# **Nucleic acid detection with CRISPR-Cas13a/C2c2**

Jonathan S. Gootenberg<sup>1,2,3,4,5\*</sup>, Omar O. Abudayyeh<sup>1,2,3,4,6,\*</sup>, Jeong Wook Lee<sup>7</sup>,

Patrick Essletzbichler<sup>1,2,3,4</sup>, Aaron J. Dy<sup>1,4,8</sup>, Julia Joung<sup>1,2,3,4</sup>, Vanessa Verdine<sup>1,2,3,4</sup>,

Nina Donghia<sup>7</sup>, Nichole M. Daringer<sup>8</sup>, Catherine A. Freije<sup>1,11</sup>, Cameron Myhrvold<sup>1,11</sup>,

Roby P. Bhattacharyya<sup>1</sup>, Jonathan Livny<sup>1</sup>, Aviv Regev<sup>1,9</sup>, Eugene V. Koonin<sup>10</sup>,

Deborah T. Hung<sup>1</sup>, Pardis C. Sabeti<sup>1,11,12,13</sup>, James J. Collins<sup>1,4,6,7,8,†</sup>, and Feng Zhang<sup>1,2,3,4†</sup>

- <sup>1</sup> Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
- <sup>2</sup> McGovern Institute for Brain Research at MIT, Cambridge, MA 02139, USA
- <sup>3</sup> Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
- <sup>4</sup> Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
- <sup>5</sup> Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
- <sup>6</sup> Department of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
- <sup>7</sup> Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA
- <sup>8</sup> Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
- <sup>9</sup> Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
- <sup>10</sup> National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA
- <sup>11</sup> Center for Systems Biology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
- <sup>12</sup> Department of Immunology and Infectious Disease, Harvard School of Public Health, Boston, MA 02115, USA
- <sup>13</sup> Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA.
- \* These authors contributed equally to this work.
- † Correspondence should be addressed to F. Z. (zhang@broadinstitute.org) and J. J. C. (jimjc@mit.edu).

## **ETHICS STATEMENT**

Human samples from Zika fever patients or individuals suspected of having Zika were obtained commercially from Boca Biolistics, and all protocols were approved by the Institutional Review Boards of Massachusetts Institute of Technology and Broad Institute of MIT and Harvard.

## **SUPPLEMENTARY MATERIALS AND METHODS**

## **Cloning of Cas13a loci and proteins for expression**

For the bacterial *in vivo* efficiency assay, Cas13a proteins from *Leptotrichia wadei* F0279 and *Leptotrichia shahii* were ordered as codon-optimized genes for mammalian expression (Genscript, Jiangsu, China) and cloned into pACYC184 backbones along with the corresponding direct repeats flanking either a beta-lactamase targeting or non-targeting spacer. Spacer expression was driven by a J23119 promoter.

For protein purification, mammalian codon-optimized Cas13a proteins were cloned into bacterial expression vector for protein purification (6x His/Twin Strep SUMO, a pET-based expression vector received as a gift from Ilya Finkelstein, University of Texas-Austin).

All plasmids used in this study are listed in Table S7.

## **Bacterial** *in vivo* **Cas13a efficiency assay**

LwCas13a and LshCas13a *in vivo* efficiency plasmids and a previously described beta-lactamase plasmid (*10*) were co-transformed into NovaBlue Singles competent cells (Millipore) at 90 ng and 25 ng, respectively. After transformation, dilutions of cells were plated on ampicillin and choramphicol LB-agar plate and incubated overnight at 37°C. Colonies were counted the next day.

## **LwCas13a protein purification**

Cas13a bacterial expression vectors were transformed into Rosetta™ 2(DE3)pLysS Singles Competent Cells (Millipore). A 16 mL starter culture was grown overnight in Terrific Broth 4 growth media (12 g/L tryptone, 24 g/L yeast extract, 9.4 g/L K<sub>2</sub>HPO, 2.2 g/L KH<sub>2</sub>PO<sub>4</sub>, Sigma) (TB), which was used to inoculate 4 L of TB for growth at 37°C and 300 RPM until an OD600 of 0.6. At this time, protein expression was induced by supplementation with IPTG (Sigma) to a final concentration of 500 uM, and cells were cooled to 18°C for 16 h for protein expression. Cells were then centrifuged at 5200 g for 15 min at 4°C. Cell pellet was harvested and stored at - 80°C for later purification.

