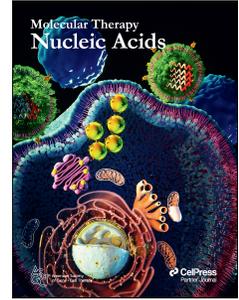


Journal Pre-proof

High-accuracy crRNA array assembly strategy for multiplex CRISPR

Xiangtong Zhao, Lixian Yang, Peng Li, Zijing Cheng, Yongshi Jia, Limin Luo, Aihong Bi, Hanchu Xiong, Haibo Zhang, Hongen Xu, Jinrui Zhang, Yaodong Zhang



PII: S2162-2531(24)00315-9

DOI: <https://doi.org/10.1016/j.omtn.2024.102428>

Reference: OMTN 102428

To appear in: *Molecular Therapy: Nucleic Acid*

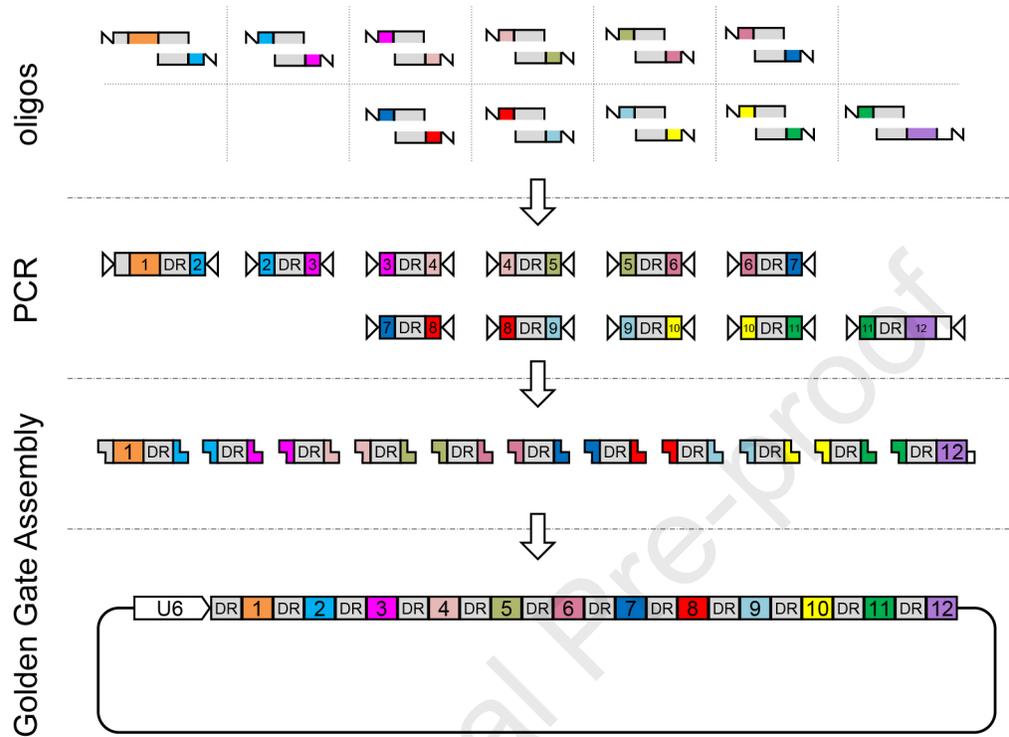
Received Date: 5 December 2023

Accepted Date: 10 December 2024

Please cite this article as: Zhao X, Yang L, Li P, Cheng Z, Jia Y, Luo L, Bi A, Xiong H, Zhang H, Xu H, Zhang J, Zhang Y, High-accuracy crRNA array assembly strategy for multiplex CRISPR, *Molecular Therapy: Nucleic Acid* (2025), doi: <https://doi.org/10.1016/j.omtn.2024.102428>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier Inc. on behalf of The American Society of Gene and Cell Therapy.



1 High-accuracy crRNA array assembly strategy for multiplex CRISPR

2 Xiangtong Zhao,^{1,2,6} Lixian Yang,^{3,6} Peng Li,⁴ Zijing Cheng,⁵ Yongshi Jia,³ Limin Luo,³ Aihong Bi,³ Hanchu
3 Xiong,³ Haibo Zhang,³ Hongen Xu,³ Jinrui Zhang¹ and Yaodong Zhang¹

4 ¹Henan Provincial Key Laboratory of Children's Genetics and Metabolic Diseases, Children's Hospital Affiliated to Zhengzhou University,
5 Henan Children's Hospital, Zhengzhou Children's Hospital, Zhengzhou, Henan, China; ²College of Public Health, Zhengzhou University,
6 Zhengzhou, Henan, China; ³Cancer Center, Department of Radiation Oncology, Zhejiang Provincial People's Hospital, Affiliated People's
7 Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, China; ⁴Department of Gastroenterology, Zhejiang Provincial People's
8 Hospital, Affiliated People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, China; ⁵School of Anesthesiology, Xuzhou
9 Medical University, Xuzhou, Jiangsu, China

10

11 ⁶These authors contributed equally

12 Correspondence: Yaodong Zhang, Henan Provincial Key Laboratory of Children's Genetics and Metabolic Diseases, Children's Hospital
13 Affiliated to Zhengzhou University, Henan Children's Hospital, Zhengzhou Children's Hospital, Zhengzhou, Henan, China

14 E-mail: syek@163.com

15 Correspondence: Xiangtong Zhao, Henan Provincial Key Laboratory of Children's Genetics and Metabolic Diseases, Children's Hospital
16 Affiliated to Zhengzhou University, Henan Children's Hospital, Zhengzhou Children's Hospital, Zhengzhou, Henan, China

17 E-mail: zhaoxiangtong@zuua.zju.edu.cn

18 Correspondence: Lixian Yang, Cancer Center, Department of Radiation Oncology, Zhejiang Provincial People's Hospital, Affiliated
19 People's Hospital, Hangzhou Medical College, Hangzhou, Zhejiang, China

20 E-mail: yanglixian@hmc.edu.cn

21

22 ABSTRACT

23 Simultaneous targeting of multiple loci with the clustered regularly interspaced short palindromic repeats (CRISPR) system, a
24 tool known as multiplex CRISPR, offers greater feasibility for manipulating and elucidating the intricate and redundant
25 endogenous networks underlying complex cellular functions. Owing to the versatility of continuously emerging CRISPR-
26 associated (Cas) nucleases and the utilization of CRISPR arrays, multiplex CRISPR has been implemented in numerous *in*
27 *vitro* and *in vivo* studies. However, a streamlined, practical strategy for CRISPR array assembly that is both convenient and
28 accurate is lacking. Here, we present a novel, highly accurate, cost-, and time-saving strategy for CRISPR array assembly.
29 Using this strategy, we efficiently assembled 12 CRISPR RNAs (crRNAs) (for AsCas12a) and 15 crRNAs (for RfxCas13d) in
30 a single reaction. CRISPR arrays driven by Pol II promoters exhibited a distinct expression pattern compared to those driven
31 by Pol III promoters, which could be exploited for specific distributions of CRISPR intensity. Improved approaches were
32 subsequently designed and validated for expressing long CRISPR arrays. The study provides a flexible and powerful tool for
33 the convenient implementation of multiplex CRISPR across DNA and RNA, facilitating the dissection of sophisticated cellular
34 networks and the future realization of multi-target gene therapy.

35

36 INTRODUCTION

37 Over the past decade, following the pioneering discovery and eukaryotic implementation of the clustered regularly interspaced
38 short palindromic repeats (CRISPR)/Cas9 system,¹⁻³ a growing number of new CRISPR systems have been developed.⁴ As a
39 substrate of the best-known Cas9, DNA is no longer the only target within our reach. With the discovery of various RNA-
40 targeting Cas nucleases,⁵⁻¹¹ RNA can now be precisely targeted, further extending our range of genetic manipulation from the
41 genome to the transcriptome. Moreover, by combining transcriptional activators,¹² repressors,¹³ base editors,¹⁴ epigenetic
42 enzymes,¹⁵ reverse transcriptases,¹⁶ and tagging molecules,¹⁷ these RNA-guided effectors have magnified the molecular
43 toolbox of basic research. Most importantly, the CRISPR systems have demonstrated unlimited potential for the treatment of a
44 growing number of incurable genetic diseases.¹⁸⁻²³

45 As indicated by the definition of CRISPR, the natural CRISPR RNA (crRNA) in bacteria and archaea is usually not
46 transcribed individually but in a clustered preform from a genomic locus called the CRISPR array, which consists of a
47 succession of direct repeats (DRs) separated by distinct spacers.⁴ Subsequently, the transcribed pre-crRNA is processed into a

48 series of mature crRNAs, each containing a single spacer and a truncated DR. Unlike Cas9, which requires an additional
49 transactivating CRISPR RNA to form a final guide RNA with the crRNA (usually replaced in practice by a chimeric single
50 guide RNA), many other CRISPR systems use mature crRNAs directly as guide RNAs. Moreover, the Cas nucleases in these
51 systems typically possess the ability to process pre-crRNAs.⁴ These two features render these Cas effectors more amenable to
52 multiplex CRISPR, because a single array is sufficient to express all the required crRNAs.

53 Current strategies for assembling customized CRISPR arrays are based on the conventional method of cloning a single
54 crRNA, except that the number of annealed oligos is multiplied. In fact, such theoretically feasible strategies may work when
55 assembling only a few crRNAs, but are quite incompetent when assembling more crRNAs, as demonstrated here. Therefore,
56 an accurate, efficient, and practical strategy for the CRISPR array assembly is required. In the present study, we designed a
57 novel cost-effective strategy for CRISPR array assembly and confirmed its superior accuracy over current alternatives using
58 the AsCas12a system. While preserving its accuracy as much as possible, several simplifications and optimizations were made
59 to make this strategy more user-friendly. Next, this strategy was used to assemble crRNAs for an RNA-targeting CRISPR
60 system, highlighting its general applicability, flexibility, and high accuracy. When attempting to harness Pol II promoters for
61 CRISPR array expression, a distinct expression pattern of CRISPR arrays transcribed by EF1a (Pol II promoter) was found,
62 compared to the more commonly used U6 (Pol III promoter). For long CRISPR arrays (>200 nt), we designed a better approach
63 based on U6 by using its edges and circumventing its drawbacks. These tandem, hierarchical arrays achieved improved
64 targeting efficiency in CRISPR systems, which require the abundant expression of crRNAs. Finally, by re-examining the
65 strategy of co-expressing Cas nuclease and the CRISPR array on a single transcript, an unavoidable detrimental effect on the
66 expression of Cas nuclease was discovered, despite its unexpectedly satisfactory targeting performance with some CRISPR
67 systems in certain applications. Further investigation of this anomaly indicated that the introduction of an upstream GFP-coding
68 sequence could enhance the expression of Pol II promoter-driven CRISPR arrays. This finding provides an alternative approach
69 for efficient expression of long CRISPR arrays.

