

# **Clustered regularly interspaced short palindromic repeats (CRISPR)- Diagnostics for differential detection of haemoglobin variant S, C, D and E.**

Patrick Adu<sup>1,2#</sup>, Prabhleen Kaur<sup>1,3</sup>, Prosad Kumar Das<sup>1,3</sup>, Parikshit Chattopadhyay<sup>1</sup>,  
Afzal C<sup>1</sup>, Debojyoti Chakraborty<sup>1,3#</sup>

<sup>1</sup>Division of Chemical and Systems Biology, CSIR-Institute of Genomics and Integrative Biology, Mathura Road, New Delhi – 110025, India.

<sup>2</sup>Department of Medical Laboratory Science, School of Allied Health Sciences, College of Health and Allied Sciences, University of Cape Coast, Cape Coast, Ghana.

<sup>3</sup>Academy of Scientific and Innovative Research (AcSIR), Ghaziabad – 201002, India.

**#Correspondence:** Patrick Adu ([patrick.adu@ucc.edu.gh](mailto:patrick.adu@ucc.edu.gh)) or Debojyoti Chakraborty ([debojyoti.chakraborty@igib.in](mailto:debojyoti.chakraborty@igib.in))

## Abstract

Although haemoglobin variants are prevalent in low- and middle-income countries, the exact disease burden remains unknown due to a lack of diagnostic capacity. Traditionally, routine clinical haemoglobin variant diagnostics have relied on electrophoresis, which separates the haemoglobins based on size and charge differences. However, electrophoresis-based assays are limited in depth and coverage due to their inability to separate co-migrating variants at the pH employed. Importantly, for genetic counselling of pre-marital couples and prenatal screening of inherited haemoglobinopathy risk, molecular-based assays are required. Here, we leveraged the specificity and dual mismatch intolerance of en31FnCas9 to achieve differential identification of haemoglobin variants S, C, D, and E. Moreover, we demonstrate reliable differential detection of homozygous, heterozygous and compound states of haemoglobin variants due to en31FnCas9's intolerance to dual mismatch in the respective gRNA-target DNA complementarity. Furthermore, we coupled our en31FnCas9-based haemoglobin variant detection to the signal enhancement of recombinase polymerase amplification (RPA) to achieve reliable differential detection of haemoglobin variants from non-invasive saliva and urine samples in <60 minutes. Taken together, our study demonstrates the feasibility of functionalization of CRISPR-based diagnostics as a point-of-care technique towards achieving the democratisation of haemoglobin variant diagnosis in resource-limited settings.

## Introduction

Haemoglobinopathies are genetically inherited disorders of haemoglobin that may result in reduced haemoglobin quantities (thalassaemia) or structural haemoglobin types other than what is physiological for an individual's age (structural variants). With approximately 7% of the global population (>500 million people) estimated to be carrying haemoglobinopathies [1, 2], haemoglobinopathies are a significant public health problem. Although various hypotheses have been proposed as the basis for the disproportionate sub-Saharan Africa (SSA) and South-East Asia (SEA) representation in the haemoglobinopathies prevalence data, the evolutionary-based theory of selective pressures imposed by the *Plasmodium* parasites may provide a plausible explanation for the co-selection of haemoglobinopathies in *Plasmodium falciparum* endemic regions [3]. The presentations of the various haemoglobinopathies are clinically variable. Thalassaemia, a quantitative globin chain disorder, is haematologically characterised by microcytic red blood cells and increased RBC haemolysis due to precipitation of excess beta or alpha globin chains [4]. Contrastingly, structural haemoglobin variants present as normocytic haemolytic anaemia with variable clinical presentation depending on the exact haemoglobin variant inherited [5]. The most common beta haemoglobin variants of clinical significance are S (an exon 1 point mutation, Glutamic acid to Valine substitution,  $\beta^{6GAG \rightarrow GTG}$ ), C (an exon 1 point mutation, Glutamic acid to Lysine substitution,  $\beta^{6GAG \rightarrow AAG}$ ), D (an exon 3 point mutation, Glutamic acid to Glutamine substitution,  $\beta^{121GAA \rightarrow CAA}$ ), and E (an exon 1 point mutation, Glutamic acid to Lysine substitution,  $\beta^{26GAG \rightarrow AAG}$ ). It should be noted that since the inheritance of each of these haemoglobin variants is independent of each other, compound inheritance is a real clinical possibility, particularly in endemic areas.

Since these haemoglobinopathies are genetically inherited, clinical diagnosis should ideally be on genotype-based platforms. These genotype-based assays, such as PCR-based assays (e.g. allele-specific PCR, Gap-PCR) and various sequencing platforms, are, however, not routinely available in regions where haemoglobinopathies are prevalent. In SSA and across the majority of Low- and Middle-Income Countries (LMICs), the lack of molecular-based diagnostic capacity

across the healthcare ecosystem has necessitated over-reliance on phenotype-based assays such as haemoglobin solubility and cellulose acetate electrophoresis [6, 7]. Although these phenotype-based assays are economical, their inherent limitations are such that resolving compound inherited variants is not always possible. For example, haemoglobin variants S, D and G are known to co-migrate under the alkaline conditions employed in cellulose acetate electrophoresis, just as haemoglobin variants C, E and O also co-migrate [7, 8]. Thus, in the absence of additional diagnostic assays, underdiagnosis of patients in resource-limited settings is a practical clinical challenge. To overcome this limitation of the routinely employed cellulose acetate electrophoresis, some healthcare facilities combine it with acid electrophoresis, which not only increases the sample-to-result time but also increases the cost for patients. Efforts to expand accessibility to affordable diagnostics have seen the deployment of many point-of-care test devices like the Gazelle, HemoTypeSC, and Sickie Scan; however, none of these overcome the inherent limitations of the clinically widely used cellulose acetate electrophoresis method [9, 10]. Given the diagnostic challenges, it is nearly impossible to estimate the epidemiology and/or disease burden of haemoglobinopathies in LMICs. Consequently, the absence of such concrete data also means that estimating the economic contribution of haemoglobinopathies to the healthcare system in LMICs to aid planning is not an easy task.