All subsequent steps of the protein purification were performed at 4°C. Cell pellet was crushed and resuspended in lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 1 mM DTT, pH 8.0) supplemented with protease inhibitors (Complete Ultra EDTA-free tablets), lysozyme, and benzonase followed by sonication (Sonifier 450, Branson, Danbury, CT) with the following conditions: amplitude of 100 for 1 second on and 2 seconds off with a total sonication time of 10 min. Lysate was cleared by centrifugation for 1 hr at  $4^{\circ}$ C at 10,000 g and the supernatant was filtered through a Stericup  $0.22 \mu$ m filter (EMD Millipore). Filtered supernatant was applied to StrepTactin Sepharose (GE) and incubated with rotation for 1 hr followed by washing of the protein-bound StrepTactin resin three times in lysis buffer. The resin was resuspended in SUMO digest buffer (30 mM Tris-HCl, 500 mM NaCl 1 mM DTT, 0.15% Igepal (NP-40), pH 8.0) along with 250 Units of SUMO protease (ThermoFisher) and incubated overnight at 4<sup>o</sup>C with rotation. Digestion was confirmed by SDS-PAGE and Coomassie Blue staining and the protein eluate was isolated by spinning the resin down.

For further cation exchange and gel filtration purification, protein was loaded onto a 5 mL HiTrap SP HP cation exchange column (GE Healthcare Life Sciences) via FPLC (AKTA PURE, GE Healthcare Life Sciences) and eluted over a salt gradient from 130 mM to 2M NaCl in elution buffer (20 mM Tris-HCl, 1 mM DTT, 5% glycerol, pH 8.0). The resulting fractions were tested for presence of LwCas13a by SDS-PAGE, and fractions containing the protein were pooled and concentrated via a Centrifugal Filter Unit to 1 mL in S200 buffer (10 mM HEPES, 1 M NaCl, 5 mM  $MgCl<sub>2</sub>$ , 2 mM DTT, pH 7.0). The concentrated protein was loaded onto a gel filtration column (Superdex® 200 Increase 10/300 GL, GE Healthcare Life Sciences) via FPLC. The resulting fractions from gel filtration were analyzed by SDS-PAGE and fractions containing LwCas13a were pooled and buffer exchanged into Storage Buffer (600 mM NaCl, 50 mM Tris-HCl pH 7.5, 5% glycerol, 2mM DTT) and frozen at -80°C for storage.

## **Nucleic acid target and crRNA preparation**

Nucleic acid targets were PCR amplified with KAPA Hifi Hot Start (Kapa Biosystems), gel extracted, and purified using MinElute gel extraction kit (Qiagen). Purified dsDNA was incubated with T7 polymerase overnight at 30°C using the HiScribe T7 Quick High Yield RNA Synthesis kit (New England Biolabs) and RNA was purified with the MEGAclear Transcription Clean-up kit (Thermo Fisher)

For preparation of crRNAs, constructs were ordered as DNA (Integrated DNA Technologies) with an appended T7 promoter sequence. crRNA DNA was annealed to a short T7 primer (final concentrations 10 uM) and incubated with T7 polymerase overnight at 37°C using the HiScribe T7 Quick High Yield RNA Synthesis kit (New England Biolabs). crRNAs were purified using RNAXP clean beads (Beckman Coulter) at 2x ratio of beads to reaction volume, with an additional 1.8x supplementation of isopropanol (Sigma)