70 RESULTS

71 Design of a novel strategy for CRISPR array assembly

72 Given that it is more convenient and cost-efficient when probing targeting efficiency, a catalytically dead AsCas12a fused to
73 an artificial VP64-p65-Rta transcription activator (dAsCas12a-VPR, abbreviated as d12a-VPR hereafter) was used for
74 subsequent experiments. An enhanced variant (denAsCas12a-VPR, abbreviated as den12a-VPR hereafter) was also used.²⁴

75 Similar to Cas9-based CRISPR activators,²⁵⁻²⁸ the combination of multiple crRNAs enabled the synergistic activation of
76 endogenous gene targets mediated by the d12a-VPR and den12a-VPR (Figure S1). Next, we assessed the accuracy of different
77 strategies for the assembly of crRNAs for AsCas12a. Current strategies for the assembly of multiple crRNAs typically consist
78 of two main procedures (Figure S2): annealing of pre-designed, single-stranded DNA oligos to form double-stranded DNA
79 (dsDNA) with the desired sticky ends, followed by sequential ligation into cloning vectors. Because of their similar principles,
80 we classified them as sticky end-based strategies. The accuracy of these assemblies relies on the precise sequential ligation of
81 the successfully annealed oligos. In practice, however, not all oligos eventually end up annealed to their respective partners as
82 desired, i.e., a fair number of oligos remain single-stranded after annealing. These single-stranded oligos can be a source of
83 trouble, as their natively-exposed ends are identical to the sticky ends of the annealed dsDNA, meaning that any site of
84 sequential ligation can be occupied by them, leading to irreversible premature termination of the assembly. This intrinsic defect
85 is likely the dominant factor limiting the number of crRNAs that can be assembled into an array in a single reaction using sticky
86 end-based strategies. The preliminary experiments suggested that only up to six crRNAs could be efficiently assembled in a
87 single assembly (data not shown).

88 To overcome this limitation, we designed a novel strategy for CRISPR array assembly that aimed to increase both accuracy
89 and maximal number by eliminating the perturbations of single-strand oligos (Figure 1). In brief, several additional bases
90 containing a BsaI recognition site were appended to the 5' end of each single-strand oligo, so that the designated inner sticky-
91 end was not exposed until a dsDNA was formed and cut by BsaI (Figure S3). These short dsDNA segments can be generated
92 mainly by two alternative methods: (1) annealing of complete complementary oligo pairs, which are particularly long because
93 of the need for BsaI recognition sites at both ends of the resulting dsDNA, and (2) a polymerase chain reaction (PCR)-based
94 approach using much shorter oligo pairs that are partially complementary to each other or to a certain template. We chose the
95 latter because it is cost-effective and an annealing-free strategy could be tested that would be substantially different from
96 conventional sticky end-based methods.

97 No additional procedures follow the PCR reaction, only optional routine recovery of DNA segments, and subsequent
98 standard Golden Gate Assembly. To this end, we named this novel approach a Golden Gate Assembly (GGA)-based strategy.

99

100 **High-accuracy assembly of CRISPR array with GGA-based strategy**

101 There are several versions of detailed conventional sticky end-based protocols which differ slightly from each other
102 (with/without predigestion before ligation or redigestion after ligation). Their accuracy was compared by assembling six
103 crRNAs. Whereas no correct clones were obtained when the ligation mixture was directly transformed into *Escherichia coli*,
104 additional redigestion with BsaI between ligation and transformation dramatically increased the accuracy of the selected clones
105 (Figure S4). Compared to the remaining two procedures, the assembly using standard Golden Gate cycles exhibited
106 improvement or non-inferiority without the procedure of predigesting or recovering the cloning vector (Figure S4); hence, this
107 specified sticky end-based strategy was used for the following evaluation and comparison.

108 Next, using our GGA-based strategy, an array containing six crRNAs was assembled with parallel controls using a sticky
109 end-based strategy for comparison. Whereas all 60 clones were correct when using the GGA-based strategy, assembly using
110 the sticky end-based strategy achieved a mean accuracy of only 37% ($P < 0.0001$, GGA versus sticky end) (Figure 2A, S5A).
111 When the number of crRNAs was increased to seven, the accuracy of the sticky end-based assembly strategy decreased sharply,
112 with only one correct out of 60 clones. Although a marked reduction was observed, the GGA-based assembly strategy retained
113 a mean accuracy of 73% (Figure 2B, S5B).

114 Having demonstrated the superiority of our GGA-based strategy over conventional sticky end-based strategies, the number
115 of crRNAs to be assembled was increased until the accuracy decreased to a level substantially lower than 50%. Following this
116 criterion, and considering that the extremely low accuracy in assembling seven crRNAs was already practically meaningless,
117 the sticky end-based strategy was no longer evaluated for assembling more crRNAs. Using the GGA-based strategy, arrays of
118 9 or 12 crRNAs were efficiently assembled, with mean accuracies of 67% and 43%, respectively (Figure 2C, S5C). For the
119 assembly of the 12 crRNAs, colonies from an additional replicate experiment were sequenced to confirm our results and
120 investigate the source of errors in incorrect inserts during the revision of this manuscript (Plasmid #37; Table S22). We found
121 that 14/15 full length inserts obtained the desired sequences with no errors (Figure S6). A variety of unusual constructs were
122 seen for inserts of small size, with one particular error arising from the apparent misassembly of the fusion site from fragment
123 “GGGTCTCCGACTGCCACAAGTGCTAATTCCTACTCTTGTAGGTAATGAATGTGTGCGGAGACC” with fragment
124 “GGGTCTCCTAGCCAGCCAATTCCTACTCTTGTAGGTAAGTCCAGGAGACC.” It is unclear whether these errors
125 represent mistakes generated during golden gate assembly owing to *in-vivo* repair events or a mixture of the two.

126 Although arrays containing more than 12 crRNAs may still be assembled in a single reaction, in view of the gradual
127 decrease in assembly accuracy as the number of crRNAs increased, the practical value of such experiments would be limited;

128 thus, we ceased here and determined the maximum number of crRNAs for AsCas12a that could be assembled into an array in
129 a single reaction to be 12.

130 The resulting array of 12 crRNAs contained spacers targeting the promoters of six endogenous human genes (two crRNAs
131 for each gene): *HBB*, *HBG1*, *IL1B*, *IL1RN*, *TTN*, *RHOXF2*. Together with this array, the den12a-VPR achieved robust
132 activation of these six genes simultaneously in HEK293T cells from 20–2000-fold (Figure 2D).

133

134 **Simplification and optimization of the GGA-based strategy for CRISPR array assembly**

135 Although our GGA-based assembly strategy exhibited greater accuracy than the alternatives, the requirement for an additional
136 recovery procedure for PCR products may hinder its wider application. To address this, we sought to circumvent this procedure
137 while preserving assembly accuracy as much as possible. Instead of the time-consuming recovery required to obtain purified
138 dsDNA segments in the standard assembly protocol, the PCR products were diluted and added directly to the Golden Gate
139 assembly reaction. From a wide range of PCR primer concentrations and volumes of PCR products pipetted for subsequent
140 assembly, the best accuracy was achieved when using a final primer concentration of 2.5 μ M for PCR and 0.1 μ L (i.e., 1 μ L
141 after 10-fold dilution) from each of the resulting PCR mixtures for a 20- μ L assembly reaction (for the designated DNA
142 polymerase; data not shown). Assemblies of 6, 7, 9, and 12 crRNAs were performed following this reaction condition. The
143 mixing of unwanted PCR components into the subsequent assembly reaction compromised the final accuracy, but only to a
144 modest degree, with the exception of the assembly of 12 crRNAs, from which the correct clones were barely obtained (Figure
145 S7). In addition, assembly failures could also be attributed to erroneous amplification products (e.g., primer dimers), which
146 could be actively assembled because they also contain BsaI sites that produce compatible overhangs. Nevertheless, this
147 simplified recovery-free version of our GGA-based strategy outperformed current alternative assembly strategies without the
148 requirement for additional procedures and could be a rational choice when assembling no more than nine crRNAs.

149 Another shortcoming that may complicate the assembly strategy is the DR sequence of AsCas12a. The 19nt DR of
150 AsCas12a has a relatively low melting temperature (44–45 °C, T_m) due to its low GC content (~26%). This intrinsic property
151 makes it challenging to design initial primers for subsequent PCR reactions, since oligos must be extended beyond the DR
152 sequence to obtain a minimal appropriate melt temperature (T_m ; 50–55 °C). To address this issue, we sought to introduce
153 mutations into the original wild-type DR, with the aim of increasing its T_m value to a proper extent (> 50 °C) without

154 compromising its function. Several previous studies have made such attempts but with distinct aims.²⁹⁻³¹ Based on their findings,
155 we generated three variants of wild-type DR (Figure 3A). Of these candidates, one variant with a T_m of 49 °C showed non-
156 inferiority in targeting efficiency compared to the wild-type DR (Figure 3B). Although the T_m of this variant was still slightly
157 below 50 °C, it was already feasible to use the DR sequence directly as the end of primers and 45 °C as the annealing
158 temperature for subsequent PCR reactions simplified the primer design procedure (Figure 3C). Based on this DR variant, arrays
159 of nine and 12 crRNAs were successfully assembled, with mean accuracies of 75% and 15%, respectively (Figure 3D, S8).
160 Moreover, the assembly of an array containing nine crRNAs was accomplished efficiently when both mutant DR and the
161 recovery-free version of the GGA-based strategy were used (Figure S9).

162

163 **High-accuracy assembly of CRISPR array for a Cas13d nuclease**

164 Whereas the previous content focused on AsCas12a, the potential applications of the assembly strategy can be extended beyond
165 this specific CRISPR system. Owing to its flexibility and generalizability, this crRNA array assembly strategy can theoretically
166 applied to most CRISPR systems in which crRNA contains a DR followed by a spacer, especially when the Cas nuclease has
167 the ability to process its own crRNA.

