Methods xxx (xxxx) xxx

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Paired dCas9 design as a nucleic acid detection platform for pathogenic strains

Yihao Zhang [a,d], Yu Wang [d], Luze Xu [d], Chunbo Lou [b], Qi Ouyang [a,c], Long Qian [a,*]

[a] Center for Quantitative Biology and Peking-Tsinghua Joint Center for Life Sciences, Peking University, Beijing 100871, China
[b] CAS Key Laboratory of Quantitative Engineering Biology, Guangdong Provincial Key Laboratory of Synthetic Genomics, Shenzhen Key Laboratory of Synthetic Genomics, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[c] The State Key Laboratory for Artificial Microstructures and Mesoscopic Physics, School of Physics, Peking University, Beijing 100871, China
[d] Peking University Team for 2015 International Genetically Engineered Machine Competition (iGEM 2015), Beijing 100871, China

ARTICLE INFO

Keywords:
Nucleic acid detection
dCas9
Biosensor
Molecular Diagnostics

ABSTRACT

The wide application of molecular beacon probes in specific DNA detection, especially in the fast prototyping of pathogen DNA detection kits in point-of-care diagnostics, has been hindered by the nonflexible choice of target sequences and the unstable fluorophore output. We developed an *in vitro* DNA detection system consisting of a pair of dCas9 proteins linked to split halves of luciferase, named the Paired dCas9 (PC) reporter. Co-localization of the reporter pair to a ~46 bp target sequence defined by two single guide RNAs (sgRNAs) activated luciferase which subsequently generated highly intensified luminescent signals. Combined with an array design and statistical analyses, the PC reporter system could be programmed to access sequence information across the entire genome of the pathogenic *Mycobacterium tuberculosis* H37Rv strain. These findings suggest great potential for the PC reporter in effective and affordable *in vitro* nucleic acid detection technologies. In this article we highlighted the systems design from our previous research work on the PC reporter (Zhang et al, 2015) with a focus on methodology.

1. Introduction

The sensitive detection of specific nucleic acids (NA) of interest in clinical or environmental samples is fundamental to many areas of modern life. These include molecular diagnostics, public health surveillance, food safety, biosafety, etc. The ongoing COVID-19 pandemic have been controlled in many countries thanks to the rapid development and massive deployment of the viral nucleic acid detection kits that allow for large scale population screens to identify infected individuals. Given the recurrent outbreaks of infectious diseases worldwide in recent years, point-of-care nucleic acid detection devices are in urgent need especially for far-reaching and under-developed areas.

The earliest NA detection methods are those based on target specific amplification, these include regular polymerase chain reaction (PCR) and its many updated or adapted versions such as quantitative PCR, loop-mediated isothermal amplification (LAMP) and recombinase polymerase amplification (RPA) [1–3]. Through successive optimizations over the past 30 years, PCR-based NA detection methods, as well as detection through direct sequencing, are well established and standardized now. Yet, they often involve sophisticated primer designs and require professional instruments and trained personnel to perform the tests. Moreover, there is an inherent trade-off between sensitivity and specificity for PCR-based detection methods [4]. Later on, for a brief period of time, molecular beacons (MBs) were developed for targeted nucleic acid detection. The beacons recognize a target by sequence-specific hybridization, which induces a conformational change that sends off a signal. For example, an RNA aptamer linked to a quenched fluorophore changes to the open conformation when it hybridizes to the target, thereby releasing the fluorescence. MBs are much simpler in ingredients, but their design and thereby the targetable sequences are constrained by many factors contributing to the desired conformations and molecular energetics [5]. In 2013, the discovery of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) technology swept the field of genome engineering for their extraordinary programmability and applicability. Our work, which was originally reported in 2016, represents one of the first NA detection methods based on the CRISPR technology. Since then, many CRISPR-based NA detection methods have emerged, in particular those utilizing the newly

* Corresponding author.
E-mail address: long.qian@pku.edu.cn (L. Qian).

https://doi.org/10.1016/j.ymeth.2021.06.003
Received 30 November 2020; Received in revised form 3 February 2021; Accepted 1 June 2021
Available online 4 June 2021

discovered Cas proteins such as Cas13a, Cas12a and Cas12b for the most recent work [6]. In 2016, the first CRISPR-based NA detection system fabricated on a strip of paper was published and used to genotype the Zika Virus which was the causative agent for a tremendous outbreak in Africa [7]. Reportedly, the entire design and test process of the paper-based detector took only 7 h, and the detector was able to distinguish similar strains with single nucleotide resolution, highlighting the programmability, specificity, low cost and ease of use contributed by CRISPR technology. In 2019, the most updated CRISPR-based NA detector had 1,000 times improvement in sensitivity and costed only $0.61. Up to now, several CRISPR-based NA detection technologies have been patented and are in the developmental pipeline for commercialization (https://sherlock.bio/)[8].

This article revisits our first generation of Paired dCas (PC) reporter system for pathogenic NA detection. We lay out detailed protocols and performance measures and give a comprehensive comparison between the PC reporter and later CRISPR-based NA detection technologies in the discussion section. In design, we combined the nuclease deactivated dCas9 and split firefly luciferase (termed NFluc and CFluc, respectively), with the former module conferring specificity of a total of ~ 46 bp precise target match, and the latter module generating a highly sensitive signal detectable to the naked eye. Specifically, when the NFluc-dCas9:sgRNA complex and the CFluc-dCas9:sgRNA complex simultaneously bound to adjacent sites on a target DNA, the two half fragments of luciferase tethered to each dCas9 were brought into proximity and the enzymatic activity was reconstituted, sending off a measurable bioluminescent signal (Fig. 1).

2. System Construction and Characterization

2.1. Protein production and purification

The codon-optimized DNA sequences of firefly luciferase (Fluc) and dCas9 were purchased from Generay Biotech (Shanghai) (For protein sequences, see Supplementary Fig. 1). *BsaI* restriction sites were inserted at both ends of each construct for cloning. NFluc-dCas9, CFluc-dCas9, dCas9-NFluc, and dCas9-CFluc were cloned into pET21a using Golden Gate Assembly with the linker between Fluc and dCas9 being 2xGGGGS. A 6 × His-tag was inserted into either N- or C- terminus of each construct for protein purification.