## **NASBA isothermal amplification**

NASBA was performed as previously described (*1*). For a 20  $\mu$ L total reaction volume, 6.7  $\mu$ L of reaction buffer (Life Sciences, NECB-24),  $3.3 \mu$ L of Nucleotide Mix (Life Sciences, NECN-24), 0.5  $\mu$ L of nuclease-free water, 0.4  $\mu$ L of 12.5  $\mu$ M NASBA primers, 0.1  $\mu$ L of RNase inhibitor (Roche, 03335402001) and 4  $\mu$ L of RNA input (or water for the negative control) were assembled at 4°C and incubated 65°C for 2 min and then 41°C for 10 min. 5  $\mu$ L of enzyme mix (Life Sciences, NEC-1-24) was added to each reaction, and the reaction mixture was incubated at  $41^{\circ}$ C for 2 hr.

## **Recombinase Polymerase Amplification**

Primers for RPA were designed using NCBI Primer blast (*27*) using default parameters, with the exception of amplicon size (between 100 and 140 nt), primer melting temperatures (between 54°C and 67°C), and primer size (between 30 and 35 nt). Primers were then ordered as DNA (Integrated DNA Technologies).

RPA and RT-RPA reactions run were as instructed with TwistAmp® Basic or TwistAmp® Basic RT (TwistDx), respectively, with the exception that 280 mM MgAc was added prior to the input template. Reactions were run with 1  $\mu$ L of input for 2 hr at 37°C, unless otherwise described.

## **LwCas13a collateral detection**

Detection assays were performed with 45 nM purified LwCas13a, 22.5 nM crRNA, 125 nM quenched fluorescent RNA reporter (RNAse Alert v2, Thermo Scientific),  $2 \mu L$  murine RNase inhibitor (New England Biolabs), 100 ng of background total human RNA (purified from HEK293FT culture), and varying amounts of input nucleic acid target, unless otherwise indicated, in nuclease assay buffer (40 mM Tris-HCl, 60 mM NaCl, 6 mM  $MgCl<sub>2</sub>$ , pH 7.3). If the input was amplified DNA including a T7 promoter from a RPA reaction, the above Cas13a reaction was modified to include 1 mM ATP, 1 mM GTP, 1 mM UTP, 1 mM CTP, and 0.6  $\mu$ L T7 polymerase mix (New England Biolabs). Reactions were allowed to proceed for 1-3 hr at 37°C (unless otherwise indicated) on a fluorescent plate reader (BioTek) with fluorescent kinetics measured every 5 min.

The single reaction combining RPA-DNA amplification, T7 polymerase conversion of DNA to RNA and Cas13a detection was performed by integrating the reaction conditions above with the RPA mix. Briefly, a 50  $\mu$ L single reaction assay consisted of 0.48  $\mu$ M forward primer, 0.48  $\mu$ M reverse primer, 1x RPA rehydration buffer, varying amounts of DNA input, 45 nM LwCas13a recombinant protein, 22.5 nM crRNA, 250 ng background total human RNA, 200 nM substrate reporter (RNase alert v2),  $4 \mu L$  murine RNase inhibitor (New England Biolabs),  $2 \text{ mM ATP}, 2$ mM GTP, 2 mM UTP, 2 mM CTP, 1  $\mu$ L T7 polymerase mix (New England Biolabs), 5 mM  $MgCl<sub>2</sub>$ , and 14 mM MgAc.

## **Digital droplet PCR quantification**

To confirm the concentration of ssDNA 1 and ssRNA 1 standard dilutions used in Figure 1C-D and for comparison of SHERLOCK sensitivity, we performed digital-droplet PCR (ddPCR). For DNA quantification, droplets were made using the ddPCR Supermix for Probes (no dUTP) (BioRad) with PrimeTime qPCR probes/primer assays (IDT) designed to target the ssDNA 1 sequence. For RNA quantification, droplets were made using the one-step RT-ddPCR kit for probes with PrimeTime qPCR probes/primer assays designed to target the ssRNA 1 sequence. Droplets were generated in either case using the QX200 droplet generator (BioRad) and transferred to a PCR plate. Droplet-based amplification was performed on a thermocycler as described in the kit protocol and nucleic acid concentrations were subsequently determined via measurement on a QX200 droplet reader.