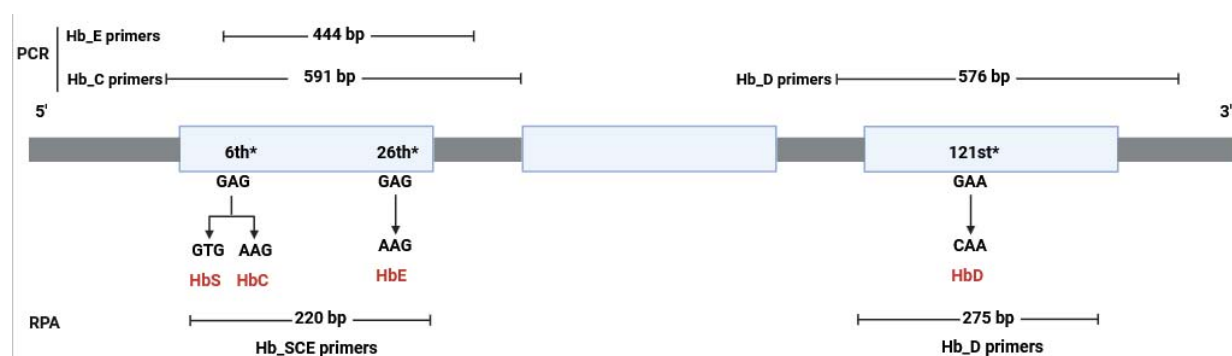
CRISPR-based diagnostic platforms (CRISPR-Dx) offer unique opportunities for expanding molecular-based diagnostic access to resource-limited settings. Given the specificity of the guide RNA (gRNA) to its target nucleic acid sequences, CRISPR-based assays can be used to enable differential detection of haemoglobin variants in heterozygous, single homozygous, or compound inherited haemoglobin variant states. Notably, >1000 point mutations [5, 11] have been identified as leading to a variety of haemoglobin variants, creating a clinical scenario that necessitates diagnostic platforms with functionalized multiplexing potential for exhaustive detection. Previously, our laboratory demonstrated that FnCas9, based on its sensitivity to mismatch intolerance [12], can enable the detection of sickle haemoglobin (Hb S) in the FELUDA assay [13]. Here, we demonstrate how the engineered variant-31 of FnCas9 (en31FnCas9) [14] can differentially detect haemoglobin variants S, C, D and E as well as wild-type sequences of the respective

regions of these variants. Additionally, we demonstrate the feasibility of coupling haemoglobin variant detection to recombinase polymerase amplification (RPA) to minimise the test turnaround time to facilitate the potential deployment as a point-of-care system. Together, we anticipate that further optimisation of our reported haemoglobin variant detection assay could lead to a robust, sensitive and specific molecular-based assay that can be implemented as a POCT to enable population-based screening in a proactive approach towards reducing the penetrance of haemoglobinopathies.

## Materials and methods

### Study design

The schematic of the relative positions of the haemoglobin variants in the specific exon of the beta globin chain is illustrated in Figure 1. The primer sequences used for the study reported herein are shown in Table 1.



**Figure 1: A schematic of the relative positions of the haemoglobin S, C, D, and E variants in the beta globin chain.** The Figure is not drawn to scale. (RPA: Recombinase Polymerase Amplification; PCR: Polymerase Chain Reaction; \*indicates the codon within which the respective haemoglobin variant point mutation occurs). The figure was created in <https://BioRender.com>.

**Table 1: Oligonucleotides used in the present study**

Name	Sequence
gRNA	
HbS_2_6	AGTAACGGCAGACTTATCCAC
HbC_15_18	CCTTAGTAGTCAGATGCACCA
HbC_17_18	CCTTCGGAGTCAGATGCACCA
HbC_16_18	CCTTATGAGTCAGATGCACCA
Diag_C_16C	CCTCATGAGTCAGATGCACCA
HbD_15_19	AACAATGCACCCACCAAGTGC
HbD_17_19	AACAGTTCACCCACCAAGTGC
Diag_HbD_16C	AAGAACTCACCCACCAAGTGC

HbE_2_6	CGTGGATGAAGTTGGCGGTAA
HbE_15_16	GTGGTACGGCCCTGGGCAGGT
HbE_16_17	GTGGCAAGGCCCTGGGCAGGT
HbE_16_19T	GTIGTAAGGCCCTGGGCAGGT
Hb_E_14_16	GTGGTAATGCCCTGGGCAGGT
Diag_HbE_17C	GTGGCGAGGCCCTGGGCAGGT
<b>PCR primers</b>	
Hb_C_591_FP	TCCTAAGCCAGTGCCAGAAG
Hb_C_591_RP	CCTTGAGGTTGTCCAGGTGAG
Hb_D_576_FP	GTA TCA TGC CTC TTT GCA CC
Hb_D_576_RP	GCA CTG ACC TCC CAC ATT CC
Hb_E_444_FP	GCTGTCATCACTTAGACCTC
Hb_E_444_RP	CCAAAGGACTCAAAGAACCT
<b>Site-directed mutagenesis</b>	
HbC_SDM_FP	GGTGCATCTGACTCCTAAGGAGAAGTCTGCCGTTACTGCCCTG
HbC_SDM_RP	CAGGGCAGTAACGGCAGACTTCTCCTTAGGAGTCAGATGCACC
HbD_SDM_FP	CCATCACTTTGGCAAACAATTCACCCACCAAGTGCAGGCTGCC
HbD_SDM_RP	GGCAGCCTGCACTGGTGGGGTGAATTGTTTGCCAAAGTGATGG
HbE_SDM-FP	GAACGTGGATGAAGTTGGTGGTAAGGCCCTGGGCAGGTTGG
HbE_SDM-RP	CCAACCTGCCCAGGGCCTTACCACCAACTTCATCCACGTTT
<b>RPA primers</b>	
<b>Hb_SCE region</b>	
RPA_219bp_ASCE_FP2	CAGGGCAGAGCCATCTATTGCTTACATTTGC
RPA_219_ASCE_RP2	CATGCCCAGTTTCTATTGGTCTCCTTAAACC
RPA_220bp_ASCE_FP	ACATTTGCTTCTGACACAACCTGTGTTCACTAG
RPA_220bp_ASCE_RP	CCCAAGAGTCTTCTCTGTCTCCACATGCCAG
RPA3_185_ASCE_FP	GCTTCTGACACAACCTGTGTTCACTAGCAACCTC
RPA3_185_ASCE_RP	CCAGTTTCTATTGGTCTCCTTAAACCTGTCTTG
<b>Hb_D region</b>	
RPA_123bp_D_FP2	CAGCTCCTGGGCAACGTGCTGGTCTGTGTGC
RPA_123bp_D_RP2	CTTGTGGGCCAGGGCATTAGCCACACCAGC
RPA_131_D_FP	CACAGCTCCTGGGCAACGTGCTGGTCTGTG
RPA_131bp_D_RP	GTGATACTTGTGGGCCAGGGCATTAGCCAC
RPA_D_319bp_FP	GCTAATAGCAGCTACAATCCAGCTACCATTC
RPA_D_319bp_RP	TAGTAGTTGGACTTAGGGAACAAAGGAACC
RPA_D_234bp_FP	GGTTGGGATAAGGCTGGATTATTCTGAGTC
RPA_D_234bp_RP	ATTGGACAGCAAGAAAGCGAGCTTAGTGATAC
RPA_D_275_bp_FP	GCTAATAGCAGCTACAATCCAGCTACCATTC
RPA_D_275_bp_RP	ATTGGACAGCAAGAAAGCGAGCTTAGTGATAC

