotides, and the length range of the element **b** is $3 \text{ nt} \leq \mathbf{b} < 20 \text{ nt}$;

the element **c** consists of a non-A nucleotide, and the nucleotide is selected from T, C and G nucleotides;

the element **d** consists of any two or more consecutive nucleotides, and the nucleotides are selected from A, T, C and G nucleotides, wherein the nucleotides at a 5' terminus and a 3' terminus of the element **d** are not A nucleotides, and the element **d** does not comprise 3 or more consecutive A nucleotides; the length range of the element **d** is $2 \text{ nt} \leq \mathbf{d} \leq 20 \text{ nt}$;

wherein the element **a** and the element **b** are not adjacent, and the element **c** and the element **d** are not adjacent, and

the poly (A) tail coding sequence does not comprise any two elements **b** that are adjacent to each other, does not comprise any two elements **c** that are adjacent to each other, and does not comprise any two elements **d** that are adjacent to each other.

2. The DNA molecule according to claim 1, wherein the length of the poly (A) tail coding sequence is 101-200 nt, 101-150 nt, 120-150 nt, 130-140 nt, 120-135 nt or 123-125 nt.

3. The DNA molecule according to claim 1 or 2, wherein the 3' terminus of the poly (A) tail coding sequence is an A nucleotide or a non-A nucleotide.

4. The DNA molecule according to any one of claims 1-3, wherein the length of the element **a** is ≤ 80 nt.

5. The DNA molecule according to any one of claims 1-3, wherein the element **a** is 30-70 nt, 35-65 nt, 40-60 nt or 45-55 nt in length, preferably 60 nt in length.

6. The DNA molecule according to any one of claims 1-5, wherein 50% or more of the polynucleotides of element **a** are located in the 5' portion or 3' portion of the poly (A) tail coding

sequence.

7. The DNA molecule according to any one of claims 1-6, wherein the element **b** is 3-10 nt, 10-19 nt, 12-15 nt, 14-17 nt or 16-19 nt in length, preferably 19 nt in length.

8. The DNA molecule according to any one of claims 1-7, wherein the number of the elements **b** is 2-10, preferably 2-5, and more preferably 3.

9. The DNA molecule according to any one of claims 1-8, wherein the element **c** is G.

10. The DNA molecule according to any one of claims 1-9, wherein the number of the elements **c** is 2-10, 3-8, 4-6 or 2-5, preferably 2.

11. The DNA molecule according to any one of claims 1-10, wherein the element **d** comprises a palindromic sequence.

12. The DNA molecule according to any one of claims 1-11, wherein the element **d** is 3-18 nt, 5-16 nt, 4-10 nt or 6-12 nt in length, preferably 6 nt in length.

13. The DNA molecule according to any one of claims 1-12, wherein the element **d** is any one or more selected from the following sequences: GATATC (SEQ ID NO: 15), GTATAC (SEQ ID NO: 16), GAATCT (SEQ ID NO: 17), GCATATGACT (SEQ ID NO: 18) and GATATCGTATAC (SEQ ID NO: 19).

14. The DNA molecule according to any one of claims 1-13, wherein the element **d** is any one or more selected from the following nucleotide sequences: SEQ ID NO: 15, SEQ ID NO: 16, and SEQ ID NO: 17.

15. The DNA molecule according to any one of claims 1-14, wherein the nucleotide sequence of the element **d** is represented by SEQ ID NO: 15.

16. The DNA molecule according to any one of claims 1-15, wherein the number of the element **d** is 0-5, preferably 1-3, and more preferably 1.

17. The DNA molecule according to any one of claims 1-16, wherein when the element **c** and the element **d** exist simultaneously, the total number of the element **c** and the element **d** is 2-15, preferably 3-5, and more preferably 3.

18. The DNA molecule according to any one of claims 1-17, wherein the 3' portion of the poly (A) tail coding sequence, preferably the 1/2 portion of the poly (A) tail coding sequence near the 3' terminus, comprises one or more non-A nucleotides.

19. The DNA molecule according to any one of claims 1-18, wherein the structure of the poly (A) tail coding sequence is:

element **a**-element **c**-element **b**-element **c**-element **b**-element **c**-element **b**-element **c**-element **b**;

element **b**-element **c**-element **b**-element **c**-element **a**-element **d**-element **b**-element **c**-element **b**-element **c**-element **b**;

element **b**-element **c**-element **b**-element **c**-element **b**-element **d**-element **a**-element **c**;

element **a**-element **d**-element **b**-element **c**-element **b**-element **c**-element **b**; or

element **b**-element **c**-element **b**-element **c**-element **b**-element **d**-element **a**.

20. The DNA molecule according to any one of claims 1-19, wherein the structure of the poly (A) tail coding sequence is:

element **a**-element **d**-element **b**-element **c**-element **b**-element **c**-element **b**; the element **a** is 60 nt in length, the element **b** is 16-19 nt in length, and the element **d** is 6 nt in length.