## **Quantitative PCR (qPCR) analysis with TaqMan probes**

To compare SHERLOCK quantification with other established methods, we performed qPCR on a dilution series of ssDNA 1. A TaqMan probe and primer set (Table S6) were designed against ssDNA 1 and synthesized with IDT. Assays were performed using the TaqMan Fast Advanced Master Mix (Thermo Fisher) and measured on a Roche LightCycler 480.

## **Real-time RPA with SYBR Green II**

To compare SHERLOCK quantification with other established methods, we performed RPA on a dilution series of ssDNA 1. To quantitate accumulation of DNA in real-time, we added 1x SYBR Green II (Thermo Fisher) to the typical RPA reaction mixture described above, which provides a fluorescent signal that correlates with the amount of nucleic acid. Reactions were allowed to proceed for 1 hr at 37°C on a fluorescent plate reader (BioTek) with fluorescent kinetics measured every 5 min.

## **SHERLOCK freeze-drying and paper deposition**

Glass fiber filter paper (Whatman, 1827-021) was autoclaved for 90 min (Consolidated Stills and Sterilizers, MKII) and blocked in 5% nuclease-free BSA (EMD Millipore, 126609-10GM) overnight. After rinsing the paper once with nuclease-free water (Life technologies, AM9932), RNases were removed via incubation with  $4\%$  RNAsecure<sup>TM</sup> (Life technologies, AM7006) at 60°C for 20 min, and the paper was rinsed three more times with nuclease-free water to remove traces of RNAsecure. Treated papers were dried for 20 min at 80°C on a hot plate (Cole-Parmer, IKA C-Mag HS7) prior to use. 1.8  $\mu$ L of Cas13a reaction mixture as indicated earlier was put onto the disc (2 mm) that was placed in black, clear bottom 384-well plate (Corning, 3544). For the freeze-dried test of SHERLOCK, the plate containing reaction mixture discs was flash frozen in liquid nitrogen and was freeze-dried overnight as previously described (*1*). RPA samples were diluted 1:10 in nuclease-free water, and 1.8  $\mu$ L of the mixture was loaded onto the paper discs and incubated at 37°C using a plate reader (BioTek Neo).

## **Lentivirus Preparation and Processing**

Lentivirus preparation and processing was performed as previously described (1). Briefly, 10  $\mu$ g pSB700 derivatives that include a ZIKV or DENV RNA fragment,  $7.5 \mu$ g psPAX2, and  $2.5 \mu$ g pMD2.G were transfected into HEK293FT cells (Life Technologies, R7007) using the HeBS-CaCl<sub>2</sub> method. 28 hr after changing media to fresh DMEM supplemented with  $10\%$  FBS,  $1\%$ penicillin-streptomycin and 4 mM GlutaMAX (ThermoFisher Scientific), the supernatant was filtered using a 0.45  $\mu$ m syringe filter. ViralBind Lentivirus Purification Kit (Cell Biolabs, VPK-104) and Lenti-X Concentrator (Clontech, 631231) were used to purify and prepare lentiviruses from the supernatant. Viral concentration was quantified using QuickTiter Lentivirus Kit (Cell Biolabs, VPK-112). Viral samples were spiked into 7% human serum (Sigma, H4522), were heated to 95°C for 2 min and were used as input to RPA.

## **Isolation and cDNA purification of ZIKV human serum samples**

Suspected ZIKV positive human serum or urine samples were inactivated with AVL buffer (Qiagen) and isolation of RNA was achieved with QIAamp Viral RNA minikit (Qiagen). Isolated RNA was converted into cDNA by mixing random primers, dNTPs, and sample RNA followed by heat denaturation for 7 min at 70˚C. Denatured RNA was then reverse transcribed with Superscript III (Invitrogen) incubated at 22-25<sup>°</sup>C for 10 min, 50<sup>°</sup>C for 45 min, 55<sup>°</sup>C for 15 min, and 80˚C for 10 min. cDNA was then incubated for 20 min at 37˚C with RNAse H (New England Biolabs) to destroy RNA in the RNA:cDNA hybrids.