168 To confirm this concept, we used this crRNA array assembly strategy to assemble crRNAs from another distinct CRISPR
169 system, Cas13d, a family of RNA-targeting Cas nucleases.⁴ Of these, RfxCas13d was chosen for subsequent assessment.⁹ The
170 commonly-used crRNA for RfxCas13d is composed of a 36nt DR and a spacer of 22–30 nt. An intermediate spacer length of
171 26 nt was selected for the following CRISPR array assembly. In contrast to that of AsCas12a, the DR of RfxCas13d was much
172 longer and had a modest GC content (50%). These properties render the assembly of its crRNAs more suitable for our strategy:
173 (1) design of the primers was much easier than that of AsCas12a, as all the forward and reverse primers could use the same
174 ~20 base 3' section; (2) using current alternatives for such a CRISPR system with long DR sequence would be more costly,
175 since it would be unavoidable to purchase particularly long oligos (>60 bases), which are usually several times more expensive
176 per base than the shorter routine oligos due to their complex production process. The detailed procedure was the same as that
177 for the assembly of crRNAs for AsCas12a, with two minor modifications (Figure S10). The resulting procedure was reminiscent
178 of several previous strategies for assembling sgRNA cassettes used in Cas9-based multiplex CRISPR.³²⁻³⁴

179 The GGA-based assembly strategy performed better with RfxCas13d than with AsCas12a, as an array of up to 15 crRNAs
180 could be assembled in one pot with acceptable accuracy (Figure 4A, S11). This improved performance may be due to purer
181 dsDNA segments or other uncertain reasons. Using this array, the simultaneous cleavage of 15 endogenous transcripts was
182 attempted. Forty-eight hours after transfection, most of these transcripts were efficiently cleaved and detected using the
183 corresponding primers flanking the cleavage sites (Figure 4B).

184

185 **Distinct expression patterns of Pol II/III promoter-driven CRISPR arrays**

186 Although the array of 15 crRNAs used for RfxCas13d reached a length of ~1200nt, the U6 promoter still seemed capable of
187 driving its transcription. This disagreed with the conventional conception that Pol III promoters (for example, U6, H1, and 7SK)
188 are typically used for the transcription of small RNAs, although RNA up to 800 nucleotides have also been reported to be
189 efficiently transcribed by U6.³⁵ One solution to this indeterminate restriction was to replace Pol III promoters with Pol II
190 promoters (e.g., CMV and EF1a), as they are capable of transcribing much longer RNAs. Several studies have evaluated the
191 ability of harnessing Pol II promoters to express crRNAs or synthetic sgRNAs.^{36,37}

192 To determine whether CRISPR arrays transcribed by Pol II promoters work as efficiently as those transcribed by Pol III
193 promoters, HEK293T cells were transfected with den12a-VPR and an array of 12 crRNAs driven by either U6 (U6-array) or
194 EF1a (EF1a-array), followed by quantification of the expression of the corresponding six targets 48 h post-transfection. No
195 straightforward determination of superiority or inferiority could be drawn from the results, because each had its own strengths
196 and weaknesses (Figure 5A). Northern blots of mature crRNAs showed that U6 outperformed EF1a in transcribing the first
197 four crRNAs, particularly the first two, but was inferior for all subsequent crRNAs (Figure 5B). When targeting individual
198 genes using two crRNAs, U6 outperformed EF1a for both *HBB* and *RHOXF2* (Figure 5C, D), suggesting that U6 was superior
199 to EF1a when driving the expression of short CRISPR arrays. Because we could not rule out the possibility that the crRNAs
200 expressed from U6 cells were already saturated, the disparity observed here might have been underestimated. Indeed, a larger
201 gap was observed by varying the ratio of den12a-VPR to the array (Figure S12). However, no disparity between these promoters
202 was observed during gene editing using nuclease-active AsCas12a (Figure 5E). The reason might be that the number of targets
203 in the genome was limited, and once edited with an indel, CRISPR elements were no longer needed, meaning that a small
204 amount of crRNAs was already sufficient, and more would be redundant.

205 To determine whether this phenomenon represented a general pattern, we compared it to other CRISPR systems. Forty-
206 eight hours after transfection with RfxCas13d and an array of 15 crRNAs driven by either U6 or EF1a, no conclusions could
207 be drawn except that the U6-array showed superiority in targeting the first four transcripts (Figure S13A). When harnessed for
208 the expression of a single crRNA, EF1a exhibited a substantially lower target cleavage efficiency than U6 (Figure S13B).

209 In contrast to the EF1a-array, the location of the crRNAs in the U6-array appeared to have a greater influence on their
210 own expression. When the internal order of the CRISPR array was reversed, the U6-array experienced a more drastic change
211 in multiplex CRISPR efficiency (Figure S14). To obtain an intuitive and precise correlation between transcriptional strength
212 and distance from the promoter, the two crRNAs targeting the HBB promoter were placed at a series of different loci while the
213 overall component was kept constant (Figure 5F, G). The first few crRNAs of the U6-array achieved very high levels of
214 expression, which could not be reached by EF1a. Thereafter, the intensity of transcription decreased continuously with
215 increasing distance from U6. In contrast, the distance from the EF1a promoter had a much weaker effect on crRNA expression,
216 except that the first few crRNAs might undergo inadequate expression.

217 These results suggest that Pol II and III promoters are not functionally equivalent when driving the expression of CRISPR
218 arrays. While U6 favored the expression and execution of the first few crRNAs within ~200 nt, EF1a seemed to distribute its
219 relatively milder strength more evenly. For single crRNA or small CRISPR arrays (<200 nt), especially when using the
220 RfxCas13d system, harnessing Pol II promoters may still be an inadequate alternative unless mild expression was acceptable.

221 **Improve the targeting efficiency by manipulating the architecture of long CRISPR arrays**

222 Having determined the superiority of U6 over EF1a when transcribing the first few crRNAs within ~200nt, we decided to find
223 a better approach for the transcription of longer CRISPR arrays than simply selecting one from either U6 or EF1a. By splitting
224 long CRISPR arrays into several short arrays, each of them could be assigned a U6 promoter, thus using the edge of U6 in the
225 strength of transcribing short RNAs while circumventing its disadvantage in transcribing long RNAs. The principle and
226 workflow were similar to those previously described in the current study, with the only difference being that small PCR
227 segments containing crRNAs and longer PCR segments containing U6 were mixed and assembled in one reaction (Figure S15).
228 However, we failed to assemble the correct array that matched our expectations for our maiden attempt (data not shown). This
229 may be attributed to the introduction of repeated U6 elements, which make the resulting array more prone to recombination
230 events.

231 In previous studies, we amplified an array of 12 or 15 crRNAs as an intact sequence from a purified plasmid that already
232 contained the entire array (considered the second round of PCR) and then cloned this unique insert into a new vector (considered
233 the second round of assembly), which is a simple procedure to transfer an existing array from one vector to another. Although
234 the accuracy of assembling 12 or 15 individual crRNAs into one array in the first round was as low as 15% (Figure 3D), the
235 second round always achieved an accuracy >80% (data not shown). Although we were unable to provide an explicit
236 interpretation of this phenomenon, we exploited it by performing a second round of assembly to reboot the unsuccessful
237 assembly of the hierarchical CRISPR array. For the sake of time, the assembly mixture of the first round was directly used as
238 the PCR template for the second round instead of the validated and purified plasmid (Figure S16). The mean accuracy of the
239 array assembly with 12 crRNAs improved to 88% (Figure S17). However, amplifying the entire array from the crude mixture
240 of previous assembly may not always succeed, especially when the destination fragment spans more than a dozen crRNAs. An
241 array of 20 crRNAs was successfully assembled in two rounds; however, we performed two individual amplifications with two
242 pairs of primers in the second round (Figure S17).

243 To improve accuracy by performing an additional round of assembly, we decided to confront the obstacle of assembling
244 hierarchical CRISPR arrays again. In the second round, an RfxCas13d CRISPR array consisting of four U6-crRNAs units was
245 successfully assembled with mean accuracy of 33% (Figure S18B), which had been shown to be impossible in the first round.
246 Similarly, this strategy substantially increased the accuracy of the denAsCas12a CRISPR array assembly consisting of three
247 U6-crRNAs units (Figure S18A).

248 In contrast to arrays driven by a single U6 or EF1a, the tandem array driven by three U6 promoters achieved robust
249 transcriptional activation of all six targets in the denAsCas12a-VPR mediated multiplex CRISPR system (Figure 6A, B).
250 Similarly, for most of the 15 endogenous transcripts, RfxCas13d showed more prominent cleavage efficiency when working
251 with a tandem array driven by four U6 (Figure 6C, D).

252 **Enhance the expression of Pol II promoter-driven CRISPR arrays by introducing an upstream GFP-coding sequence**

253 Since Pol II promoters are capable of driving the expression of CRISPR arrays in addition to protein-encoding genes, one might
254 consider the possibility of co-expressing the Cas protein and CRISPR array on a single transcript. To achieve this type of co-
255 expression, the coding sequence of the Cas protein must be positioned upstream of the CRISPR array, followed by a poly(A)
256 tail. However, this arrangement poses a problem: processing of the CRISPR array inevitably leads to the separation of the Cas
257 protein mRNA from the poly(A) tail, ultimately resulting in inadequate expression of the Cas protein and unsatisfactory

258 CRISPR efficiency. A previous study attempted to address this issue by harnessing a putative mRNA-stabilizing element called
259 “Triplex”.³⁷ Their results showed that the introduction of “Triplex” between EGFP and CRISPR array completely rescued the
260 loss of EGFP fluorescence due to the processing of downstream array. We are concerned about this kind of “complete rescue,”
261 as it challenges the irreplaceable role of poly(A) tails in stabilizing eukaryotic mRNA.³⁸

262 To test this hypothesis, we investigated whether this specific co-expression pattern is a viable option for multiplex CRISPR.
263 Consistent with the conventional understanding, removal of the poly(A) tail remarkably reduced EGFP expression at both the
264 mRNA and protein levels (Figure S19A, B), emphasizing the crucial role of the poly(A) tail in stabilizing mRNA. To examine
265 the mRNA-stabilizing efficacy of Triplex and the well-known woodchuck hepatitis virus post-transcriptional regulatory
266 element (WPRE), we cloned EGFP downstream of the EF1a promoter, followed by an array of 12 crRNAs for AsCas12a, with
267 or without a Triplex or WPRE element, between EGFP and the CRISPR array. The resulting plasmids were transfected into
268 HEK293T cells with the control vector or den12a-VPR. Forty-eight hours post-transfection, as a result of the processing of
269 crRNAs, the simultaneous expression of the den12a-VPR dramatically reduced EGFP mRNA and fluorescence. Distinct from
270 a previous study,³⁷ the additional introduction of Triplex or WPRE resulted in mild, if any, improvement in the mRNA or
271 fluorescence of EGFP (Figure S19C, D).

272 The targeting efficiency of co-expressing the den12a-VPR and CRISPR arrays on a single transcript was directly evaluated
273 next (Figure S20A). Co-expression via direct integration resulted in a marked reduction in den12a-VPR mRNA/protein
274 expression but not in the ultimate CRISPRa efficiency, which increased (Figure S20B-D). This discrepancy between decreased
275 den12a-VPR expression and enhanced CRISPRa efficiency might be a result of increased expression of the CRISPR array due
276 to the insertion of the upstream den12a-VPR. When the EF1a-array and den12a-VPR were delivered by individual constructs,
277 the expression of the CRISPR array and the final CRISPRa efficiency were enhanced by inserting an EGFP-coding sequence
278 upstream of the CRISPR array, but not by inserting a random stuffer (Figure 7A). Northern blots of mature crRNAs verified
279 an increase in expression levels, especially in the initial part (Figure 7B). This phenomenon is not unique, as CMV-driven
280 CRISPR arrays can also be enhanced using this approach (Figure 7C, D). Moreover, enhanced CRISPRa efficiency was
281 achieved by increasing the ratio of array-coding plasmids to those expressing the den12a-VPR (Figure S21A). Even the original
282 version (i.e., den12a-VPR), which is supposed to possess considerably lower activity and requires a smaller number of crRNAs,
283 showed improved CRISPRa efficiency when the ratio of the EF1a-array was increased or combined with the EF1a-GFP-array
284 (Figure S21B).