The dCas9 fusion protein constructs cloned into pET21a vectors were expressed in the *Escherichia coli* BL21 (DE3) strain. Transformants of the *E. coli* BL21 strain were grown overnight at 37 °C and diluted twice before induction to achieve higher expression level. Bacterial cultures were induced at OD_{600} 0.6–0.8 with 1 mM isopropyl β-D-1-thio-galactopyranoside (IPTG, ordered from J&K®) at final concentration of 200 μM and incubated at 16 °C overnight.

Cells were pelleted by centrifugation at 6,000 rpm (~4,200 g) for 10 min at 4 °C. The pellets were resuspended in lysis buffer (20 mM Tris, 500 mM NaCl, pH 8.0) and disrupted with ultrasonication. After centrifugation (15,000 rpm or ~ 15,000 g, 50 min, 4 °C), the supernatant was passed through 0.22 μm polyvinylidene fluoride (PVDF) filter (Millipore #SLGV033RS). Cleared supernatant was incubated in HisTrap™ HP column (Cytiva, #17–5248-02), and equilibrated with lysis buffer at 4 °C for 2 hr. The loaded column was washed with lysis buffer and eluted in the same buffer supplemented with 0–0.5 M imidazole gradient. The peak fractions containing the fusion proteins were then loaded onto HiTrap® desalting column (Cytiva, #17–1408-01) and washed with desalting buffer (20 mM HEPES, 150 mM KCl, pH7.5) to remove imidazole. The desalted sample was then concentrated with protein concentrator (Millipore #ACS510024) of 100 kDa cutoff.

For all purified proteins, the size and purity were confirmed by SDS-PAGE (Supplementary Fig. 2). The proteins were further concentrated to a range of 300–1000 μg/mL and quantified by Bradford protein assay (Tiangen, #PA102) with spectrophotometry (Thermo Scientific Multiskan™ FC Microplate Photometer), then aliquoted and stored in 50% glycerol at −80 °C.

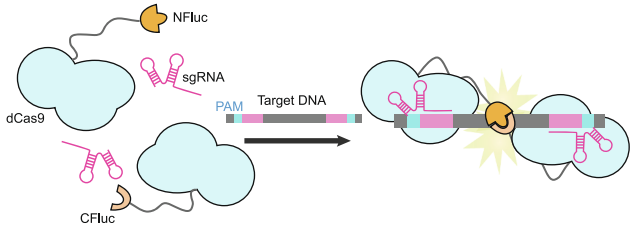

Fig. 1. Schematic of the paired dCas9 (PC) reporter system. dCas9 (light blue) was fused with fragments of split luciferase (NFluc and CFluc, orange), respectively, to form two kinds of split luciferase-dCas9:sgRNA complexes. In the presence of target DNA, the complexes bound to two adjacent recognition sites (pink), bringing NFluc and CFluc into proximity to produce bioluminescent signals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. sgRNA production and purification

sgRNA expression constructs were cloned into plasmid pSB1C3 (iGEM Parts Registry) by Gibson Assembly; and protospacers were then inserted into the plasmid by Golden Gate Assembly to generate sgRNA expression cassettes consisting of the T7 promoter, sgRNA, and T7 terminators (Supplementary Fig. S3). sgRNA was obtained from *in vitro* transcription (HiScribe T7 Quick High Yield RNA Synthesis Kit, NEB #E2050S) of the PCR linearized expression cassette DNA. The linearized sgRNA expression cassette DNA was mixed with T7 RNA polymerase and NTP in a reaction system (1 μg template DNA, 2 μL T7 RNA polymerase mix, 6.7 mM each NTP in 30 μL) and incubated at 37 °C for 2 hr. The template DNA was later removed by 2 μL DNaseI (2000 U/mL, NEB #M0303S) digestion at 37 °C for 15 min.

Synthesized sgRNAs were cleaned up by Biomed RNAclean Kit (Biomed RA109-01), and stored in DEPC treated water at −80 °C. The concentration of synthetized sgRNAs was quantified by spectrophotometry (Thermo Scientific NanoDrop™ 2000).

2.3. In vitro measurement of luminescence

Since the dCas9 fusion protein could be programmed by either one of the C-sgRNA and the N-sgRNA, wrong pairings between dCas9 fusion proteins and the sgRNAs would abrogate the correct target identity. Therefore, NFluc-dCas9 and CFluc-dCas9 were pre-incubated with their corresponding sgRNAs, respectively. To obtain Fluc-dCas9-sgRNA complex, purified Fluc-dCas9 protein and sgRNA were mixed and incubated at 25 °C for 10 min. Each Fluc-dCas9-sgRNA system for *in vitro* DNA detection consisted of 3 μL sgRNA (diluted to 1000 nM) and 2 μL 0.25 μM fusion protein in 8.5 μL of 1 × HEPES buffer (20 mM HEPES, 150 mM KCl, pH 7.5).

For the *in vitro* DNA detection assay, target DNA samples were added into the mixture of NFluc-dCas9:sgRNA and CFluc-dCas9:sgRNA reaction system (See Supplementary Fig. S4 for the primary target DNA sequence used for experimental optimization of the PC reporter system in Section 2.4 and verification in Section 2.5). Basically, each 13.5 μL reaction system of NFluc-dCas9:sgRNA and CFluc-dCas9:sgRNA were combined, and 3 μL diluted target DNA (30 nM, unless otherwise specified) was added into the mixture and incubated at 37 °C for 30 min (30 μL in total). The reconstituted luciferase binding target DNA was detected and measured by luminescence assay. The described mixture of Fluc-dCas9, sgRNA and target DNA was transferred to a prewarmed (to 37 °C) 96-well plate with black flat-bottom. Luminescence was measured immediately after the addition of 100 μL prewarmed (to 37 °C) Luciferase Assay Reagent (Promega #E4550) using microplate reader (Thermo Scientific Varioskan Flash). Of note, the black flat-bottom 96-well plates shouldn't be reused to avoid contamination

from residual split luciferase.

2.4. Optimization of PC reporter system

2.4.1. The location of His-Tag

Despite the fact that His-tags are widely used for protein purification, it is well-known that they might have considerable impacts on protein structure. Hence, we first tested the activity of C-terminally and N-terminally His-tagged split luciferase-dCas9 fusion proteins to see whether the location of His-tag mattered. We found that when the His-tag is attached to NFluc or CFluc, the activity of fusion protein dramatically decreased; this is probably due to the interruption of luciferase reconstitution by the His-tags. When the His-tag was attached to dCas9, the enzymatic activity was reconstituted (Fig. 2). Therefore, split luciferase-dCas9 fusion proteins with His-tagged dCas9 were chosen for further study.