Diagnostic HbC\_16C, diagnostic HbD\_16C, and HbE\_17 refer to gRNA synthesised to specifically cleave the respective WT sequence of the haemoglobin variant but not the haemoglobin variant itself. The underlined nucleotides indicate mismatches; the numbered positions (gRNA) in bold print indicate the position where the point mutation causing the respective haemoglobin variant occurs. In contrast, the numbered position in normal font indicates where a mismatch was introduced to enable differentiation from the wild-type sequence.

## En31Fncas9 Protein purification

The proteins used in this study were purified as reported previously [12]. The plasmid for the en31FnCas9 protein was expressed in *Escherichia coli* Rosetta2 (DE3) (Novagen), and cultured at 37°C in LB medium (supplemented with 50 mg/ml

kanamycin and 35 mg/ml chloramphenicol) until OD<sub>600</sub> reached 0.6. Subsequently, protein expression was induced by the addition of 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG). The Rosetta2 (DE3) cells were then cultured overnight (at 18 °C) and harvested by centrifugation. After resuspension of the cells in a lysis buffer (20 mM HEPES, pH 7.5, 500 mM KCl, 5% glycerol supplemented with 1× PIC [Roche], 100 µg/mL lysozyme), the *E. coli* cells were lysed by sonication. After centrifugation, the lysate was affinity-purified by Ni-NTA beads (Roche). The eluted en31FnCas9 protein was further purified by Ion exchange chromatography in which we passed the diluted eluate through pre-equilibrated SP Sepharose beads (Cytiva). The concentration of purified protein was measured by a Pierce BCA protein assay kit (Thermo Fisher Scientific), keeping BSA as a standard. The purified proteins were stored at -80 °C until further use. For all *in vitro* cleavage assays, the en31FnCas9 was reconstituted at 5 µM concentrations, and 1 µl was used in a 10 µl IVC reaction to achieve a 500 nM protein concentration.

### **Oligonucleotides used in the study**

All the oligonucleotides (Table 1) used in this study were synthesised by Eurofins Genomics, India.

### **Plasmid Construction**

Human Genomic DNA was extracted from the blood using the DNeasy Blood & Tissue kit (Qiagen) as per the manufacturer's instructions. WT sequences corresponding to 444 bp (for Hb S and E), 591 bp (Hb C), and 576 bp (Hb D) were PCR-amplified and column-purified for subsequent cloning experiments. The column-purified PCR products of the respective WT sequences of the regions where the haemoglobin variant (S, C, D and E) point mutation occurs were cloned into the TOPO-TA vector (Thermo Fisher Scientific) and then used for generating haemoglobin variants through site-directed mutagenesis.

### **Site-directed mutagenesis to generate haemoglobin variants S, C, D, and E:**

Haemoglobin variants S, C, D, and E were generated within the TOPO-TA plasmid backbone (Acharya et al., 2019) by QuickChange II site-directed mutagenesis (SDM) kit (Agilent) following the manufacturer's protocol. The success of the SDM-generated haemoglobin variants was confirmed through Sanger sequencing.

## **Sanger Sequencing:**

The sequencing reaction was carried out in 10µl volume (containing 0.5µl purified DNA, 0.8µl sequencing reaction mix, 2µl 5X dilution buffer, and 0.6µl forward/reverse primer) using the Big Dye Terminator v3.1 cycle sequencing kit (ABI, 4337454). The amplification conditions were: 3 min denaturation at 95°C, followed by 40 cycles of (10 secs at 95°C, 10 secs at 55°C, 4 min at 60°C) and 10 min extension. Subsequently, the PCR product was purified by mixing with 12µl of 125 mM EDTA (pH 8.0) and incubating at RT for 5 min. 50µl of absolute ethanol and 2µl of 3M NaOAc (pH 4.8) were then added, incubated for 10 min at RT and centrifuged for 30 min at 3800rpm, followed by an invert spin at <300rpm to discard the supernatant. The pellet was washed twice with 100 µl of 70% ethanol at 4000rpm for 15 min, and the supernatant was discarded by invert spin. The pellet was air-dried, dissolved in 12 µl of Hi-Di formamide (Thermo Fisher, 4311320), denatured at 95°C for 5 min followed by a snap chill, and linked to an ABI 3130xl sequencer. Base-calling was carried out using sequencing analysis software (v5.3.1) (ABI, US), and the sequence was analyzed using Chromas v2.6.5 (Technelysium, Australia).

## **NaOH-based genomic DNA extraction for RPA assay**

Genomic DNA (gDNA) extraction from human blood, saliva, or urine was performed through a previously published manual NaOH extraction protocol [15]. Briefly, saliva or urine was mixed with an equal volume of 1X PBS and centrifuged at 13000 rpm. For blood samples, an equal volume of RNase-free water was added to haemolyse the red blood cells, and centrifuged to sediment nucleated cells. After discarding the supernatant, this was followed by three 1X PBS washes. After the final washing, the pellet was lysed with 40µl of 0.2M NaOH and incubated at 70°C for 10 minutes for urine and saliva pellet, (5 minutes for the blood pellet). Subsequently, the reaction was neutralised with 180µl of 0.04 Tris-HCl (pH 7.5), and centrifuged at 13000rpm, after which the supernatant was transferred into a fresh vial. A total volume of 1 – 2 µl supernatant was used in the PCR or RPA reaction or stored at -20 °C until further use.

## **Annealing-extension PCR for making sgRNA template**



For each gRNA production, DNA template for *in vitro* transcription (IVT) was produced through overlap extension PCR by assembling the following on ice: 12.5 µl forward primer (10 µM), 12.5 µl reverse primer (10 µM), 5 µl 10X PCR buffer, 4 µl dNTPs (2.5mM), 0.5 µl Taq pol, 15.5 µl nuclease-free water to make a total volume of 50 µl reaction. The forward primer was unique for each haemoglobin variant; the reverse primer was, however, common for all the haemoglobin variants (Table 1). The amplification conditions were: Initial denaturation (95°C, 3min), 35 cycles of 95°C (10s), 55°C (30s), 72°C (30s), followed by 7 minutes' final extension. The size of the PCR product was verified on a 2% agarose gel.