285 Regarding the additional insertion of Triplex or WPRE when co-expressing den12a-VPR and CRISPR arrays on a single
286 transcript, WPRE seemed to slightly rescue den12a-VPR mRNA/protein, whereas Triplex did not (Figure S20B, C). However,
287 neither improved the efficiency of CRISPRa compared to direct integration (Figure S20D).

288 Taken together, the strategy of co-expressing the Cas protein and CRISPR array on a single transcript may be a viable
289 option for some CRISPR systems under certain circumstances, but not always. According to our results, an important
290 determinant is that the optimal ratio of the Cas effector to its crRNA varies across CRISPR systems. A systematic and rigorous
291 evaluation is needed before applying this approach to other CRISPR systems not tested here, especially for *in vivo* delivery,
292 which was lacking in our study.

293

294

295 **DISCUSSION**

296 Although up to 15 crRNAs can be assembled in a single reaction using our method, some previous studies have used arrays
297 containing more than 20 crRNAs.³⁷ However, these methods typically perform more than one round of assembly or rely on
298 expensive long dsDNA segments purchased from companies. Based on purified construct coding the array of 12 crRNAs
299 assembled in the first round (while the assembly product of the first round was used as template; Figure S16), a second round
300 of PCR and subsequent assembly we performed (Figure S22A). Although recombination events occurred during clonal
301 expansion, because the same template was used for all three PCR reactions, the array of 36 crRNAs was efficiently assembled
302 with high accuracy (Figure S22B, C). Owing to the lack of appropriate application examples, we did not determine the
303 maximum number of crRNAs that could be assembled in two rounds of assembly. Given the remarkable accuracy of assembling
304 36 crRNAs, the successful assembly of more crRNAs (i.e., 40, 50, or even more) within two rounds is possible.

305 Speaking of unwanted recombination events during the transformation of plasmids into competent *E.coli*, we were puzzled
306 by its “elusiveness”. For convenience, even for the transformation of lentiviral constructs, we routinely use DH5 α , a commonly
307 used recA1 mutant competent *E.coli* strain, instead of stb13 or NEB stable, which are supposed to further reduce recombination
308 events between repeat elements in the plasmid. A lower temperature of 30 °C for bacterial culture is also rarely used. However,
309 we had never encountered any recombination events between the two long terminal repeats (LTRs) in any the lentiviral
310 constructs. Specific to the current study, the accuracy of the assembly of crRNAs was not improved by using stb13 or NEB
311 stable, nor by culturing at 30 °C (data not shown). Compared to the transformation of the newly-assembled mixture,
312 recombination events were rare during the re-transformation of repeat-containing plasmids extracted from bacteria (data not
313 shown). We are not sure whether most of the negative colonies were the result of unwanted recombination or simply failed
314 assembly. As the reagents associated with Golden Gate cloning have been continuously upgraded over the years,^{39,40} the limits
315 on the maximum number of crRNAs that can be assembled in a single reaction, if the failure of assembly is currently the
316 primary constraint, may be pushed further in the future.

317 In contrast to previous reports,^{24,37} the current results indicate that EF1 α is not functionally equivalent to U6 in the
318 transcription of CRISPR arrays or single crRNA. Despite the differences between their strengths when transcribing transcripts
319 of varying lengths, additional modifications (e.g., 7-methylguanosine (m7G) cap and poly(A) tail) may also affect the
320 subcellular localization or processing of the CRISPR array, thereby affecting its final execution. We attempted to quantitatively
321 compare the transcription levels of the precursor crRNA arrays using RT-qPCR (Figure S23). To generate amplified segments

322 of the appropriate length, each primer pair must span three to four crRNAs. This intrinsic property causes the final result to be
323 affected by the positions of both the forward and reverse primers. For these RT-qPCR experiments, because we used reverse
324 qPCR primers to produce cDNA in the reverse transcription reaction, the quantitative results were dominated by the expression
325 level of the reverse primer locus. This may not be appropriate for the CRISPR array driven by U6 because the most abundant
326 short transcripts containing only one or two crRNAs were not detected.

327 While the loss of the poly(A) tail by different approaches led to reduced EGFP expression, we noticed that the reduction
328 seemed to be more drastic when the loss of the poly(A) tail was caused by processing the downstream CRISPR array, compared
329 to simply removing the SV40-poly(A) signal in the construct (Figure S19). Although we are unsure whether the former
330 accurately represents the actual effect of poly(A) deficiency, the latter may underestimate it. When SV40-poly(A) was removed
331 from the construct, Pol II promoters did not stop transcribing upon completion of EGFP transcription, but instead continued to
332 transcribe uncertain downstream sequences until terminated for other reasons. Thus, the 3' terminus of the final EGFP mRNA
333 was not directly exposed, but was “protected” by downstream RNA of uncertain length, which may delay the degradation of
334 EGFP mRNA by acting as a stuffer. In lentiviral constructs, the poly(A) signal between the two LTRs must be removed, as it
335 would lead to the premature termination of viral RNA during virus packaging. Consequently, the sequence integrated into the
336 host genome after lentiviral infection does not have a poly(A) signal. A WPRE element is frequently introduced upstream of
337 the 3' LTR, which is supposed to stabilize and facilitate the translation of mRNA.⁴¹ Whereas the improvement in the expression
338 of upstream coding genes in our previous results with transient expression was mild (Figure S19), WPRE markedly enhanced
339 the expression of upstream EGFP in the context of lentiviral delivery, which is consistent with the conventional concept. In
340 contrast, the triplexes failed under these circumstances (Figure S23).

341 Although the previously reported co-expression strategy inevitably resulted in reduced expression of the Cas effector in
342 our re-evaluation, it did not necessarily lead to compromised CRISPR efficiency, at least in the denAsCas12a-VPR mediated
343 multiplex activation system. However, regarding the genomic DNA-targeting system with the primordial nuclease-active
344 AsCas12a (or enAsCas12a), it is currently unknown whether this particular co-expression approach is compatible.

345 One of the shortcomings of this study, which we must emphasize, is that all the evaluations were performed *in vitro* using
346 cell lines, and these results may not precisely represent the *in vivo* performance of the CRISPR systems and expression methods
347 assessed here. Therefore, a rigorous *in vivo* evaluation is warranted before further application, especially when harnessing Pol
348 II promoters to express CRISPR arrays or when using the co-expression strategy. The findings will substantially simplify the

349 preparatory work prior to *in vivo* manipulation of multiple targets and boost the exploration of potential applications using
350 multiplex CRISPR.

351

352

Journal Pre-proof

353 MATERIALS AND METHODS**354 Plasmid construction**

355 The coding sequences of AsCas12a, VPR, and RfxCas13d were amplified from pY108 (Addgene #84739), pXR001 (#84739),
356 and lenti-EF1a-dCas9-VPR-Puro (#99373) plasmids, respectively. The amplified fragments were cloned into destination
357 vectors using standard digestion–ligation or the Gibson method. Gibson cloning was used to introduce the desired mutations
358 into protein-coding plasmids. The CRISPR arrays were assembled using direct ligation or the standard Golden Gate method.
359 The junction sets of the four-base overhangs were determined based on their ligation fidelity predicted by the online tool
360 NEBridge Ligase Fidelity Viewer (<https://ggtools.neb.com/viewset/run.cgi>).³⁹ Detailed protocols are provided in the
361 Supplemental Methods. All constructs, including those encoding the CRISPR arrays used in the editing experiments, were
362 validated using Sanger sequencing. The detailed sequences of these constructs are listed in Table S22.

363 Assembly of CRISPR arrays with GGA-based strategy

364 Detailed protocols can be found in the Supplemental Methods. Briefly, PCR was performed using predesigned oligo pairs
365 (synthesized by Tsingke Biotech). Next, an optional recovery step was performed, if necessary. The Golden Gate assembly
366 reaction was set up with purified segments (or diluted PCR products) and a destination cloning vector. The reagents used were
367 BsaI-HF v2 (NEB, R3733) and T4 DNA Ligase (NEB, M0202). The Golden Gate Assembly program was run using a
368 thermocycler: (37 °C 5 min → 16 °C 5 min) × 30 cycles, followed by 60 °C for 5 min. If reactions were performed overnight,
369 a 4 °C terminal hold was added to the program and the step at 60 °C for 5 min until the day before transformation. An amount
370 of 2–10 µl of the assembly reaction was transformed into competent cells.

371 Cell culture and transient transfection

372 HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS and 1%
373 penicillin/streptomycin at 37 °C with 5% CO₂. For transient transfection, 2×10⁵ HEK293T cells/well were seeded in 24-well
374 plates. When 60–80% confluence was reached the following day, a total of 500 ng plasmids were transfected into each well
375 using Lipo8000™ Transfection Reagent (Beyotime) according to the manufacturer's protocol. Transfected cells were supplied
376 with fresh, complete DMEM (containing FBS) every 24h, and harvested 48–72 h after transfection for downstream experiments.

377 Lentivirus production and infection

378 For lentivirus packaging, 5×10^5 HEK293T cells were seeded per well into 12-well plates. When 80% confluence was reached
379 80% the following day, 1 μ g of mixed plasmids (transfer: psPAX2: pMD2.G = 4:3:1) were transfected into each well using
380 Lipo8000™ (Beyotime). The medium was replaced with 1 mL fresh, complete DMEM 12–24 h after transfection. Then, 48 h
381 after transfection, another 1 mL of complete DMEM was added. At 72 h after transfection, lentivirus-containing supernatant
382 was harvested, filtered through a 0.45- μ m filter (or centrifuged at $2,000 \times g$ for 5 min) and stored at 4 °C (or -80 °C for long-
383 term preservation).

384 For lentivirus infection, 1.5×10^5 HEK293T cells were seeded per well into 24-well plates. When 40–60% confluence was
385 reached the following day, the medium was replaced with 0.5 mL fresh complete DMEM and 1.5 mL lentivirus-containing
386 supernatant. The medium was replaced with fresh, complete DMEM 12–24 h after infection and refreshed/supplemented every
387 24 h until cells were harvested 72 h after infection.