## **Bacterial genomic DNA extraction**

For experiments involving CRE detection, bacterial cultures were grown in lysogeny broth (LB) to mid-log phase, then pelleted and subjected to gDNA extraction and purification using the Qiagen DNeasy Blood and Tissue Kit, using the manufacturer's protocol for either Gram negative or Gram positive bacteria, as appropriate. gDNA was quantified by the Quant-It dsDNA (Thermo Scientific) assay on a Qubit fluorometer (Thermo Scientific) and its quality assessed via 200-300 nm absorbance spectrum on a Nanodrop spectrophotometer.

For experiments discriminating between *E. coli* and *P. aeruginosa,* bacterial cultures were grown to early stationary phase in Luria-Bertani (LB) broth. 1.0 mL of both *E. coli* and *P. aeruginosa* were processed using the portable PureLyse bacteria gDNA extraction kit (Claremont BioSolutions). 1X binding buffer was added to the bacterial culture before passing through the battery-powered lysis cartridge for three minutes. 0.5X binding buffer in water was used as a wash solution before eluting with 150  $\mu$ L of water.

## **Genomic DNA extraction from human saliva**

2 mL of saliva was collected from volunteers, who were restricted from consuming food or drink 30 min prior to collection. Samples were then processed using QIAamp® DNA Blood Mini Kit (Qiagen) as recommended by the kit protocol. For boiled saliva samples, 400  $\mu$ L of phosphate buffered saline (Sigma) was added to 100  $\mu$ L of volunteer saliva and centrifuged for 5 min at 1800 g. The supernatant was decanted and the pellet was resuspended in phosphate buffered saline with 0.2% Triton X-100 (Sigma) before incubation at 95°C for 5 min. 1  $\mu$ L of sample was used as direct input into RPA reactions.

#### **Synthetic standards for human genotyping**

To create standards for accurate calling of human sample genotypes, we designed primers around the SNP target to amplify ~200 bp regions from human genomic DNA representing each of the two homozygous genotypes. The heterozygous standard was then made by mixing the homozygous standards in a 1:1 ratio. These standards were then diluted to equivalent genome concentrations ( $\sim$ 0.56 fg/ $\mu$ L) and used as input for SHERLOCK alongside real human samples.

#### **Detection of tumor mutant cell free-DNA (cfDNA)**

Mock cfDNA standards simulating actual patient cfDNA samples were purchased from a commercial vendor (Horizon Discovery Group). These standards were provided as four allelic fractions (100% WT and 0.1%, 1%, and 5% mutant) for both the BRAF V600E and EGFR L858R mutants. 3  $\mu$ L of these standards were provided as input to SHERLOCK.

#### **Analysis of SHERLOCK fluorescence data**

To calculate background subtracted fluorescence data, the initial fluorescence of samples was subtracted to allow for comparisons between different conditions. Fluorescence for background conditions (either no input or no crRNA conditions) were subtracted from samples to generate background subtracted fluorescence.

crRNA ratios for SNP or strain discrimination were calculated to adjust for sample-to-sample overall variation as follows:

$$
crRNA A_i ratio = \frac{(m+n)A_i}{\sum_{i=1}^{m} A_i + \sum_{i=1}^{n} B_i}
$$

where  $A_i$  and  $B_i$  refer to the SHERLOCK intensity values for technical replicate i of the crRNAs sensing allele A or allele B, respectively, for a given individual. Since we typically have four technical replicates per crRNA,  $m$  and  $n$  are equal to 4 and the denominator is equivalent to the sum of all eight of the crRNA SHERLOCK intensity values for a given SNP locus and

individual. Because there are two crRNAs, the crRNA ratio average across each of the crRNAs for an individual will always sum to two. Therefore, in the ideal case of homozygosity, the mean crRNA ratio for the positive allele crRNA will be two and the mean crRNA ratio for the negative allele crRNA will be zero. In the ideal case of heterozygosity, the mean crRNA ratio for each of the two crRNAs will be one.