2.4.2. Configuration of the PC reporter

Considering that dCas9 is a huge protein and is structurally complex, we needed to figure out the specific configuration the PC reporter should adopt. We first fused split luciferase to either the N- or the C-terminus of dCas9 (Fig. 3A) to test its reconstitution efficiency. Then, provided that the initial binding of dCas9 depends on the protospacer adjacent motif (PAM, a short 3′ motif adjacent to target sequence) [9] four sets of sgRNA orientation settings were also tested (Fig. 5B). One set had two PAM sequences placed distal from the spacer sequence between the sgRNA pair, with the 5′ end of the sgRNA adjacent to the spacer (PAM-out), while another set had both PAM sequences placed adjacent to the spacer (PAM-in). In the PAM-direct 1 and PAM-direct 2 sets, both sgRNA orientations were combined, with one PAM sequence adjacent to and another distal from the spacer (Fig. 3B). In total, 16 pairs of split luciferase-dCas9:sgRNA complexes were constructed and tested (Fig. 3C). Most of these pairs worked well. Among them, we selected NFluc-dCas9/CFluc-dCas9, the most robust protein construct with minimal intergroup difference, and “PAM-out”, the sgRNA orientation exhibiting the strongest luminescence, for further study.

As the reconstitution of enzymatic activity was constrained by the distance between NFluc and CFluc[10] we varied the interval length between the two target fragments from 5 bp to 107 bp (Fig. 4) to find how it affected the performance of the PC reporter system. Results showed that the PC reporter worked with a preference for spacer length of about 21 bp when two fusion proteins were in the best spatial position for reconstitution when bound on DNA. As the spacer length increased, the luminescence intensity decreased at the length around 28 bp and went stronger again at around 33 bp. In this periodic pattern, the signal peak and trough happened to appear per DNA helix turn. We speculated that this optimal 21 ± 2 bp spacing might be related to the length of the fusion protein linker. The luminescence was significantly dampened as the spacer length exceeded three helix turns, since two split luciferase fragments were too far apart for the reconstitution.

With all the above pre-testing, the final design of PC reporter system was determined, with both fragments of split luciferase fused to the N-terminus of dCas9, the PAM-out sgRNA orientation, and an optimal spacer length of 21 ± 2 bp.

2.5. Specificity and sensitivity test

We first carried out a specificity test using the PC reporter system to distinguish the target plasmid DNA (iGEM Part Registry BBa_K909009) from different kinds of non-specific controls. The results showed that the luminescence intensity of the positive target is significantly higher than a non-specific control DNA lacking one or both target segments (Fig. 5A). Also, a sensitivity test of PC reporter system was conducted with serially diluted target DNA samples, and the results show that the PC reporter reliably detected 0.05 nM of plasmid DNA or crude PCR product (Fig. 5B and 5C). Furthermore, we speculated that multiple PC reporters recognizing more sites on the same target DNA would result in enhanced bioluminescence and push the detection limit of the system. The results confirmed our speculation; however, the system was likely saturated as a large number of target segments was introduced (Fig. 5D).

2.6. Applying the PC reporter to crude amplified nucleic acid products

We took advantage of the programmability of sgRNA-targeting and demonstrated one example using the PC reporter to detect the genome of *Mycobacterium tuberculosis* (*Mtb*), which is the major pathogenic agent for human tuberculosis (Fig. 6). We wanted to figure out if the PC reporter had the potential to be applied to clinical cases where patients might be in an early stage of tuberculosis and thus genome copies of *Mtb* in their body fluid samples could be scarce. Serially diluted *Mtb* genomic DNA samples were amplified through 35 cycles of PCR and tested with the PC reporter. Equivalently one genome/500 μL (~30 aM) pre-PCR

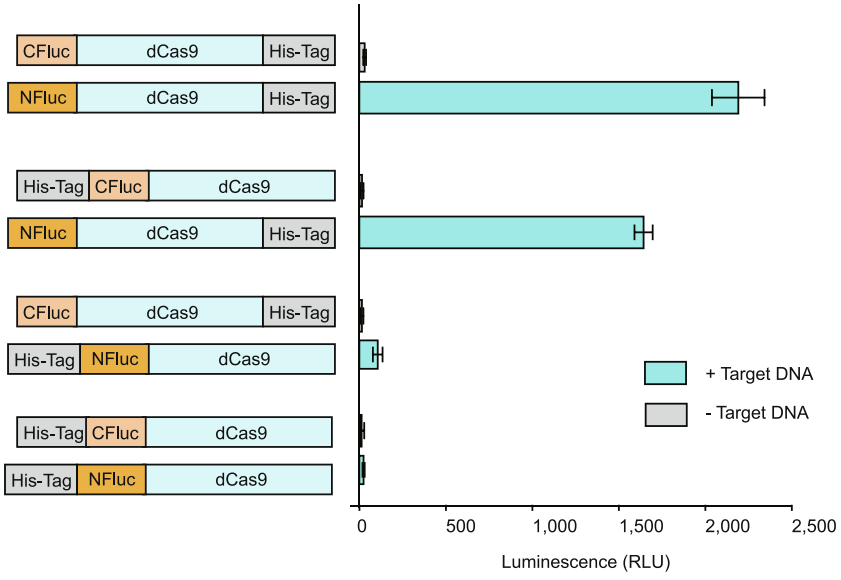

Fig. 2. The impact of His-tag location on the reconstitution of luciferase activity. When attached to dCas9, His-tag shows minimal impact on the complementation of NFluc and CFluc. Error bars denote s.e.; n = 3.