388 **Extraction of RNA and RT-qPCR**

389 Total RNA was extracted using RNAiso Plus (Takara) 48–72 h after transient transfection or lentiviral infection. An amount
390 of 1 μ g total RNA was then reverse transcribed using ReverTra Ace® qPCR RT Kit (Toyobo) with supplied primer mix or
391 gene-specific primers (when designated, and sequences are listed in Table S2), followed by qPCR using ChamQ Universal
392 SYBR qPCR Master Mix (Vazyme) and the primers listed in Table S1. The qPCR was performed using a LightCycler 480 II
393 (Roche). Quantification of RNA expression was normalized to ACTB (unless otherwise specified) and calculated using the
394 $\Delta\Delta$ Ct method.

395 **Electrophoresis of oligonucleotides**

396 DNA oligonucleotide electrophoresis was performed on a 20% polyacrylamide gel in 1 \times TBE buffer. The gel was stained with
397 Gel Red (Beyotime) for 1 h at room temperature and scanned using an ultraviolet transilluminator after washing.

398 **Quantification of gene editing**

399 A total of 72 h after transfection, genomic DNA was extracted using a Rapid Animal Genomic DNA Isolation Kit (Sangon
400 Biotech) following the manufacturer's protocol. The extracted DNA was used as a template to amplify the target region flanking
401 the edited site, using specific primers. The PCR amplicons were purified using the AxyPrep PCR Clean-up Kit (Axygen)
402 following the manufacturer's protocol. Purified DNA was subjected to Sanger sequencing using the forward PCR primer as

403 the sequencing primer. The resulting “.ab1” files were uploaded to obtain the final quantitative spectrum of the indels using the
404 online tool: <https://ice.synthego.com>.⁴²

405 **Northern blot**

406 Northern blots of mature crRNAs were based on the Biotin–Streptavidin system and performed using the Biotin Northern Blot
407 Kit (for Small RNA) (Beyotime, R0220) according to the manufacturer’s protocol. The 5’ biotin-labelled DNA probes were
408 synthesized by Tsingke Biotech. The probe sequences are listed in Table S21. The 5s rRNA was used as an internal control
409 and was detected by agarose gel electrophoresis.

410 **Western blot**

411 A total of 72 h after transient transfection or lentivirus infection, cells were lysed in RIPA buffer (MedChem Express)
412 supplemented with PMSF and BeyoZonase™ Super Nuclease (Beyotime). Total protein concentration was quantitated using
413 Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific). All lysates were diluted to a concentration of 1250 ng/μL, then
414 mixed with 5× loading buffer (a final protein concentration of 1 μg/μL) and boiled for 5 min. Equal amounts of protein were
415 loaded and separated using sodium dodecyl sulphate–polyacrylamide gel electrophoresis and then transferred to a
416 polyvinylidene difluoride membrane. Membranes were blocked with 5% nonfat milk in TBST buffer for 1 h at room
417 temperature, and incubated with appropriate dilutions of primary antibody overnight at 4 °C. The next day, the membrane was
418 washed three times with TBST and incubated with a horseradish peroxidase-conjugated secondary antibody for 1 h at room
419 temperature. After three washes, chemiluminescent signals in the membrane were captured using a CCD camera.

420 **Statistical analysis**

421 Values are reported as mean or mean ± standard deviation as indicated in the appropriate figure legends. Unless otherwise
422 stated, at least three biological replicates were used for each experiment. When comparing two groups, statistical differences
423 were determined using the unpaired Student’s *t*-test. One-way analysis of variance was used to assess the significance of
424 differences between more than two groups. Two-way analysis of variance was used to compare two factors. Statistical
425 significance was set at $P < 0.05$. Prism 6.01 was used for all statistical analyses.

426

427 **DATA AND CODE AVAILABILITY**

428 The authors declare that the data supporting the results of this study are available in the article and supplemental material.
429 Additional resource information is available from the corresponding author upon request.

430 KEYWORDS

431 Multiplex CRISPR, crRNA, CRISPR array, Golden Gate Assembly, AsCas12a, RfxCas13d, Pol II promoter, and Pol III
432 promoter

433 ACKNOWLEDGEMENTS

434 We thank the members of the Institute of Immunology (Zhejiang University School of Medicine) for their support in this study.
435 This work was financially supported by National Natural Science Foundation of China (32300642 and 82102814), Zhejiang
436 Provincial Natural Science Foundation of China (LQ24C060009 and LQ22H160053), and postdoctoral grants for scientific
437 research from the Zhejiang Provincial People's Hospital (C-2023-BSH28).

438 AUTHOR CONTRIBUTIONS

439 XZ: conception and design, collection of data, data analysis, manuscript writing; LY: conception and design, data collection,
440 data analysis, manuscript writing, financial support; PL, ZC, YJ, LL, AB, HX, HZ, and JZ: collection of data, data analysis,
441 manuscript editing, administrative, technical, material, or financial support; YZ: conception and design, data analysis, financial
442 support, revision, and final approval of the manuscript.

443 DECLARATION OF INTERESTS

444 The authors declare no competing interests.

445 SUPPLEMENTAL INFORMATION

446 Supplemental information can be found online at.

447

448 REFERENCES

- 449 1. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA
450 endonuclease in adaptive bacterial immunity. *Science*. 337, 816-821.
- 451 2. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex
452 genome engineering using CRISPR/Cas systems. *Science*. 339, 819-823.
- 453 3. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome
454 engineering via Cas9. *Science*. 339, 823-826.
- 455 4. Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath,
456 P., et al. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol*. 18, 67-
457 83.
- 458 5. Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova,
459 E., Minakhin, L., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*. 353,
460 aaf5573.
- 461 6. East-Seletsky, A., O'Connell, M.R., Knight, S.C., Burstein, D., Cate, J.H., Tjian, R., and Doudna, J.A. (2016). Two distinct RNase
462 activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*. 538, 270-273.
- 463 7. Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdine, V., Cox, D.B.T., Kellner, M.J., Regev,
464 A., et al. (2017). RNA targeting with CRISPR-Cas13. *Nature*. 550, 280-284.
- 465 8. Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., and Zhang, F. (2017). RNA editing with
466 CRISPR-Cas13. *Science*. 358, 1019-1027.
- 467 9. Konermann, S., Lotfy, P., Brideau, N.J., Oki, J., Shokhirev, M.N., and Hsu, P.D. (2018). Transcriptome Engineering with RNA-
468 Targeting Type VI-D CRISPR Effectors. *Cell*. 173, 665-676 e614.
- 469 10. Ozcan, A., Krajcski, R., Ioannidi, E., Lee, B., Gardner, A., Makarova, K.S., Koonin, E.V., Abudayyeh, O.O., and Gootenberg, J.S.
470 (2021). Programmable RNA targeting with the single-protein CRISPR effector Cas7-11. *Nature*. 597, 720-725.
- 471 11. Colognori, D., Trinidad, M., and Doudna, J.A. (2023). Precise transcript targeting by CRISPR-Csm complexes. *Nat Biotechnol*.
- 472 12. Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S.,
473 Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. 517, 583-588.
- 474 13. Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik,
475 M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*. 159, 647-661.
- 476 14. Anzalone, A.V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime
477 editors. *Nat Biotechnol*. 38, 824-844.
- 478 15. Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome
479 editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol*. 33, 510-517.
- 480 16. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram,
481 A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*. 576, 149-157.
- 482 17. Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in
483 gene expression and fluorescence imaging. *Cell*. 159, 635-646.
- 484 18. Ma, H., Marti-Gutierrez, N., Park, S.W., Wu, J., Lee, Y., Suzuki, K., Koski, A., Ji, D., Hayama, T., Ahmed, R., et al. (2017). Correction
485 of a pathogenic gene mutation in human embryos. *Nature*. 548, 413-419.

- 486 19. Wang, D., Li, J., Song, C.Q., Tran, K., Mou, H., Wu, P.H., Tai, P.W.L., Mendonca, C.A., Ren, L., Wang, B.Y., et al. (2018). Cas9-
487 mediated allelic exchange repairs compound heterozygous recessive mutations in mice. *Nat Biotechnol.* *36*, 839-842.
- 488 20. Beyret, E., Liao, H.K., Yamamoto, M., Hernandez-Benitez, R., Fu, Y., Erikson, G., Reddy, P., and Izpisua Belmonte, J.C. (2019). Single-
489 dose CRISPR-Cas9 therapy extends lifespan of mice with Hutchinson-Gilford progeria syndrome. *Nat Med.* *25*, 419-422.
- 490 21. Maeder, M.L., Stefanidakis, M., Wilson, C.J., Baral, R., Barrera, L.A., Bounoutas, G.S., Bumcrot, D., Chao, H., Ciulla, D.M., DaSilva,
491 J.A., et al. (2019). Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat Med.* *25*,
492 229-233.
- 493 22. Nelson, C.E., Wu, Y., Gemberling, M.P., Oliver, M.L., Waller, M.A., Bohning, J.D., Robinson-Hamm, J.N., Bulaklak, K., Castellanos
494 Rivera, R.M., Collier, J.H., et al. (2019). Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy.
495 *Nat Med.* *25*, 427-432.
- 496 23. Santiago-Fernandez, O., Osorio, F.G., Quesada, V., Rodriguez, F., Basso, S., Maeso, D., Rolas, L., Barkaway, A., Nourshargh, S.,
497 Folgueras, A.R., et al. (2019). Development of a CRISPR/Cas9-based therapy for Hutchinson-Gilford progeria syndrome. *Nat Med.* *25*,
498 423-426.
- 499 24. Kleinstiver, B.P., Sousa, A.A., Walton, R.T., Tak, Y.E., Hsu, J.Y., Clement, K., Welch, M.M., Horng, J.E., Malagon-Lopez, J., Scarfo,
500 I., et al. (2019). Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and
501 base editing. *Nat Biotechnol.* *37*, 276-282.
- 502 25. Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013). CAS9 transcriptional
503 activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.* *31*, 833-838.
- 504 26. Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA-guided activation of endogenous
505 human genes. *Nat Methods.* *10*, 977-979.
- 506 27. Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R.
507 (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* *23*,
508 1163-1171.
- 509 28. Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G.,
510 Leong, K.W., et al. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods.* *10*, 973-976.
- 511 29. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost,
512 J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell.* *163*, 759-771.
- 513 30. Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1
514 also processes precursor CRISPR RNA. *Nature.* *532*, 517-521.
- 515 31. DeWeirdt, P.C., Sanson, K.R., Sangree, A.K., Hegde, M., Hanna, R.E., Feeley, M.N., Griffith, A.L., Teng, T., Borys, S.M., Strand, C.,
516 et al. (2021). Optimization of AsCas12a for combinatorial genetic screens in human cells. *Nat Biotechnol.* *39*, 94-104.
- 517 32. Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-
518 processing system. *Proc Natl Acad Sci U S A.* *112*, 3570-3575.
- 519 33. Shao, S., Chang, L., Sun, Y., Hou, Y., Fan, X., and Sun, Y. (2018). Multiplexed sgRNA Expression Allows Versatile Single
520 Nonrepetitive DNA Labeling and Endogenous Gene Regulation. *ACS Synth Biol.* *7*, 176-186.
- 521 34. Breunig, C.T., Durovic, T., Neuner, A.M., Baumann, V., Wiesbeck, M.F., Koflerle, A., Gotz, M., Ninkovic, J., and Stricker, S.H. (2018).
522 One step generation of customizable gRNA vectors for multiplex CRISPR approaches through string assembly gRNA cloning (STAgR).
523 *PLoS One.* *13*, e0196015.
- 524 35. Zhou, H., Huang, C., and Xia, X.G. (2008). A tightly regulated Pol III promoter for synthesis of miRNA genes in tandem. *Biochimica
525 et biophysica acta.* *1779*, 773-779.
- 526 36. Xu, L., Zhao, L., Gao, Y., Xu, J., and Han, R. (2017). Empower multiplex cell and tissue-specific CRISPR-mediated gene manipulation
527 with self-cleaving ribozymes and tRNA. *Nucleic Acids Res.* *45*, e28.