#### **SUPPLEMENTARY TEXT**

#### **Characterization of LwCas13a cleavage requirements.**

The protospacer flanking site (PFS) is a specific motif present near the target site that is required for robust ribonuclease activity by Cas13a. The PFS is located at the 3' end of the target site and was previously characterized for LshCas13a by our group as H (not G) (*10*). Although this motif is akin to a protospacer adjacent motif (PAM), a sequence restriction for DNAtargeting Class 2 systems, it is functionally different as it not involved in preventing selftargeting of CRISPR loci in endogenous systems. Future structural studies of Cas13a will likely elucidate the importance of the PFS for Cas13a:crRNA target complex formation and cleavage activity.

We purified the recombinant LwCas13a protein from *E. coli* (Fig. S1D-E) and assayed its ability to cleave a 173-nt ssRNA with each possible protospacer flanking site (PFS) nucleotide (A, U, C or G) (fig. S1F). Similar to LshCas13a, LwCas13a can robustly cleave a target with A, U, or C PFS, with less activity on the ssRNA with a G PFS. Although we see weaker activity against ssRNA 1 with a G PFS, we still see robust detection for the two target sites with G PFS motifs (Table S4; rs601338 crRNA and Zika targeting crRNA 2). It is likely that the H PFS is not required under every circumstance and that in many cases strong cleavage or collateral activity can be achieved with a G PFS.

## **Discussion of Recombinase Polymerase Amplification (RPA) and other isothermal amplification strategies.**

Recombinase polymerase amplification (RPA) is an isothermal amplification technique consisting of three essential enzymes: a recombinase, single-stranded DNA-binding proteins (SSBs), and a strand displacing polymerase. RPA overcomes many technical difficulties present in other amplification strategies, particularly polymerase chain reaction (PCR), by not requiring temperature regulation as the enzymes all operate at a constant temperature around 37°C. RPA replaces temperature cycling for global melting of the double-stranded template and primer annealing with an enzymatic approach inspired by *in vivo* DNA replication and repair. Recombinase-primer complexes scan double-stranded DNA and facilitate strand exchange at complementary sites. The strand exchange is stabilized by SSBs, allowing the primer to stay

bound. Spontaneous disassembly of the recombinase occurs in its ADP-bound state, allowing a strand-displacing polymerase to invade and extend the primer, allowing amplification without complex instrumentation unavailable in point-of-care and field settings. Cyclic repetition of this process in a temperate range of 37-42°C results in exponential DNA amplification. The original formulation published uses the Bacillus subtilis Pol I (Bsu) as the strand-displacing polymerase, T4 uvsX as the recombinase, and T4 gp32 as the single-stranded DNA binding protein (*18*), although it is unclear what components are in the current formulation sold by TwistDx used in this study.

Additionally, RPA has a number of limitations:

- 1) Although Cas13a detection is quantitative (fig. S2), real-time RPA quantitation can be difficult because of its rapid saturation when the recombinase uses all available ATP. While real-time PCR is quantitative because of the ability to cycle amplification, RPA has no mechanism to tightly control the rate of amplification. Certain adjustments can be made to reduce amplification speed, such as reducing available magnesium or primer concentrations, lowering the reaction temperature, or designing inefficient primers. Although we see some instances of quantitative SHERLOCK, such as in Fig. 2A-F, it is not always the case and depends on the template.
- 2) RPA efficiency can be sensitive to primer design. The manufacturer typically recommends designing longer primers to ensure efficient recombinase binding with average GC content (40-60%) and screening up to 100 primer pairs to find highly sensitive primer pairs. We have found with SHERLOCK that we only have to design two primer pairs to achieve an attomolar test with single molecule sensitivity. This robustness is likely due to the additional amplification of signal by constitutively active Cas13a collateral activity that offsets any inefficiencies in amplicon amplification. This quality is particularly important for our bacterial pathogen identification in Fig. 2G-H. We have experienced issues with amplifying highly structured regions such as the 16S rRNA gene sites in bacterial genomes because there is no melting step involved in RPA. Thus, secondary structure in primers becomes an issue, limiting amplification efficiency and thus sensitivity. We believe we were successful despite these RPA-specific issues

because of additional signal amplification from Cas13a.