### ***In Vitro* Transcription (IVT) Reaction**

All the gRNAs for haemoglobin variants were 21-nucleotide sequences with a single base mismatch insertion to ensure that although there was only a single mismatch with the haemoglobin variant, there was a dual mismatch with the WT sequence (see Table 1). *In vitro* transcription for gRNAs/crRNAs was done using the MegaScript T7 Transcription kit (ThermoFisher Scientific) following the manufacturer's protocol. Briefly, IVT was set up in a 20 µl total volume comprising 2µl of 10X Reaction Buffer, 2µl of UTP (75mM), 2µl of ATP (75mM), 2µl of GTP (75mM), 2µl of CTP (75mM), 5µl extension PCR product as template, 2µl T7 polymerase and 3 µl of nuclease-free water as per the manufacturer's protocol. IVT reactions were incubated at 37°C overnight (>16 hours) for sgRNA synthesis. Following DNase I treatment (Thermo Fisher Scientific) to remove residual DNA, gRNA clean-up was undertaken using Monash RNA purification columns as per the manufacturer's protocol. The respective gRNA concentration was quantified on a NanoDrop spectrophotometer (Thermo Fisher Scientific). Purified IVT sgRNAs/crRNAs were stored at -20°C until further use. For all *in vitro* cleavage assays, the sgRNA was reconstituted at 5 µM concentrations, and 1 µl was used in a 10 µl IVC reaction to achieve a 500 nM sgRNA concentration (unless otherwise stated).

### **Detection via *In Vitro* Cleavage (IVC)**

PCR-amplified DNA products of respective haemoglobin variants were used as substrates for *in vitro* cleavage assays as previously published [12]. 9 µl RNP of the en31FnCas9 protein and respective gRNA was reconstituted by combining 10x CutSmart buffer (1µl), 50mM MgCl (2µl), 5 µM en31FnCas9 protein (1µl), 5 µM gRNA (1.1 µl), and 3.9 µl nuclease-free water (at 25°C for 20 minutes). For each IVC

assay, 1  $\mu$ L of purified target DNA amplicon (500 nM to achieve a final 50 nM concentration in a 10  $\mu$ L reaction volume) was added to the preformed RNP and incubated at 37°C for 5 – 30 minutes, as the case may be; cleaved products were visualised on the 2% agarose gel.

## RPA

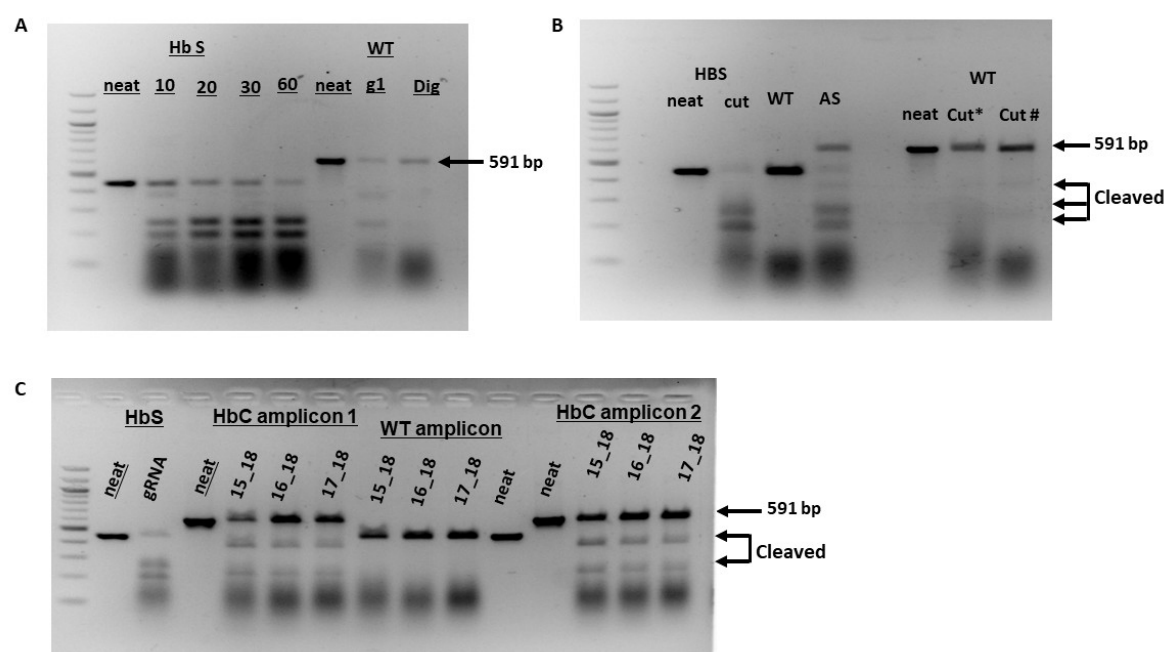
For screening of RPA primers, three forward primers and three reverse primers flanking the site for gRNA for Hb S, C, and E regions (for Hb D region, five primer pairs were tested) were designed (Figure 1 for location and Table 1 for primer sequence). Each primer was 30 – 34 bp long, and amplicons resulting from all primer combinations were between 120 and 319 bp long. The TwistAmp® Basic kit was purchased from TwistDx Ltd. (Hertfordshire, AL, UK). Basic Kit (TwistDx) and used for isothermal amplification of gDNA following the manufacturer's protocol with few modifications. Briefly, RPA reactions were undertaken in a final 25  $\mu$ L volume comprising 20  $\mu$ L rehydrated pellet (29.5  $\mu$ L of primer-free rehydration buffer, one pellet and 9.5  $\mu$ L nuclease-free water), 1  $\mu$ L 0.25  $\mu$ M forward, 1  $\mu$ L 0.25  $\mu$ M reverse primer, 2  $\mu$ L of NaOH-extracted gDNA, and 1  $\mu$ L 280mM magnesium acetate (last added). After vortexing and a brief spin of the mixture, it was incubated at 37 °C for 20 – 30 minutes. Subsequently, the amplification products were either column-purified and resolved on a 2% agarose gel or used directly for en31FnCas9–sgRNA IVC reaction.