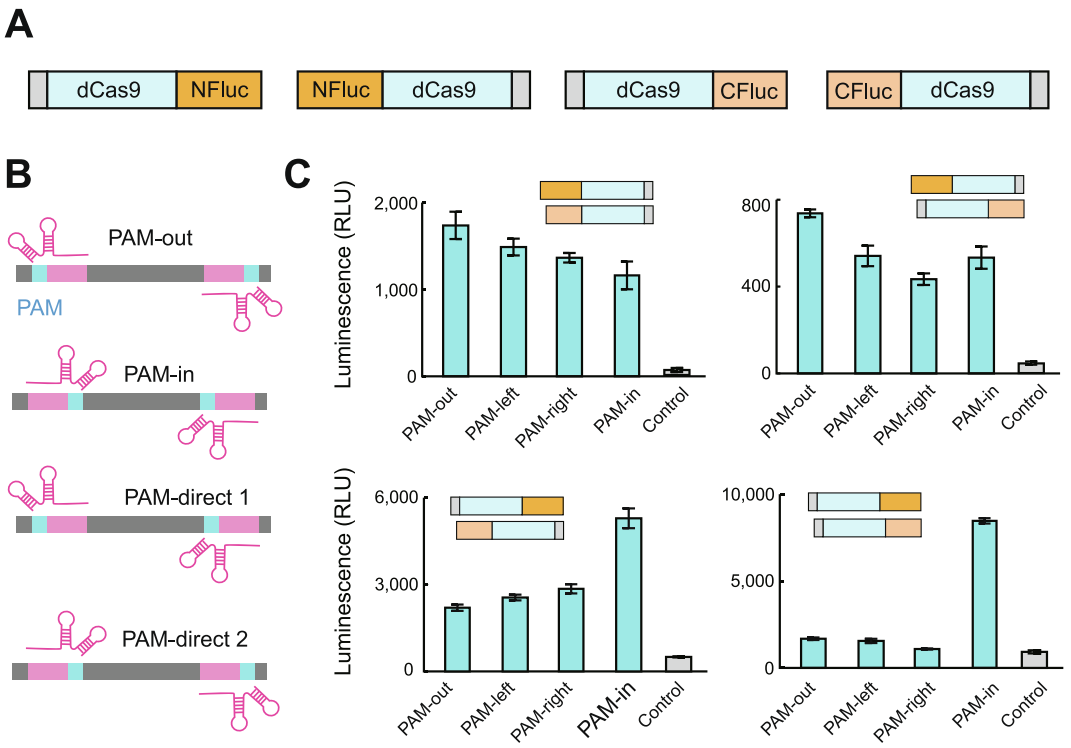

Fig. 3. Construction and optimization of PC reporter system. (A) Four different split luciferase-dCas9 fusion constructs. (B) Four different sgRNA orientation settings. In orientation PAM-out, the pair of PAM sequences were distal from the spacer sequence, with the 5′ end of the sgRNA adjacent to the spacer; in orientation PAM-in, the pair of PAM sequences were adjacent to the spacer sequence, with the 3′ end of the sgRNA in proximity to the spacer; in orientation PAM-direct 1 and PAM-direct 2, one PAM sequence was adjacent to and another distal from the spacer. (C) Experimental results of PC reporter systems of 16 different (4 different protein constructs × 4 different sgRNA orientation settings) configurations. Error bars denote s.e.; n = 3.

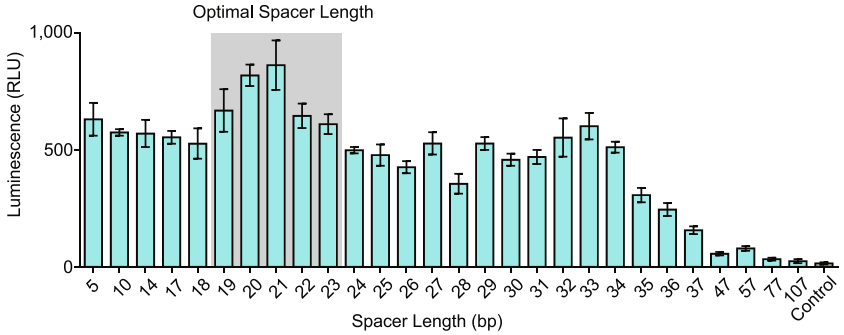

Fig. 4. The effect of spacer length variation on the performance of the PC reporter. Spacer was defined as the sequence between the sgRNA pairs. The spacer length varied from 5 bp to 107 bp. Note that the effect of spacer length exhibited a period of 10–11 bp, which was consistent with the structure of the DNA double helix. Error bars denote s.e.; n = 3.

was detected with 11.6-fold luminescence gain in the post-PCR reaction mixture, demonstrating excellent sensitivity for the test.

3. Array-Based detection

In clinical practice, the situations are much more complicated considering the fact that strain mutations, sample variations and other environmental factors would cause misdiagnosis in single target tests. Therefore, we designed a detection array to extract more sequence information by targeting multiple sites on the *Mtb* genome in one time.

3.1. Search for sgRNA pairs through the M. Tuberculosis genome

We developed an algorithmic pipeline to identify possible target sequence pairs and large genomic regions for the detection of a specific genome by the array-based PC reporter assay (Fig. 7A). The corresponding code package *MarkerFinder* can be found as a Supplementary Item, with detailed explanations of each module listed in Supplementary Information.

For target DNA recognition, the CRISPR/dCas9 system requires a PAM sequence in the form of 5′-NGG-3′ downstream the 20-bp segment that is complementary to the matching region of sgRNA. Since our experimental results showed that the PAM-out orientation (5′-CCN $(N)_{20}$-…-$(N)_{20}$NGG-3′) was highly efficient for the PC reporter system, we restricted our target finding algorithm to this particular orientation.

First, to find out all possible sgRNA-matching segments on *M. tuberculosis* H37Rv genome, we took advantage of Python 3.4.3 built-in regular expression to search for upstream (‘(?<=cc).(?=.{20})’) and downstream (‘(?<=.{20}).(?=gg)’) segments followed by the PAM sequence, separately. A human oral metagenome (HOMG) database was