- 528 37. Campa, C.C., Weisbach, N.R., Santinha, A.J., Incarnato, D., and Platt, R.J. (2019). Multiplexed genome engineering by Cas12a and
529 CRISPR arrays encoded on single transcripts. *Nat Methods*. *16*, 887-893.
- 530 38. Passmore, L.A., and Coller, J. (2022). Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nat Rev Mol Cell Biol*.
531 *23*, 93-106.
- 532 39. Potapov, V., Ong, J.L., Kucera, R.B., Langhorst, B.W., Bilotti, K., Pryor, J.M., Cantor, E.J., Canton, B., Knight, T.F., Evans, T.C., Jr.,
533 et al. (2018). Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly.
534 *ACS Synth Biol*. *7*, 2665-2674.
- 535 40. Pryor, J.M., Potapov, V., Kucera, R.B., Bilotti, K., Cantor, E.J., and Lohman, G.J.S. (2020). Enabling one-pot Golden Gate assemblies
536 of unprecedented complexity using data-optimized assembly design. *PLoS One*. *15*, e0238592.
- 537 41. Zufferey, R., Donello, J.E., Trono, D., and Hope, T.J. (1999). Woodchuck hepatitis virus posttranscriptional regulatory element enhances
538 expression of transgenes delivered by retroviral vectors. *J Virol*. *73*, 2886-2892.
- 539 42. Conant, D., Hsiao, T., Rossi, N., Oki, J., Maures, T., Waite, K., Yang, J., Joshi, S., Kelso, R., Holden, K., et al. (2022). Inference of
540 CRISPR Edits from Sanger Trace Data. *CRISPR J*. *5*, 123-130.

541

542 FIGURE LEGENDS

543 **Figure 1. Schematic workflow of Golden Gate Assembly-based CRISPR array assembly strategy**

544 First, PCRs were set up and PCR reactions were run with predesigned oligo pairs that were partially complementary to each other. Next, an
545 optional recovery step was performed if necessary (because of the relatively short DR length of AsCas12a, some of the resulting dsDNA
546 segments were too short to be recovered using common DNA recovery kits; therefore, we used a relatively crude method via ethanol
547 precipitation. Although single-stranded oligos or dNTPs might still be included, most of the salts were removed, and the DNA polymerase
548 was inactivated). Finally, a standard Golden Gate assembly reaction was set up and run with the purified segments (or diluted PCR products)
549 and a destination cloning vector.

550

551 **Figure 2. High-accuracy assembly of CRISPR array with GGA-based strategy**

552 (A, B) Accuracies of the assembly of six (A) or seven (B) crRNAs using the conventional sticky end-based or novel GGA-based strategy.
553 Values shown as mean, $n = 3$ independent experiments. (C) Accuracies of the assembly of nine or 12 crRNAs using the GGA-based strategy.
554 Values shown as mean with $n \geq 3$. (D) Quantification of relative mRNA expression over the non-targeting control in HEK293T cells 48 h
555 after transfection with a single plasmid containing CBh-driven denAsCas12a-VPR and a U6-driven CRISPR array of 12 crRNAs targeting
556 the indicated six genes. CBh, a robust Pol II promoter. Values shown as mean \pm SD with $n = 3$.

557

558 **Figure 3. Simplification of the GGA-based strategy for CRISPR array assembly**

559 (A) Schematics of wild-type and three mutant DR variants of the AsCas12a system. (B) Quantification of relative *HBB* expression over the
560 non-targeting control in HEK293T cells 48 h after transfection with den12a-VPR and arrays with DR variants depicted in (A). (C) An example
561 showcasing the streamlining of the initial primer design by using a mutant DR with higher GC content. (D) Accuracies in the assembly of
562 nine or 12 crRNAs with mutant DR using the GGA-based strategy. Values shown as mean with $n = 3$.

563

564 **Figure 4. High-accuracy assembly of crRNAs for a Cas13d nuclease**

565 (A) Accuracies of the assembly of 9, 12, or 15 crRNAs for RfxCas13d using the GGA-based strategy. Values shown as mean with $n \geq 3$. (B)
566 Quantification of relative mRNA expression compared to the non-targeting control in HEK293T cells 48 h after transfection with a plasmid

567 containing EF1a-RfxCas13d and a U6-driven CRISPR array of 15 crRNAs targeting the indicated 15 endogenous transcripts. Values shown
568 as mean with $n = 3$.

569 **Figure 5. Distinct expression patterns of Pol II/III promoter-driven CRISPR arrays**

570 (A) Quantification of relative mRNA expression over the non-targeting control in HEK293T cells 48 h after transfection with EF1a-
571 denAsCas12a-VPR and an array of 12 crRNAs driven by either U6 (U6-array) or EF1a (EF1a-array). Values shown as mean \pm SD with $n =$
572 3. (B) Representative northern blot images of mature crRNAs processed from CRISPR arrays driven by U6 or EF1a. Blots of crRNA-6 were
573 not detected because the corresponding probe was not working well. (C, D) Quantification of the indicated gene expression over the non-
574 targeting control in HEK293T cells 48 h after transfection with EF1a-denAsCas12a-VPR and a U6- or EF1a-driven array of 2 crRNAs
575 targeting the promoter of *HBB* (C) or *RHOXF2* (D). (E) Quantification of gene editing efficiency in HEK293T cells 72 h after transfection
576 with enAsCas12a and a *DNMT1*-targeting crRNA driven by either U6 or EF1a. (F, G) Quantification of relative *HBB* expression over the
577 non-targeting control in HEK293T cells 48h after transfection with EF1a-denAsCas12a-VPR and the indicated arrays (12 \times) driven by U6 (F)
578 or EF1a (G), where HBB-targeting crRNAs were positioned at different loci. Values shown as mean with $n = 3$.

579

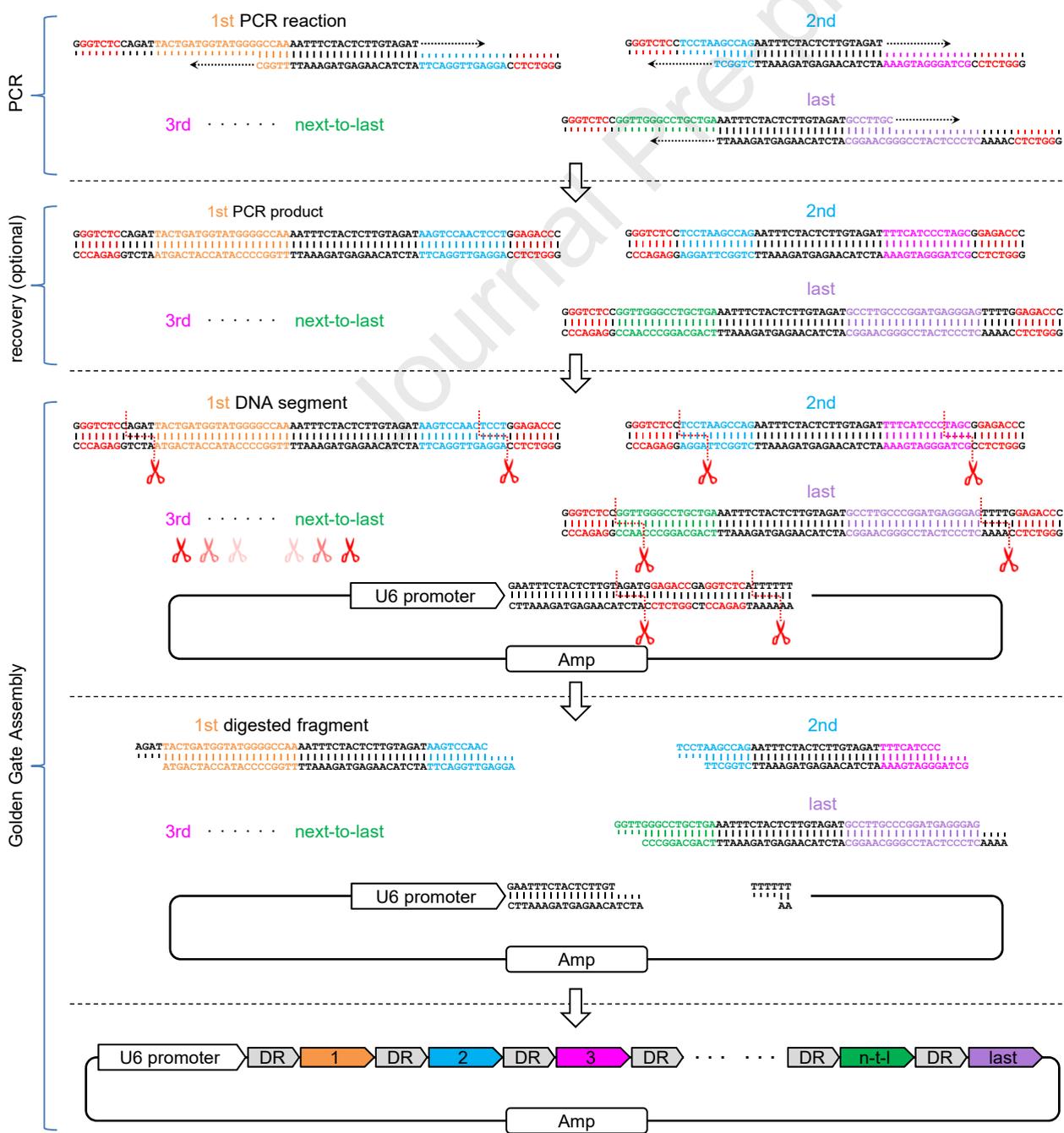
580 **Figure 6. Improved targeting efficiency by optimizing the internal architecture of the CRISPR array**