## Results

### Differential detection of haemoglobin S, and C variants from wild-type sequence.

The point mutations leading to haemoglobin S ( $\beta^{6GAG \rightarrow GTG}$ ) and C ( $\beta^{6GAG \rightarrow AAG}$ ) variants occur in the sixth codon of the beta globin chain. Thus, differential detection of these haemoglobin variants requires a highly specific assay, given that there is only one nucleotide difference between Hb S and C. Here, we show that when using PAM distal gRNA for Hb S variant that has a single mismatch introduced at position 6 [12], en31FnCas9 enables *in vitro* differential detection of Hb S from wild-type haemoglobin sequence (Figure 2A & B). In Figure 1B, we mimicked the

heterozygous Hb AS inheritance by combining the Hb S and WT sequence of the Hb S region in a single tube and then explored whether using a combinatorial gRNA for Hb S and WT could aid differential detections of each of the amplicons. The specificities of the gRNA enabled distinct identification of heterozygous Hb AS, indicating that these gRNAs do not interfere with each other when used simultaneously in a single IVC reaction (Figure 2B). Furthermore, under our assay conditions, detection of Hb variants could be visually seen on an agarose gel from 10 minutes of 37°C incubation (Figure 2A). Moreover, when using PAM distal gRNA specific to Hb C variant with a single mismatch introduced at the 15, or 16, or 17<sup>th</sup> position, the en31FnCas9 enabled differential detection of Hb C variant (Figure 2C). Furthermore, we also demonstrate that the gRNA used for these Hb variant detections did not show cleavage towards the wild-type sequence of the same region (Figure 2C).



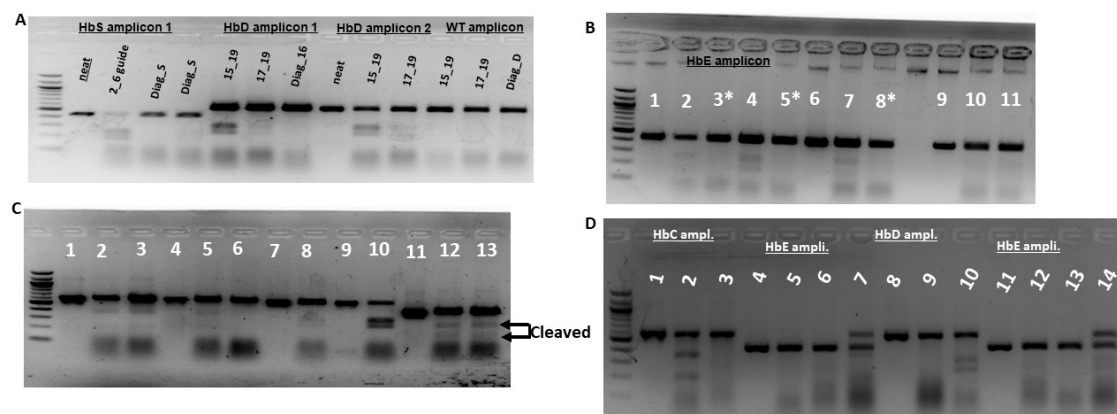
**Figure 2: gRNA for differential detection of Hb S, C and WT sequences.** A depicts a time-course IVC for Hb S amplicon (using Hb S gRNA) and WT amplicon (using WT-specific gRNA for the WT sequence of the Hb S region, and Diag\_gRNA able to detect heterozygous inheritance). B is IVC for the differential detection of homozygous S and heterozygous S

(AS). C is the optimisation of gRNA for the differential detection of homozygous Hb C inheritance. [15\_18, 16\_18, & 17\_18 identify positions relative to the NGG PAM where mismatches were introduced to enable differential detection of Hb C variant]. Note that the numbered positions in bold print indicate the position where the point mutation causing the respective haemoglobin variant occurs. In contrast, the numbered position in normal font indicates where a mismatch was introduced to enable differentiation from the wild-type sequence.

### **En31FnCas9 enables differential detection of Hb D and E variants in *in vitro* IVC.**

Haemoglobin D point mutation occurs in exon 3 of the beta globin chain ( $\beta^{121\text{GAA}} \rightarrow \text{CAA}$ ). When using PAM distal gRNA specific to the haemoglobin D variant with a mismatch introduced at the 15<sup>th</sup> or 17<sup>th</sup> position, we demonstrate that en31FnCas9 enabled differential detection of the Hb D variant (Figure 3A). Additionally, neither the 15\_19 gRNA nor the 17\_19 gRNA cleaved the wild-type sequence of the Hb D region. In Figure 3B, we demonstrate that both PAM-proximal and PAM-distal gRNA can enable the differential detection of haemoglobin variant E. The respective PAM proximal (HbE 2\_6 gRNA) and distal (HbE 16\_19 gRNA) guides did not show activity against the wild-type sequence of the HbE region, showing that these gRNAs can be employed to differentiate the homozygous state from a wild-type sequence. Since the haemoglobinopathy genes are recessively inherited, identification of both heterozygous and homozygous states is important for genetic counselling. In the present study, we demonstrate that when a single mismatch is introduced into the WT gRNA, in a position other than where the haemoglobin variant point mutation occurs, ensuring that the gRNA has two mismatches against the respective haemoglobin variant, en31FnCas9 can be used to enable differentiation of the WT sequences from the haemoglobin variants (Figure 3C). In leveraging the intolerance of en31FnCas9 to dual mismatches, we demonstrate that WT sequences of the haemoglobin variants C, D, and E can be differentially identified from their respective variants using diagnostic HbC\_16C, diagnostic HbD\_16C, and HbE\_17, respectively

(Figure 3C). Multiplexing functionalization represents an important step in achieving point-of-care test democratisation as it improves test throughput and assay affordability through the reduction of the cost per test. A crucial step in navigating the multiplexing potential of the CRISPR-Dx is ascertaining the non-interference of the functionalities and specificities of gRNAs. In this regard, we mimicked compound inherited haemoglobinopathies by combining haemoglobin variants in a single tube, and then explored the feasibility of combinatorial gRNAs to simultaneously detect these through in vitro cleavage assays (Figure 3D). Our IVC data demonstrate that potentially inherited compound haemoglobinopathies CE and DE can be reliably differentially detected using en31FnCas9 (Figure 3D).