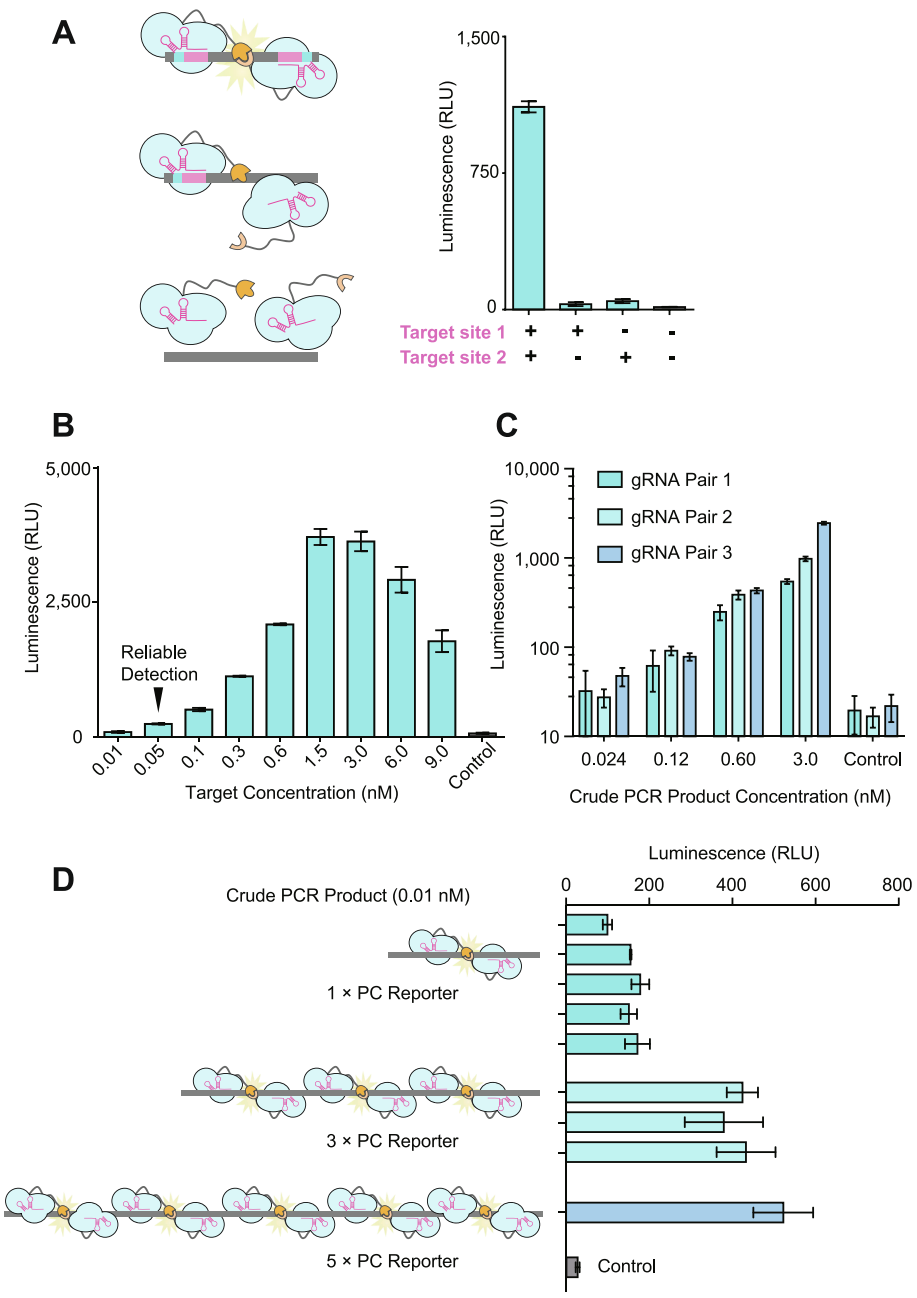

Fig. 5. Characterization of the PC reporter system. (A) Specificity test of the PC reporter. (B) Bioluminescent intensity at different concentrations of plasmid DNA as the target. (C) Bioluminescent intensity at different concentrations of crude PCR product as the target. (D) One, three, and five PC reporters were, respectively, subjected to the measurement using 0.01 nM PCR products of the target DNA. Five target sites were pre-designed and tested either individually (1x PC reporter) or in combinations (3 of 10 possible combinations for 3x PC reporter and all combined for 5x PC reporter). In A, B, C and D, error bars show s.e. of three experimental replicates.

constructed (249 rows, http://www.hmpdacc.org/HMRGD/) and all segment candidates were aligned with HOMG using blastn (BLAST-2.2.31+, NCBI) and biophyton-1.65 module. Parameters for BLAST were listed in Supplementary Table 1. Previous research has demonstrated that perfect base-pairing within 10–12 bp directly 5′ of the PAM sequence (PAM-proximal) determines Cas9 specificity, and Cas9 tolerates mismatches to a greater extent in the PAM-distal region than in the PAM-proximal region[11]. Thus, segment candidates containing 12-bp PAM-proximal sequence perfectly matching HOMG were directly eliminated.

The remaining segment candidates were BLAST'ed again to find possible off-target sites. Then, an algorithm from the CRISPR Design toolset (http://crispr.mit.edu/about) was used to score individual off-targets for each segment candidate.

$$S = \prod_{e \in M}(1 - W[e]) \times \frac{1}{(\frac{19-\bar{d}}{19} \times 4 + 1)} \times \frac{1}{n_m^2}$$

$$W = [0, 0, 0.014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613, 0.851, 0.732, 0.828, 0.615, 0.804, 0.685, 0.583]$$

In the first term, e denotes the position indices of mismatches between the segment and an off-target site, with W representing the experimentally-determined weights, i.e. deleterious effects, of mismatched positions on target recognition[11]. The second term refers to the effect of mean pairwise distance between mismatches (d). The third term represents a dampening penalty for highly mismatched off-target sites, where n_m refers to the number of mismatches. Once individual off-target sites were scored, each segment was assigned an overall score:

$$S_{guide} = \frac{1}{1 + \sum_{h \in T} S_{single}(h)}$$

where h runs over set Γ of potential off-target sites of a segment sequence. A higher overall score indicates a better segment with potentially no or very weak off-target effects. Segments with individual

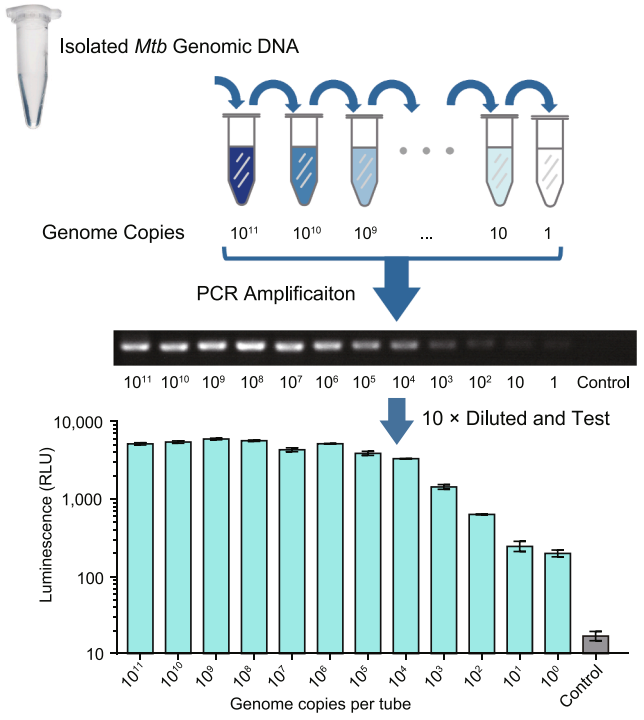

Fig. 6. Workflow and results of the PC reporter in detection of the *M. tuberculosis* (H37Rv) genome. Error bars show s.e. of three experimental replicates.

off-target site score 2% or above were eliminated, and the remaining were ranked in the order of overall score from 100% to 0% (Supplementary Fig. 5). Segments having an overall score greater than 45% were retained while others were filtered out for their lack of specificity.