3) The amplification sequence length must be short (100-200 bp) for efficient RPA. For most applications, this is not a significant issue and perhaps is even advantageous (e.g. cfDNA detection where average fragment size is 160 bp). Sometimes large amplicon lengths are important, such as when universal primers are desired for bacterial detection and the SNPs for discrimination are spread over a large area.

SHERLOCK's modularity allows any amplification technique, even non-isothermal approaches, to be used prior to T7 transcription and Cas13a detection. This modularity is enabled by the compatibility of the T7 and Cas13a steps in a single reaction allowing detection to be performed on any amplified DNA input that has a T7 promoter. Prior to using RPA, we tried nucleic acid sequence based amplification (NASBA) (*1, 17*) for our detection assay (fig. S3A). However NASBA failed to drastically improve the sensitivity of Cas13a (fig. S3B,C). Other amplification techniques that could be employed prior to detection include PCR, loop mediated isothermal amplification (LAMP) (*28*), strand displacement amplification (SDA) (*29*), helicase-dependent amplification (HDA) (*30*), and nicking enzyme amplification reaction (NEAR) (*31*). The ability to swap any isothermal technique allows SHERLOCK to overcome the specific limitations of any one amplification technique.

#### **Design of engineered mismatches.**

We previously showed that LshCas13a target cleavage was reduced when there were two or more mismatches in the target:crRNA duplex but was relatively unaffected by single mismatches, an observation we confirmed for LwCas13a collateral cleavage (fig. S8A). We hypothesized that by introducing an additional mutation in the crRNA spacer sequence, we would destabilize collateral cleavage against a target with an additional mismatch (two mismatches in total) while retaining on-target collateral cleavage, as there would only be a single mismatch. To test the possibility of engineering increased specificity, we designed multiple crRNAs targeting ssRNA 1 and included mismatches across the length of the crRNA (fig. S8A) to optimize on-target collateral cleavage and minimize collateral cleavage of a target that differs by a single mismatch. We observed that these mismatches did not reduce collateral cleavage of ssRNA 1, but significantly decreased signal for a target that included an additional mismatch (ssRNA 2). The designed crRNA that best distinguished between ssRNA 1 and 2 included synthetic mismatches close to the ssRNA 2 mismatch, in effect creating a "bubble," or distortion in the hybridized RNA. The loss of sensitivity caused by the coordination of a synthetic mismatch and an additional mismatch present in the target (i.e., a double mismatch) agrees with the sensitivity of LshCas13a and LwCas13a to consecutive or nearby double mismatches and presents a basis for rational design of crRNAs that enable single-nucleotide distinction (fig. S8B).

For mismatch detection of ZIKV and DENV strains, our full-length crRNA contained two mismatches (Fig 3A,B). Due to high sequence divergence between strains, we were unable to find a continuous stretch of 28 nt with only a single nucleotide difference between the two genomes. However, we predicted that shorter crRNAs would still be functional, and designed shorter 23nt crRNAs against targets in the two ZIKV strains that included a synthetic mismatch in the spacer sequence and only one mismatch in the target sequence. These crRNAs could still distinguish African and American strains of ZIKV (fig. S8C). Subsequent testing of 23 nt and 20 nt crRNA show that reductions of spacer length reduce activity but maintain or enhance the ability to discriminate single mismatches (fig. S9A-G)

To better understand how synthetic mismatches may be introduced to facilitate single-nucleotide mutation discrimination, we tiled the synthetic mismatch across the first seven positions of the spacer at three different spacer lengths: 28, 23, and 20 nt (fig. S9A). On a target with a mutation at the third position, LwCas13a shows maximal specificity when the synthetic mismatch is in position 5 of the spacer, with improved specificity at shorter spacer lengths, albeit with lower levels of on-target activity (fig. S9B-G). We also shifted the target mutation across positions 3-6 and tiled synthetic mismatches in the spacer around the mutation (fig. S10).