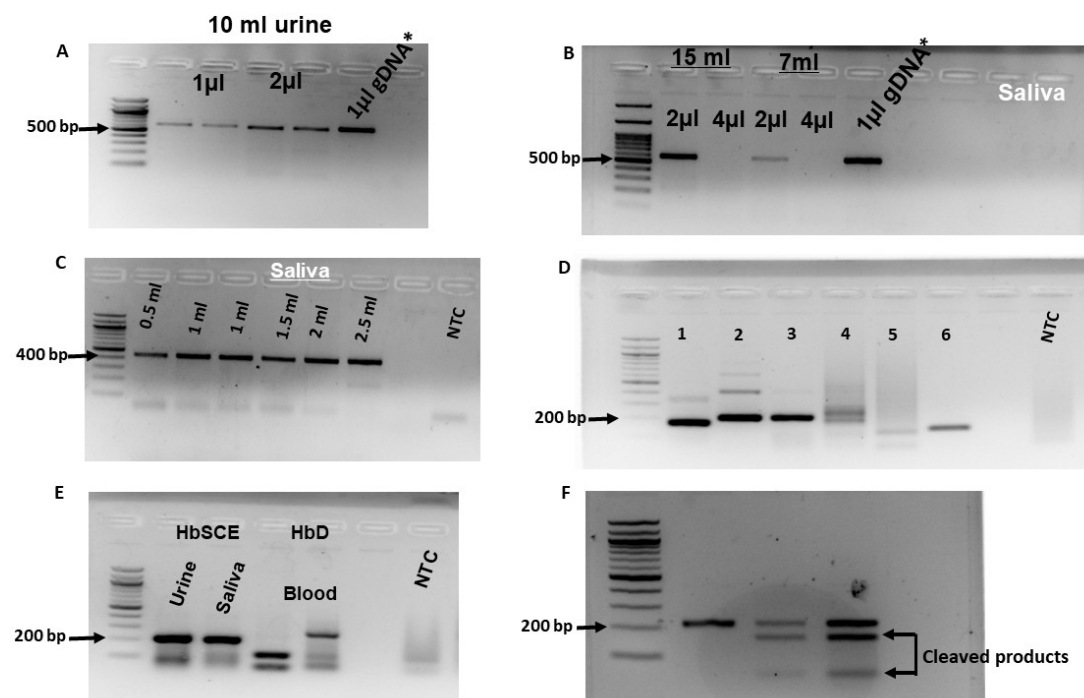
**Figure 3: gRNA for differential detection of Hb C, D, E and WT sequences.** A is the optimisation of gRNA for the differential detection of homozygous Hb D inheritance. [15\_19, & 17\_19 identify positions relative to the NGG PAM where mutations were introduced to enable differential detection of Hb D variant]. **Figure B** is differential detection of Hb E amplicon; [lanes 1, 6 & 9 are uncut; \*indicates that lane (3, 5 & 8) were cleaved with gRNA specific to WT sequence of the Hb E region; lanes 2 (gRNA\_2\_6), 4 (gRNA\_16\_19) and 7 (gRNA\_14\_16) were cleaved with Hb E specific gRNA; lane 9 is WT sequence of Hb E region uncut; lanes 10 and 11 are WT sequence of HbE region cleaved with Diag\_HbE\_16 and Diag\_HbE\_17 respectively. **Figure C** is the differential detections of Hb C, D and E (Lane 1: HbC amplicon uncut, Lane 2: HbC cleaved with HbCgRNA\_16\_18, Lane 3: HbC cleaved with HbCgRNA\_17\_18, Lane 4: WT sequence of HbC region uncut, Lane 5: WT sequence of HbC region cleaved with Diag\_HbC\_16C, Lane 6: WT sequence of HbC region cleaved with Diag\_HbC\_17C, Lane 7: HbD uncut, Lane 8: HbD cleaved with Diag\_HbD\_16C, Lane 9: WT sequence of HbD region uncut, Lane 10: WT sequence of HbD region cleaved with Diag\_HbD\_16C, Lane 11: HbE uncut, Lane 12: HbE cleaved with, HbE\_14\_16\_gRNA; Lane 13: WT sequence of HbE region cleaved with Diag\_HbE\_17. **Figure D** shows that the combinatorial use of gRNA with specificities for different

haemoglobin variants does not interfere with each other: Lane 1: HbC uncut, lane 2: HbC cleaved with HbC\_15\_18 gRNA, Lane 3: HbC cleaved with HbE\_14\_16 gRNA, lane 4: HbE uncut, lane 5: HbE cleaved with HbC\_15\_18 gRNA, lane 6: HbE cleaved with HbE\_14\_16 gRNA, Lane 7: Combined HbC & HbE variants cleaved with HbC\_15\_18 gRNA and HbE\_14\_16 gRNA, lane 8: HbD uncut, lane 9: HbD cleaved with HbE\_14\_16 gRNA, lane 10: HbD cleaved with HbD\_15\_19 gRNA, lane 11: HbE uncut, lane 12: HbE cleaved with HbE\_14\_16 gRNA, lane 13: HbE cleaved with HbD\_15\_19 gRNA, lane 14: combined HbD and E variants cleaved with HbD\_15\_19 gRNA and HbE\_14\_16 gRNA. [Note that the numbered positions in bold print indicate the position where the point mutation causing the respective haemoglobin variant occurs. In contrast, the numbered position in normal font indicates where a mismatch was introduced to enable differentiation from the wild-type sequence.]

### **RPA-based haemoglobin variant detection from less invasive urine and saliva samples**

Since there is no tool to predict the efficiency of RPA primers, we screened a series of primers for isothermal amplification of the genomic regions containing the haemoglobin S, C, E variants on one hand, and haemoglobin D on the other hand (See table 1). We initially optimised the NaOH gDNA extraction procedure [15] using 30 ml of urine donated by two individuals. As shown in Figure 4A, 1 – 2 µl of the NaOH-extracted gDNA was successfully amplified using the traditional PCR method. We were further able to successfully amplify gDNA extracted from 7 ml or 15 ml urine using the NaOH extraction method (Figure 4B). Subsequently, we were able to achieve reliable PCR amplification from as little as 500 µl saliva samples (Figure 4C). Among the many limitations of the traditional PCR-based assay for point-of-care diagnostics are the long processing times and the requirement for thermal cyclers, which are not routinely available in LMICs. Thus, we explored RPA assay for the detection of the haemoglobin variants. In Figure 4D and E, we demonstrate the feasibility of successfully implementing RPA to amplify NaOH-extracted gDNA from blood, urine and saliva. Furthermore, we demonstrate that 1 – 2 µl of the RPA product directly used in an *in vitro* cleavage assay enabled the detection of haemoglobin variant within 50 – 60 minutes (10 - 15, 20, and 20 minutes for NaOH extraction, RPA and 20 minutes for IVC, respectively). Taken together, our data demonstrate the potential of implementing a molecular-based isothermal haemoglobin variant detection assay, which is suitable for resource-limited settings.