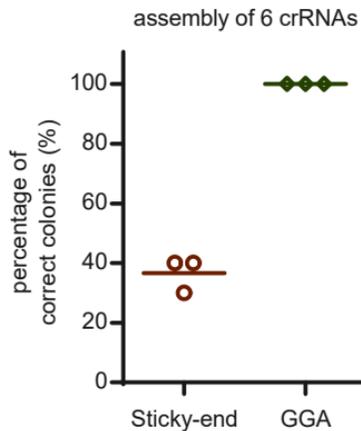
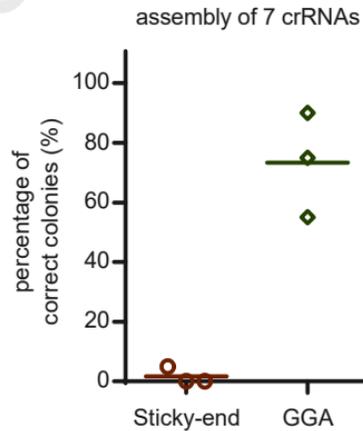
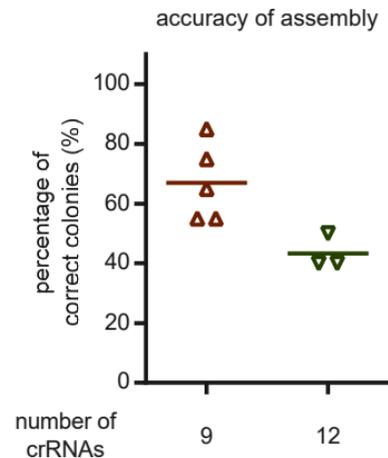
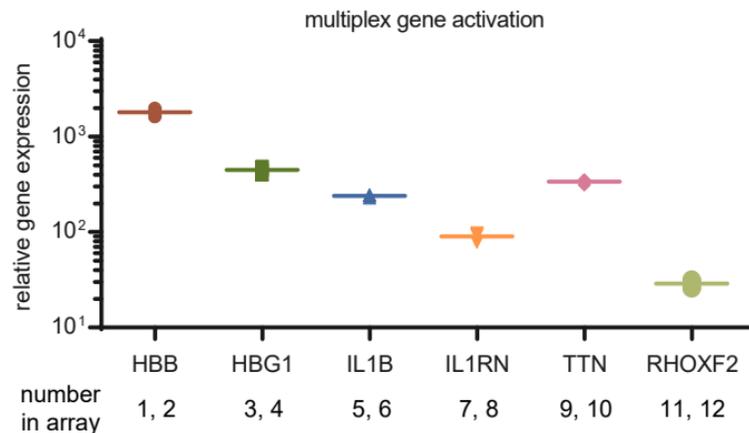
581 (A) Schematics of CRISPR arrays used with denAsCas12a-VPR: U6-array (12), EF1a-array (12), U6-4-U6-4-U6-4. (B) Quantification of
582 relative mRNA expression over the non-targeting control in HEK293T cells 48 h after transfection with EF1a-denAsCas12a-VPR and the
583 indicated arrays. Values shown as mean \pm SD with $n = 3$. (C) Schematics of CRISPR arrays used with RfxCas13d: U6-array (15), U6-4-U6-
584 4-U6-4-U6-3. (D) Quantification of relative mRNA expression of the indicated genes compared to the non-targeting control in HEK293T
585 cells 48 h after transfection with RfxCas13d and the indicated arrays. Values shown as mean \pm SD with $n = 3$.

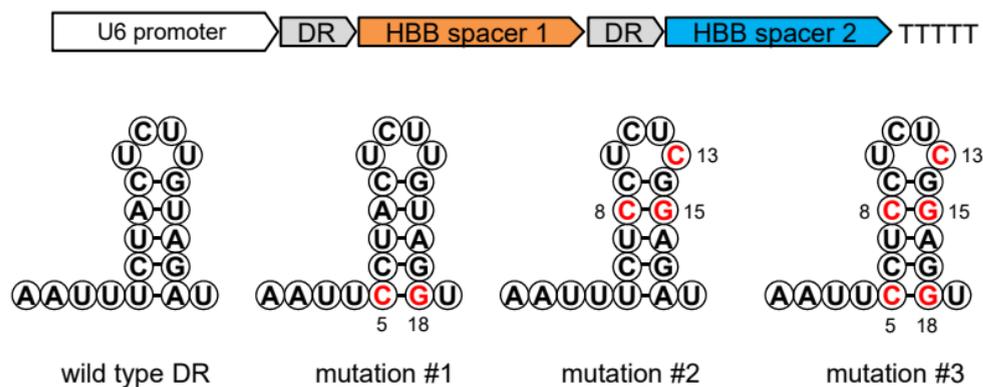
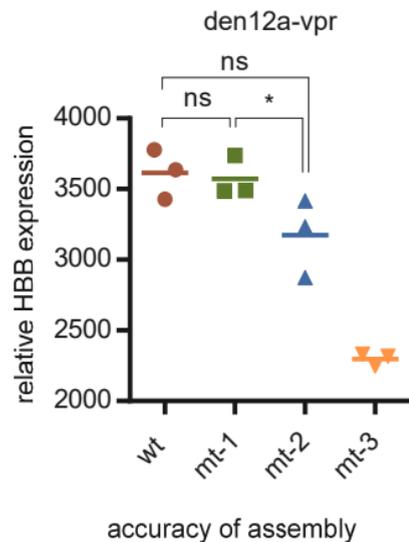
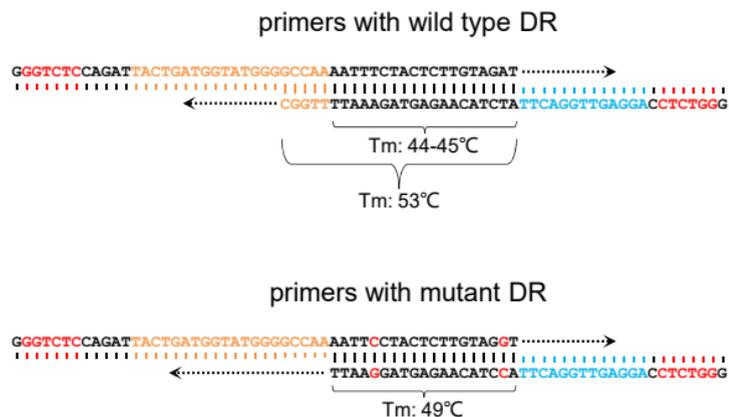
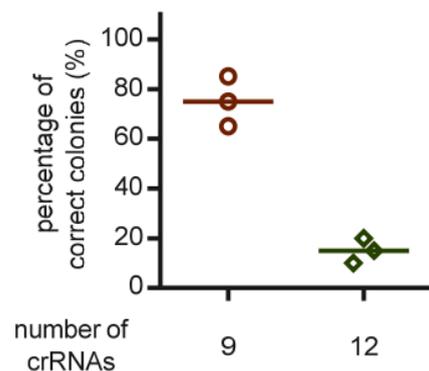
586

587 **Figure 7. Enhanced expression of Pol II promoter-driven CRISPR arrays by introducing an upstream GFP-coding sequence**

588 (A) Quantification of relative mRNA expression over the non-targeting control in HEK293T cells 48h after transfection with EF1a-den12a-
589 VPR and EF1a-driven arrays with/without upstream EGFP or stuffer. The 745-bp stuffer with a GC content of \sim 30% was cloned from the
590 pLKO.1 cloning vector (see Supplemental Table for a detailed sequence). Values are shown as mean \pm SD with $n = 3$. (B) Representative
591 northern blot images of mature crRNAs processed from EF1a- and EF1a-GFP-arrays. (C) Quantification of relative mRNA expression over
592 the non-targeting control in HEK293T cells 48 h after transfection with EF1a-den12a-VPR and CMV-driven arrays with/without upstream
593 EGFP. Values shown as mean \pm SD with $n = 3$. (D) Representative northern blot images of mature crRNAs processed from CMV- and CMV-
594 GFP-arrays.