21. The DNA molecule according to any one of claims 1-20, wherein the structure of the poly (A) tail coding sequence is:

element **b**-element **c**-element **b**-element **c**-element **b**-element **d**-element **a**; and the element **a** is 60 nt in length, the element **b** is 16-19 nt in length, and the element **d** is 6 nt in length.

22. The DNA molecule according to any one of claims 1-21, wherein the poly (A) tail coding sequence is represented by any one of SEQ ID NOs: 1-10.

23. The DNA molecule according to any one of claims 1-22, wherein the poly (A) tail coding sequence is represented by SEQ ID NO: 3 or SEQ ID NO: 4.

24. The DNA molecule according to any one of claims 1-23, wherein it is further connected to a gene of interest fragment on the 5' end of the poly (A) tail coding sequence, and the gene of interest fragment and the poly (A) tail coding sequence co-encode RNA.

25. The DNA molecule according to any one of claims 1-24, further comprising a replicon.

26. The DNA molecule according to any one of claims 1-25, further comprising a resistance gene.

27. The DNA molecule according to any one of claims 1-26, further comprising a promoter for initiating transcription of the RNA.

28. The DNA molecule according to any one of claims 1-27, wherein the gene of interest fragment comprises a 5' UTR coding sequence.

29. The DNA molecule according to any one of claims 1-28, wherein the gene of interest fragment comprises a protein coding sequence or a non-protein coding sequence.

30. The DNA molecule according to any one of claims 1-29, wherein the gene of interest fragment comprises a 3' UTR coding sequence.

31. The DNA molecule according to any one of claims 1-30, comprising a replicon, an antibiotic resistance gene, a promoter, a 5' UTR coding sequence, a protein coding sequence, and a 3' UTR coding sequence.

32. The DNA molecule according to any one of claims 1-31, wherein the protein coding sequence encodes an HPV (human papillomavirus) protein, preferably the HPV protein is derived from HPV type 16 and/or type 18.

33. The DNA molecule according to any one of claims 1-32, wherein the protein coding sequence encodes HPV E2, E6 or E7 protein, a fusion protein of E6 and E7 protein polypeptide fragments, or a fusion protein of E2, E6 and E7 protein polypeptide fragments, preferably the HPV protein is derived from HPV type 16 and/or type 18.

34. The DNA molecule according to any one of claims 1-33, wherein the protein coding sequence encodes the polypeptide represented by SEQ NO: 26.

35. The DNA molecule according to any one of claims 1-34, comprising a polynucleotide sequence represented by any one of SEQ ID NOs: 22-25, or a synonymous mutant of the polynucleotide sequence represented by any one of SEQ ID NOs: 22-25, or a polynucleotide sequence sharing more than 85% sequence identity with the polynucleotide sequence represented by any one of SEQ ID NOs: 22-25 or a synonymous mutant thereof.

36. The DNA molecule according to any one of claims 1-35, wherein the DNA molecule is a DNA plasmid.

37. A cell comprising the DNA molecule according to any one of claims 1-36.

38. The cell according to claim 37, wherein the cell is a prokaryotic cell.

39. The cell according to claim 37, wherein the cell is *Escherichia coli*.

40. An RNA molecule encoded by the DNA molecule according to any one of claims 1-36.

41. The RNA molecule according to claim 40, further comprising a 5'-cap structure, and/or some or all of the uridines in the RNA are chemically modified uridines; preferably, some or all of the uridines in the RNA are pseudouridines or 1-methyl-pseudouridines.

42. A DNA coding sequence for the poly (A) tail according to any one of claims 1-36.

43. A poly (A) tail sequence encoded by the DNA coding sequence for the poly (A) tail according to any one of claims 1-36.

44. Use of the DNA coding sequence for a poly (A) tail according to claim 42 for making the replication of a DNA molecule encoding an RNA more conservative in a host cell, wherein the poly (A) tail is located at the 3' terminus of the RNA.

45. The use according to claim 44, wherein the host cell is a prokaryotic cell, preferably *E. coli*.

46. Use of the poly (A) tail sequence according to claim 43 for regulating the expression of an RNA molecule in a host cell, wherein the poly (A) tail is located at the 3' terminus of the RNA.

47. The use according to claim 46, wherein the host cell is a eukaryotic cell, preferably a mammalian cell, and more preferably a human cell.

48. A library comprising the DNA molecule according to any one of claims 1-36.

49. A library comprising an RNA molecule encoded by the DNA molecule according to any one of claims 1-36.

50. A Method for regulating a protein expression, comprising:

introducing a plurality of DNAs in the DNA library according to claim 48 into target cells at different times and/or in different quantitative ratios; or

introducing a plurality of RNAs in the RNA library according to claim 49 into target cells at different times and/or in different quantitative ratios.

51. A hybrid molecule of DNA and RNA, comprising the same genetic information as the DNA molecule according to any one of claims 1-36, the same genetic information as the poly (A) tail according to claim 42, or the same genetic information as the RNA molecule according to claim 40.