**Figure 4: Optimisation of less invasive samples for Recombinase Polymerase Amplification assay.** A is a 10 ml of urine volume from two donors (amplified PCR products are 591 bp) used for NaOH-based gDNA extraction; B shows gDNA 7 and 15 ml saliva from two donors (amplified PCR products are 591 bp) used for NaOH-based gDNA extraction; C shows the titration of the minimal saliva volume that can be used for NaOH-based gDNA extraction for PCR amplification (amplified PCR products are 444 bp); D is optimisation of primers for RPA for the amplification of Hb SCE and Hb D regions (amplified RPA products are 190 bp, and 220 bp for HbSCE primers); E is RPA for Hb SCE region and HbD region using NaOH-based gDNA extracted from urine, saliva and blood (amplified PCR products are 220 bp for HbSCE region and 131 bp and 275 bp for HbD region); F is an in vitro cleavage assay undertaken on 1 µl or 2 µl volume of RPA amplified Hb C variant. (\*indicates 1 µl of Qiagen column purified gDNA from blood)

## Discussion

Clinically, haemoglobin variant diagnosis in LMICs has traditionally relied on phenotype-based assays such as cellulose acetate and/or acid electrophoresis, and solubility assays. While these assay platforms are economically feasible in resource-limited settings, their inherent failure to resolve co-migrating haemoglobin bands may inadvertently lead to underdiagnosis in some cases [7, 8]. The inherent sensitivity, specificity and scalability of CRISPR-based diagnostics offer unique opportunities for achieving democratisation of haemoglobinopathy diagnostics. Here, we leverage the intolerance of en31FnCas9 to dual mismatch in the PAM proximal or distal regions of the gRNA to demonstrate differential detection of haemoglobin variants S, C, D and E. Furthermore, we demonstrate that the inherent specificity of the en31FnCas9 can

also be leveraged to achieve differentiation of the homozygous and heterozygous states.

There are hundreds of genetically inherited point mutations that lead to haemoglobin variants and, in some cases, thalassaemia. Given that each of these point mutations is independently inherited, compound inherited mutations causing complex clinical presentations or symptomatology are likely scenarios in areas with high haemoglobinopathy burden. Thus, diagnostic platforms with multiplexing functionality have the potential to expand the diagnostic spectrum for these haemoglobinopathies in LMICs. Deploying such multiplex diagnostics in a POCT format in LMICs could potentially increase accessibility to neonatal screening and premarital screening programs to improve healthcare delivery as well as empower individuals to make informed choices to reduce the haemoglobinopathy burden. CRISPR diagnostics, as a molecular-based assay, has the potential to meet these key metrics given its nucleic acid-based specificity. Although our study focused on a limited number of haemoglobin variants of clinical significance, the differential *in vitro* detections achieved through the combination of gRNAs, even in heterozygous states, attest to the untapped potential of CRISPR-based diagnostics in multiplexed detection of compound haemoglobin variants. One of the key hindrances to the deployment of molecular-based assays in resource-poor settings has been the requirement for thermal cyclers to enable biomarker amplification to enhance signal detection. However, the discovery of isothermal amplification methods such as helicase-dependent amplification (HDA) [16], loop-mediated isothermal amplification (LAMP) [17], and recombinase polymerase amplification (RPA) [18] has not only enabled signal amplification without the need for thermal cyclers but also achieved sample-to-result in a relatively shorter time frame. These isothermal-based CRISPR-diagnostics have been previously harnessed for ultra-sensitive identification and diagnosis of various infectious agents [19-21]. In our present study, we demonstrate the compatibility of the haemoglobin variant detection assay with RPA-mediated signal amplification, strengthening the case for its potential point-of-care application. In the clinical settings, haemoglobin variant detection assays have traditionally required blood as a source material [7]; the expertise for phlebotomy limits the performance of these assays to hospital settings. In the present study, we have demonstrated the reliability of using non-invasive samples, such as urine and saliva,



as blood substitutes and thereby minimise the requirement for technical expertise when procuring samples. Thus, we anticipate that future protocol optimisation studies, and eventual coupling to a sensitive readout, will enable its deployment in primary healthcare ecosystems in resource-limited settings due to the convenience of saliva or urine sampling.

Except for Hb S, in which we had patient samples, our study was carried out using laboratory-generated site-directed mutagenesis-derived haemoglobin variants and may not account for the nuances associated with patients' chromosomal background. Thus, future field validation studies involving the use of samples from individuals with these inherited variants are expected to foster better assay characterisation and eventual POCT functionalisation. Notably, since Hb E mutation is associated with thalassaemia symptoms and clinical sequelae [22], our study demonstrates the potential for CRISPR-Dx as a diagnostic tool for point mutation-induced thalassaemia syndromes [23]. Thus, future functionalisation of our molecular-based CRISPR-Dx assay can complement existing NBS and population screening programs in LMICs. Previously, others have demonstrated the potential for cost-saving through the miniaturisation of CRISPR-diagnostics when deployed on a microfluidics platform to improve its sensitivity [24]. Santhiago and colleagues also demonstrated that the sensitivity of CRISPR-Dx was heavily dependent on enzyme kinetics and detector sensitivity [25]. Thus, we anticipate that further characterisation of our reported haemoglobin variant detection assay using various CRISPR-Cas orthologues and exploration of various detection platforms will lead to improved assay read-out options to facilitate the democratisation of haemoglobin variant diagnostics. Although mandatory newborn screening programs have improved patient care in developed economies, such programs have not been consistently implemented in LMICs due to cost implications. Thus, investments in research and development of molecular-based assays such as those described in the present study towards potential POCT functionalisation hold promise for achieving affordable and accessible universal newborn haemoglobinopathy screening.

## **Declarations**

## **Data availability**

The sequences of the oligonucleotides used in this study are enclosed in Table 1 of the manuscript. All the gel images on which the report is based are also enclosed in the manuscript.

## **Research ethics**

Our study was approved by the Human Ethics Committees of the CSIR – Institute of Genomics and Integrative Biology, New Delhi. The study was performed in accordance with the principles of the 1975 Helsinki Declaration as revised in 2024. All participants who donated urine, blood and/or saliva samples gave written informed consent before sampling; data were de-identified to preserve participants' anonymity.

## **Consent for publication**

All the authors and their respective affiliated institutions gave consent for publication.

## **Conflict of interest**

D.C. is listed as co-inventor for a US Patent titled, “Kinetically enhanced engineered FnCas9 and its uses thereof” which includes the engineered 31 FnCas9 variant used in this study. The patent applicant (application number is 18/049,291) is the Council of Scientific and Industrial Research, New Delhi, India, and has been granted. The authors have no other competing interests to declare.