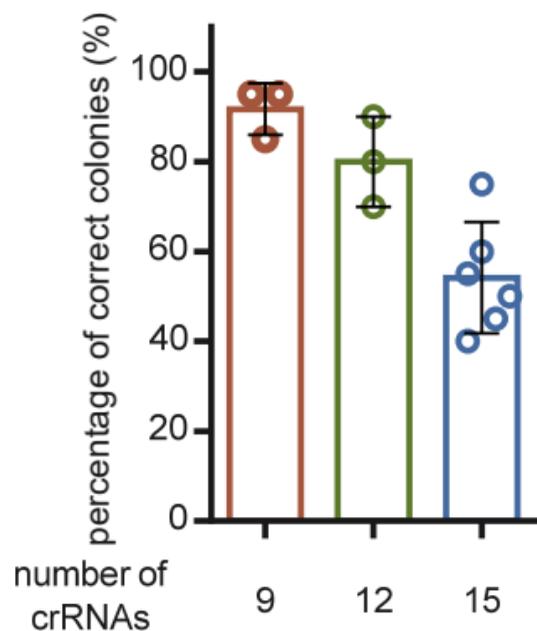


A**B****C****D**

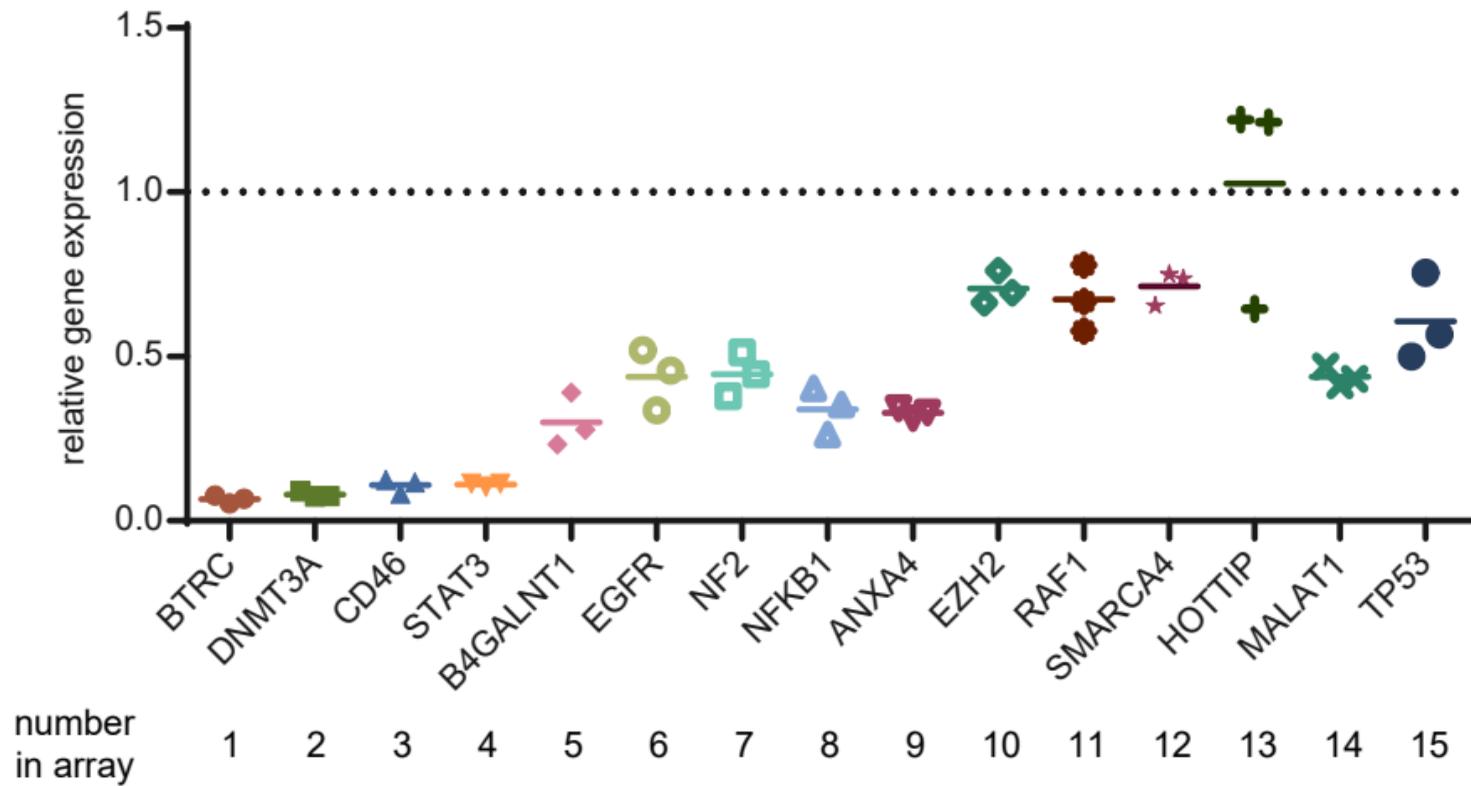
A**B****C****D**

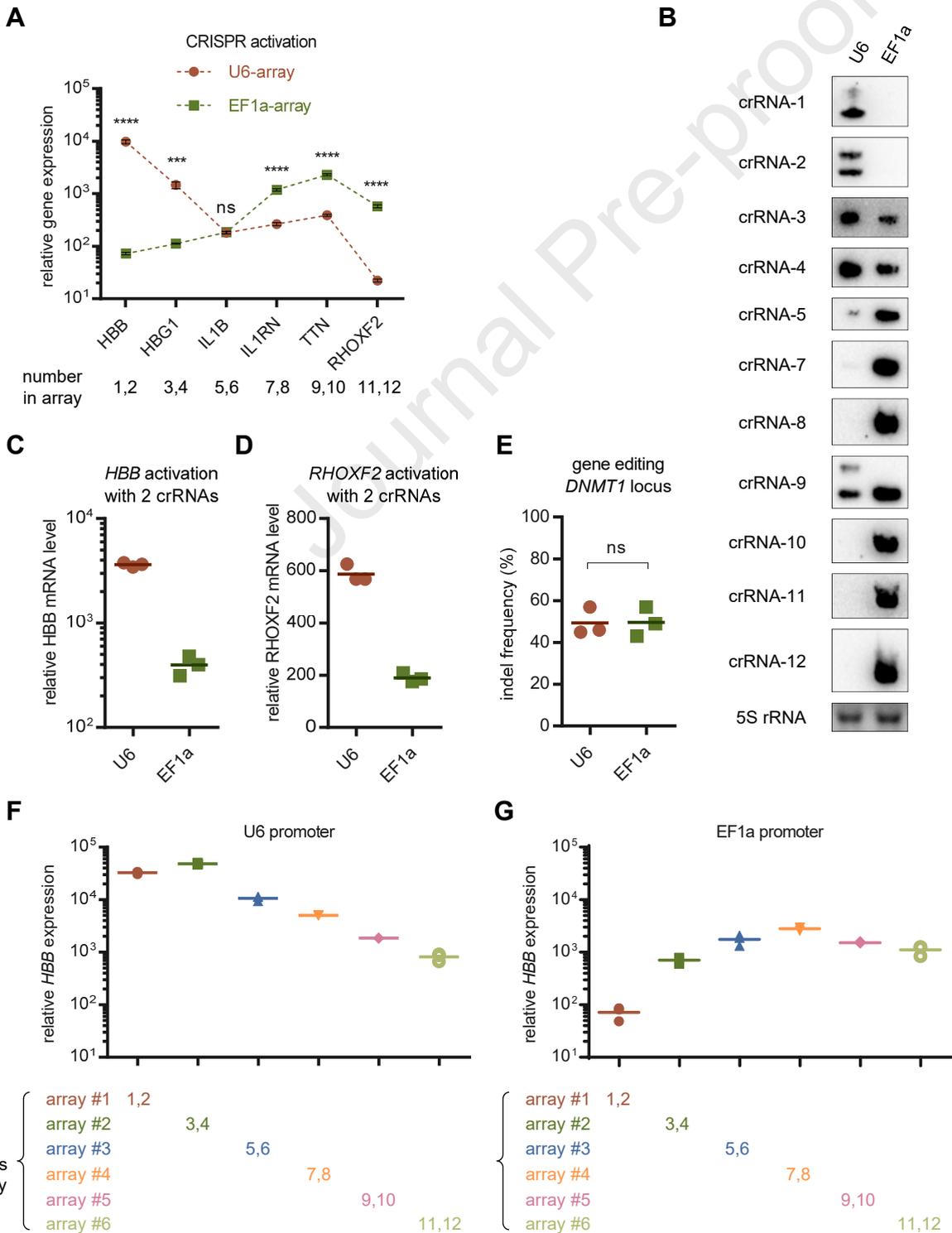
A

accuracy of assembly

**B**

multiplex RNA-targeting CRISPR





A

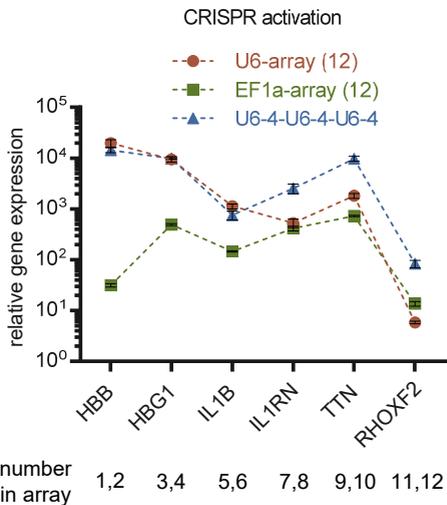
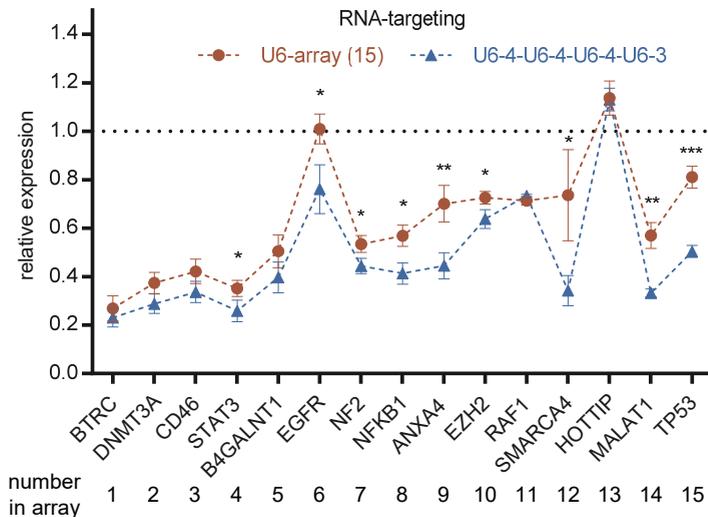
U6-array (12)



EF1a-array (12)



U6-4-U6-4-U6-4

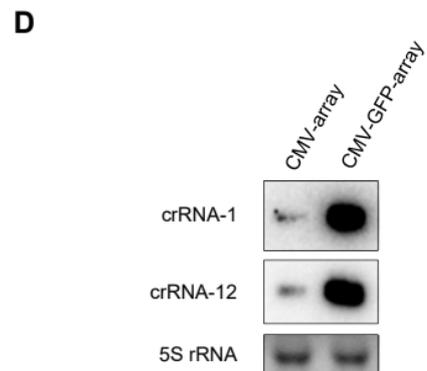
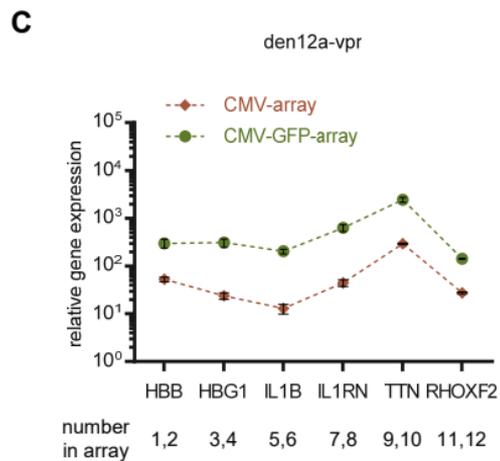
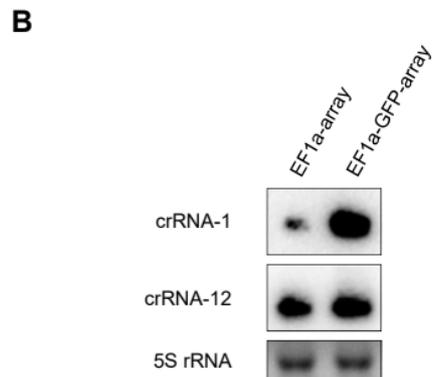
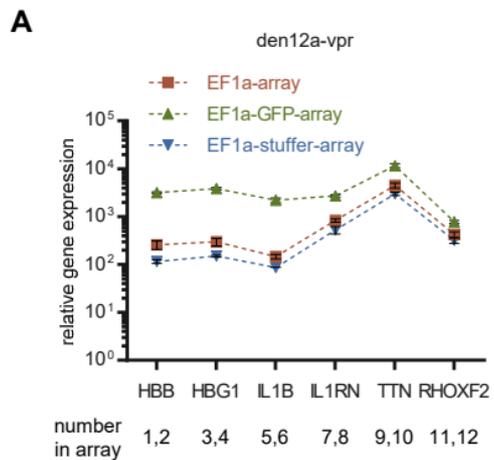
**B****D****C**

U6-array (15)



U6-4-U6-4-U6-4-U6-3





Zhao and colleagues designed a novel high-accuracy, cost- and time-saving strategy for CRISPR array assembly. Moreover, they revealed that CRISPR arrays driven by Pol II promoters exhibit a distinct expressing pattern compared to Pol III promoters. Based on these findings, we designed improved approaches for expressing long CRISPR arrays.

Journal Pre-proof