Finally, according to experimental data, upstream and downstream segments with the appropriate spacer length (19–23 bp) were paired, and named as "markers". In total, 3751 markers were selected by this criterion. For experimental convenience, four evenly distributed regions, each of ~ 3 kb long, throughout the *Mtb* genome were selected and each region contained eight markers, yielding a total of 32 markers (Fig. 7B, Supplementary Table S2). A detection array for *B. subtilis* was designed likewise, and 72 of 2583 markers were selected and distributed in nine regions (Supplementary Fig. 6, Supplementary Table S3).

3.2. High-Throughput assay and scoring

Each well on the array plate were loaded with sgRNAs targeting a distinct pair of *Mtb* markers and preincubated with the PC reporters in separate rounds. Homogeneous solutions containing *Mtb* genomic DNA or the nonspecific *E. coli* genomic DNA were subsequently applied to this detection array followed by a statistical analysis on the luminescence matrices. For data analysis, the results were presented as heat maps (Fig. 7C). The pattern of *Mtb* was readily discriminated from that of *E. coli*. Note that among 32 successfully constructed sgRNAs, relative luminescence levels varied between wells as a result of differential performances of each individual target sequence. A similar result was obtained for the array-based detection of *B. subtilis* against the *E. coli* genomic DNA control (Supplementary Fig. 6).

This was consistent with our concern that a single PC reporter's performance could fluctuate due to uncontrolled factors. Reassuringly,

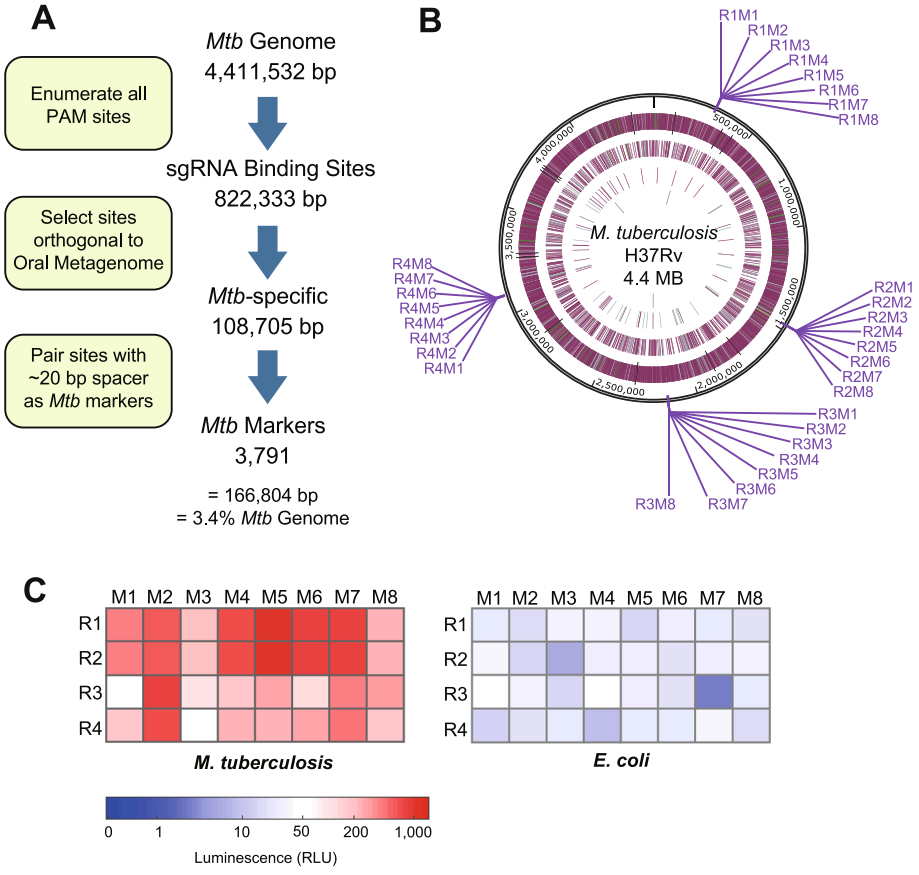

Fig. 7. Array-based detection of *Mtb*. (A) Workflow of *Mtb*-specific marker selection. (B) 32 *Mtb*-specific markers distributed among four regions in the *Mtb* genome. (C) Results of high-throughput assay for *Mtb* and *E. coli*. R denotes regions obtained from the *Mtb* genome (left) or the control genome (right); M denotes markers within each region.

the array-based PC reporter utilized more sequence information than single target detection to solve this problem. Not assuming an underlying distribution for the bioluminescent signals, we chose the nonparametric Wilcoxon Rank Sum Test of Block Design as a statistical analysis method. Data (bioluminescence) obtained from the array test were denoted X_{ij} and Y_{ij} for the experiment group (*Mtb*) and the control group (*E. coli*), respectively. The subscript i indicates the i-th marker of all m markers, and the subscript j indicates the j-th experiment in all n experimental replicates. In our case, $m = 32$ and $n = 3$. We took test results of a single marker, $\{X_i., Y_i.\}$, as a "block", and first calculated the Wilcoxon rank sum statistics W_i of the experiment group:

$$W_i = \sum_{j=1}^{n} Rank(X_{ij}; \{X_i., Y_i.\})$$

where $Rank(X; \Omega)$ denotes the rank of data X in the set Ω. The rank sum statistic W_i for equal sample sizes takes a range $[\frac{n(n+1)}{2}, \frac{n(3n+1)}{2}]$. For $n = 3$, it can be explicitly computed to have mean $E(W_i) = 10.5$ and variance $Var(W_i) = 5.25$. Note that for $n = 3$, the rank sum test is not sufficient to reject the null hypothesis that $X_i.$ and $Y_i.$ come from the same distribution, because both the minimal and maximal possible W_i (i.e. 6 and 15, respectively) correspond to a p-value = 0.1 (two-tailed). Therefore, we pooled W_i for all m markers (blocks) together. It is reasonable to assume W_i for all markers are independently and identically distributed random variables. Then, based on the central limit theorem, as$m \to \infty$, the random variable $W_{BD} = \frac{\sum_{i=1}^{m} W_i - mE(W_i)}{\sqrt{mVar(W_i)}}$ converges in distribution to a standard normal distribution $\mathcal{N}(0, 1)$. For the data of the *Mtb* array test in Fig. 7C, we obtained $W_{BD} = 9.44$, which corresponded to a p-value $= 8 \times 10^{-23}$ in the standard normal distribution.