## **Acknowledgements**

P.A. is supported by a CSIR-UNESCO-TWAS Postdoctoral Research Fellowship; P.K., PKD, PC and AC are supported by funding from the Council for Scientific and Industrial Research, India.

## **Contributions**

P.A. and D.C. conceived the project and designed the experimental protocol. P.C. and A.C. designed and performed the protein engineering experiments. P.A. and P.K. optimised the cloning, site-directed mutagenesis, and IVC assays. P.A., P.K. and P.K.D. designed and implemented bioinformatics tools for the design of gRNAs, validation, and in vitro cleavage assays. P.K. and P.A. were involved in the optimisation of RPA assays. P.A. wrote the draft manuscript. All authors critically reviewed the manuscript. All authors read and approved the final manuscript.

## References

1. Weatherall DJ: **Hemoglobinopathies worldwide: present and future.** *Current molecular medicine* 2008, **8**(7):592-599.
2. Kohne E: **Hemoglobinopathies: clinical manifestations, diagnosis, and treatment.** *Dtsch Arztebl Int* 2011, **108**(31-32):532-540.
3. Weatherall DJ, Clegg JB: **Inherited haemoglobin disorders: an increasing global health problem.** *Bulletin of the World Health Organization* 2001, **79**(8):704-712.
4. Sabath DE: **The role of molecular diagnostic testing for hemoglobinopathies and thalassemias.** *International Journal of Laboratory Hematology* 2023, **45**(S2):71-78.
5. Thom CS, Dickson CF, Gell DA, Weiss MJ: **Hemoglobin variants: biochemical properties and clinical correlates.** *Cold Spring Harb Perspect Med* 2013, **3**(3):a011858.
6. Cheesbrough M: **District laboratory practice in tropical countries:** Cambridge university press; 2006.
7. Barbara J. Bain IB, Michael A Laffan, S. Mitchell Lewis: **Dacie and Lewis Practical Haematology:** Elsevier Churchill Livingstone; 2011.
8. Adu P, Simpong N, Kontor K, Ephraim RK: **Misleading presentation of haemoglobin electrophoresis data.** *Ghana medical journal* 2017, **51**(1):36-38.
9. Hasan MN, Fraiwan A, An R, Alapan Y, Ung R, Akkus A, Xu JZ, Rezac AJ, Kocmich NJ, Creary MS *et al*: **Paper-based microchip electrophoresis for point-of-care hemoglobin testing.** *Analyst* 2020, **145**(7):2525-2542.
10. Mukherjee MB, Colah RB, Mehta PR, Shinde N, Jain D, Desai S, Dave K, Italia Y, Raicha B, Serrao E: **Multicenter Evaluation of HemoTypeSC as a Point-of-Care Sickle Cell Disease Rapid Diagnostic Test for Newborns and Adults Across India.** *Am J Clin Pathol* 2020, **153**(1):82-87.
11. Giardine B, Borg J, Higgs DR, Peterson KR, Philipsen S, Maglott D, Singleton BK, Anstee DJ, Basak AN, Clark B *et al*: **Systematic documentation and analysis of human genetic variation in hemoglobinopathies using the microattribution approach.** *Nat Genet* 2011, **43**(4):295-301.
12. Acharya S, Mishra A, Paul D, Ansari AH, Azhar M, Kumar M, Rauthan R, Sharma N, Aich M, Sinha D *et al*: **Francisella novicida Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing.** *Proceedings of the National Academy of Sciences* 2019, **116**(42):20959-20968.
13. Azhar M, Phutela R, Kumar M, Ansari AH, Rauthan R, Gulati S, Sharma N, Sinha D, Sharma S, Singh S *et al*: **Rapid and accurate nucleobase detection using FnCas9 and its application in COVID-19 diagnosis.** *Biosensors and Bioelectronics* 2021, **183**:113207.
14. Acharya S, Ansari AH, Kumar Das P, Hirano S, Aich M, Rauthan R, Mahato S, Maddileti S, Sarkar S, Kumar M *et al*: **PAM-flexible Engineered FnCas9 variants for robust and ultra-precise genome editing and diagnostics.** *Nature Communications* 2024, **15**(1):5471.
15. Rudbeck L, Dissing J: **Rapid, Simple Alkaline Extraction of Human Genomic DNA from Whole Blood, Buccal Epithelial Cells, Semen and Forensic Stains for PCR.** *BioTechniques* 1998, **25**(4):588-592.
16. Vincent M, Xu Y, Kong H: **Helicase-dependent isothermal DNA amplification.** *EMBO reports* 2004, **5**(8):795-800.
17. Tomita N, Mori Y, Kanda H, Notomi T: **Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products.** *Nature protocols* 2008, **3**(5):877-882.
18. Piepenburg O, Williams CH, Stemple DL, Armes NA: **DNA detection using recombination proteins.** *PLoS biology* 2006, **4**(7):e204.

19. Chen JS, Ma E, Harrington LB, Da Costa M, Tian X, Palefsky JM, Doudna JA: **CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity.** *Science* 2018, **360**(6387):436-439.
20. Li S-Y, Cheng Q-X, Wang J-M, Li X-Y, Zhang Z-L, Gao S, Cao R-B, Zhao G-P, Wang J: **CRISPR-Cas12a-assisted nucleic acid detection.** *Cell discovery* 2018, **4**(1):20.
21. Kellner MJ, Koob JG, Gootenberg JS, Abudayyeh OO, Zhang F: **SHERLOCK: nucleic acid detection with CRISPR nucleases.** *Nature Protocols* 2019, **14**(10):2986-3012.
22. Orkin SH, Kazazian HH, Antonarakis SE, Ostrer H, Goff SC, Sexton JP: **Abnormal RNA processing due to the exon mutation of  $\beta$ E-globin gene.** *Nature* 1982, **300**(5894):768-769.
23. Thein SL: **Molecular basis of  $\beta$  thalassemia and potential therapeutic targets.** *Blood Cells, Molecules, and Diseases* 2018, **70**:54-65.
24. Ackerman CM, Myhrvold C, Thakku SG, Freije CA, Metsky HC, Yang DK, Ye SH, Boehm CK, Kosoko-Thoroddsen T-SF, Kehe J *et al*: **Massively multiplexed nucleic acid detection with Cas13.** *Nature* 2020, **582**(7811):277-282.
25. Huyke DA, Ramachandran A, Bashkirov VI, Kotseroglou EK, Kotseroglou T, Santiago JG: **Enzyme Kinetics and Detector Sensitivity Determine Limits of Detection of Amplification-Free CRISPR-Cas12 and CRISPR-Cas13 Diagnostics.** *Analytical Chemistry* 2022, **94**(27):9826-9834.