The above analysis performed on data across a large number of markers allowed statistically significant statements to be made although single markers lacked the statistical power. With a carefully selected set of markers genome-wide, the array-based PC reporter system becomes robust to situations where either the pathogen DNA is present at trace concentrations or nonspecific DNA is ample to confound test results.

4. Conclusions and forward-looking statements

In conclusion, we designed the PC reporter system with a pair of dCas9 and split luciferase enzymes. The NA detection system benefited from the single base specificity and the superb re-programmability of dCas9, as well as the high sensitivity of enzymatic reporters with minimal background signal. The PC reporter was successfully applied to the detection of the pathogenic *M. tuberculosis* H37Rv genomic DNA.

Compared to other CRISPR-based detection methods, our paired dCas9 design extended the specificity to a 46-bp exact match with the appropriate spacing between upstream and downstream fragments. Upon target recognition, the reconstituted luciferase rapidly generated detectable bioluminescence, yielding a sensitivity of ~ 0.01 nM without nucleic acid amplification, or 30 aM with 35 cycles of PCR. This was on par with or slightly better than other CRISPR-based detection methods such as SHERLOCK[7,12] and HOLMES[8] (~0.1 nM without PCR). The readout time was within minutes. In Table 1, we compare key technical aspects of major CRISPR-based NA detection systems reported over the past five years.

We implemented the PC reporter in a parallel detection array and constructed an algorithmic pipeline for the automatic selection of genome-wide markers to identify the genome of interest. The array was used in this study to enhance the robustness of detection – it is expected to further push the PC reporter's detection limit and to allow for the testing of less-treated samples. Another potential use of the detection array is to assess the sequence diversity of a sample of possibly mixed DNA sequences. In this case, each individual marker will be against featured segments from distinct genomes of interest. The joint detection pattern could then be parsed by statistical or machine learning algorithms to elucidate the composition of the sample. Given the high specificity of the method, subtle nucleotide differences throughout the genome can be leveraged to resolve very closely related DNA sequences.

However, the first generation of PC reporter still faces limitations at this stage. Only purified nucleic acid samples were tested in our study, and we have not challenged the PC reporter with untreated samples directly collected from patients or the environment. In fact, most current CRISPR-based NA detection methods require sample pre-treatment, of which the HUDSON protocol used in [12] provides an exemplary attempt. Likewise, almost all current methods also depend on preliminary nucleic acid amplification (e.g., PCR or RT-RPA), which significantly contributes to the reported aM ~ fM sensitivities. Although signal amplification was implemented through enzymatic activities (as in our case and [13]) or by collateral cleavage by the newly discovered Cas12 and Cas13 proteins (as in [8,12,14]), the detection limits without amplification were usually in the nM range (Table 1). In this regard, electrochemical readout through novel nanomaterial may provide a solution. For example, a newly developed CRISPR-Chip device was reported to be able to detect ~ 1.7 fM target DNA without a pre-amplification step [15]. This was achieved by the integration of a graphene-based field effect transistor (gFET) which sensitively transformed dCas9 binding events to the underlying drain electrodes.

Nucleic acid detection empowered by CRISPR technology has its applications not only in pathogen detection but also in much broader biomedical applications ranging from prenatal screens and circulating tumor DNA detection to the genotyping of multidrug resistant strains, and even more general technologies involving reading out sequence information from biological or non-biological samples. The modular

Table 1
Comparison of CRISPR-based Diagnostic Methods.

Method	Target	Sample Type	Amplification	Readout	Sensitivity without Amplification	Specificity-enhancing Techniques
SHERLOCK[7]	ZIKA virus	Urine, Serum	2 hr RNA amplification	Paper-based visual detection	~30 nM	Primers for amplification, target recognition by Cas9 and toehold reaction
HUDSON + SHERLOCK [12]	ZIKA virus	Urine, Serum Saliva, Blood	20 min RPA	Fluorescence	/	Primers for amplification and target recognition by Cas13
HOLMESv2[8]	RNA virus, human cell mRNA and circular RNA	Urine	2 hr LAMP	Fluorescence	~1 nM	Primers for amplification and target recognition by Cas12b
RCH[13]	miRNA	Serum	2.5 hr isothermal amplification	Colorimetric readout	/	Probe for amplification, target recognition by dCas9, and paired dCas9 design
DETECTR[14]	HPV	Anal Swab, Cell Line	10 min RPA	1 hr detection, fluorescence	~0.1 nM	Primers for amplification, target recognition by Cas12a
PC Reporter (this work)	*M. tuberculosis*	Cell Line	<1 hr PCR amplification	Bioluminescence	~0.1 nM	Primers for amplification, target recognition by dCas9, and paired dCas9 design

design of the PC reporter grants ample room for improvement by replacing current modules by or coupling the system to novel parts or materials. For example, by introducing a separate PAM-presenting oligonucleotide (PAMer) [16] the PC reporter could be used for the detection of single strand RNA viruses such as SARS-CoV, HCV, and Ebola. The engineered or synthetically evolved Cas9 variants that overcome the constrains of the conventional PAM sequence (NGG) [17,18] can greatly expand the range of sequences that could be targeted. Finally, a variety of reporter proteins have been rationally engineered into reconstitutable pieces, including dihydrofolate reductase (DHFR) [19] β-Lactamase[20] and β-galactosidase[21]. These reporter candidates, when linked to the PC reporter, will be able to generate diverse signals such as pigmentation or electric currents that can be easily observed by naked eyes or sensitively quantified by microchips.

CRediT authorship contribution statement

Yihao Zhang: Methodology, Investigation, Formal analysis, Visualization, Writing - original draft, Writing - review & editing. **Yu Wang:** Investigation, Formal analysis, Writing - original draft. **Luze Xu:** Software, Writing - review & editing. **Chunbo Lou:** Conceptualization, Supervision, Methodology. **Qi Ouyang:** Conceptualization, Supervision. **Long Qian:** Conceptualization, Supervision, Methodology, Formal analysis, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Key R&D Program of China (No.2020YFA0906900).

Author Contributions

Q.O., C.L. and L.Q. conceived and supervised the study. Y.Z. and L.Q. conceived and organized the manuscript. Y.Z., Y.W. and L.X. drafted the System Construction and Characterization section of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymeth.2021.06.003.

References

[1] M. Arya, I.S. Shergill, M. Williamson, L. Gommersall, N. Arya, H.R. Patel, Basic principles of real-time quantitative PCR, Expert Review of Molecular Diagnostics. 5 (2014) 209–219, https://doi.org/10.1586/14737159.5.2.209.

[2] Y. Zhao, F. Chen, Q. Li, L. Wang, C. Fan, Isothermal Amplification of Nucleic Acids, Chem. Rev. 115 (2015) 12491–12545, https://doi.org/10.1021/acs.chemrev.5b00428.

[3] O. Piepenburg, C.H. Williams, D.L. Stemple, N.A. Armes, DNA Detection Using Recombination Proteins, PLoS Biology. 4 (2006), e204, https://doi.org/10.1371/journal.pbio.0040204.

[4] R. Martzy, C. Kolm, R. Krska, R.L. Mach, A.H. Farnleitner, G.H. Reischer, Challenges and perspectives in the application of isothermal DNA amplification methods for food and water analysis, Anal Bioanal Chem. 411 (2019) 1695–1702, https://doi.org/10.1007/s00216-018-1553-1.

[5] A. Moutsiopoulou, D. Broyles, E. Dikici, S. Daunert, S.K. Deo, Molecular Aptamer Beacons and Their Applications in Sensing, Imaging, and Diagnostics, Small. 15 (2019) 1902248, https://doi.org/10.1002/smll.201902248.

[6] Y. Li, S. Li, J. Wang, G. Liu, CRISPR/Cas Systems towards Next-Generation Biosensing, Trends Biotechnol. 37 (2019) 730–743, https://doi.org/10.1016/j.tibtech.2018.12.005.

[7] K. Pardee, A.A. Green, M.K. Takahashi, D. Braff, G. Lambert, J.W. Lee, et al., Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components, Cell. 165 (2016) 1255–1266, https://doi.org/10.1016/j.cell.2016.04.059.

[8] L. Li, S. Li, N. Wu, J. Wu, G. Wang, G. Zhao, et al., HOLMESv2: A CRISPR-Cas12b-Assisted Platform for Nucleic Acid Detection and DNA Methylation Quantitation, ACS Synth Biol. 8 (2019) 2228–2237, https://doi.org/10.1021/acssynbio.9b00209.

[9] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, E. Charpentier, A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity, Science. 337 (2012) 816–821, https://doi.org/10.1126/science.1225829.

[10] J.P. Guilinger, D.B. Thompson, D.R. Liu, Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification, Nat. Biotechnol. 32 (2014) 577–582, https://doi.org/10.1038/nbt.2909.

[11] P.D. Hsu, D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, V. Agarwala, et al., DNA targeting specificity of RNA-guided Cas9 nucleases, Nat. Biotechnol. 31 (2013) 827–832, https://doi.org/10.1038/nbt.2647.

[12] C. Myhrvold, C.A. Freije, J.S. Gootenberg, O.O. Abudayyeh, H.C. Metsky, A. F. Durbin, et al., Field-deployable viral diagnostics using CRISPR-Cas13, Science. 360 (2018) 444–448, https://doi.org/10.1126/science.aas8836.

[13] X.-Y. Qiu, L.-Y. Zhu, C.-S. Zhu, J.-X. Ma, T. Hou, X.-M. Wu, et al., Highly Effective and Low-Cost MicroRNA Detection with CRISPR-Cas9, ACS Synth Biol. 7 (2018) 807–813, https://doi.org/10.1021/acssynbio.7b00446.

[14] J.S. Chen, E. Ma, L.B. Harrington, M. Da Costa, X. Tian, J.M. Palefsky, et al., CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity, Science. 360 (2018) 436–439, https://doi.org/10.1126/science.aar6245.

[15] R. Hajian, S. Balderston, T. Tran, T. deBoer, J. Etienne, M. Sandhu, et al., Detection of unamplified target genes via CRISPR–Cas9 immobilized on a graphene field-effect transistor, Nat, Biomed Eng. 3 (2019) 427–437, https://doi.org/10.1038/s41551-019-0371-x.

[16] M.R. O'Connell, B.L. Oakes, S.H. Sternberg, A. East-Seletsky, M. Kaplan, J. A. Doudna, Programmable RNA recognition and cleavage by CRISPR/Cas9, Nature. 516 (2014) 263–266, https://doi.org/10.1038/nature13769.

[17] B.P. Kleinstiver, M.S. Prew, S.Q. Tsai, V.V. Topkar, N.T. Nguyen, Z. Zheng, et al., Engineered CRISPR-Cas9 nucleases with altered PAM specificities, Nature. 523 (2015) 481–485, https://doi.org/10.1038/nature14592.

[18] J.H. Hu, S.M. Miller, M.H. Geurts, W. Tang, L. Chen, N. Sun, et al., Evolved Cas9 variants with broad PAM compatibility and high DNA specificity, Nature. 556 (2018) 57–63, https://doi.org/10.1038/nature26155.

[19] L.M. Gonçalves, W.F.A. Callera, M.D.P.T. Sotomayor, P.R. Bueno, Penicillinase-based amperometric biosensor for penicillin G, Electrochemistry Communications. 38 (2014) 131–133, https://doi.org/10.1016/j.elecom.2013.11.022.

[20] A. Galarneau, M. Primeau, L.-E. Trudeau, S.W. Michnick, β-Lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein–protein interactions, Nat. Biotechnol. 20 (2002) 619–622, https://doi.org/10.1038/nbt0602-619.

[21] W.A. Mohler, H.M. Blau, Gene expression and cell fusion analyzed by lacZ complementation in mammalian cells, Proceedings of the National Academy of Sciences. 93 (1996) 12423–12427, https://doi.org/10.1073/pnas.93.22.12423.