#### **Genotyping with SHERLOCK using synthetic standards.**

Evaluation of synthetic standards created from PCR amplification of the SNP loci allows for accurate identification of genotypes (fig. S12A,B). By computing all comparisons (ANOVA) between the SHERLOCK results of an individual's sample and the synthetic standards, each individual's genotype can be identified by finding the synthetic standard that has the most similar SHERLOCK detection intensity (fig. S12C,D). This SHERLOCK genotyping approach is generalizable to any SNP locus (fig. S12E).

#### **SHERLOCK is an affordable, adaptable CRISPR-Dx platform.**

For the cost analysis of SHERLOCK (Table S2), reagents determined to be of negligible cost were omitted, including DNA templates for the synthesis of crRNA, primers used in RPA, common buffers (MgCl<sub>2</sub>, Tris HCl, glycerol, NaCl, DTT), glass microfiber filter paper, and RNAsecure reagent. For DNA templates, ultramer synthesis from IDT provides material for 40 *in vitro* transcription reactions (each being enough for  $\sim$ 10,000 reactions) for  $\sim$ \$70, adding negligible cost to crRNA synthesis. For RPA primers, a 25 nmole IDT synthesis of a 30 nt DNA primer can be purchased for ~\$10, providing material adequate for 5000 SHERLOCK reactions. Glass microfiber paper is available for \$0.50/sheet, which is sufficient for several hundred SHERLOCK reactions. 4% RNAsecure reagent costs \$7.20/mL, which is sufficient for 500 tests.

In addition, for all experiments, except the paper-based assays, 384-well plates were used (Corning 3544), at the cost of \$0.036/reaction. Because of the negligible cost, this was not included in the overall cost analysis. Additionally, SHERLOCK-POC does not require the use of a plastic vessel, as it can easily be performed on paper.

The readout method for SHERLOCK used in this publication was a plate reader equipped with either a filter set or a monochromator. As a capital investment, the cost of the reader was not included in the calculation, as the cost precipitously decreases as more reactions are run on the instrument and is negligible.

For POC applications, cheaper and portable alternatives could be used, such as hand-held spectrophotometers (*32*) or portable electronic readers (*1*), which reduce the cost of instrumentation to <\$200. While these more portable solutions will reduce the speed and ease of readout as compared to bulkier instruments, they allow for more broad use.

## **SUPPLEMENTARY FIGURES**



#### **Figure S1. LwCas13a is capable of RNA-guided RNA interference and cleavage.**

- (A) Schematic of the CRISPR/Cas13a locus from *Leptotrichia wadei*. Representative crRNA structures from LwCas13a and LshCas13a systems are shown.
- (B) Schematic of *in vivo* bacterial assay for Cas13a activity. A protospacer is cloned upstream of the beta-lactamase gene in an ampicillin-resistance plasmid, and this

construct is transformed into *E. coli* expressing Cas13a in conjunction with either a targeting or non-targeting spacer. Successful transformants are counted to quantify activity.

- (C) Quantitation of LwCas13a and LshCas13a *in vivo* activity. (n=2 biological replicates; bars represent mean  $\pm$  s.e.m.)
- (D) Final size exclusion gel filtration of recombinant LwCas13a protein.
- (E) Coomassie blue stained acrylamide gel of LwCas13a stepwise purification.
- (F) Activity of LwCas13a against different PFS targets. LwCas13a was targeted against fluorescent RNA with variable 3' PFS flanking the spacer, and reaction products were visualized on denaturing gel. LwCas13a shows a slight preference against a G PFS.



## **Figure S2. Detection with LwCas13a is quantitative.**

(A) Fluorescence measurements from Cas13a detection without amplification are correlated with input RNA concentration. (n=2 biological replicates; bars represent mean  $\pm$  s.e.m.)



#### **Figure S3. Nucleic acid amplification with NASBA followed by Cas13a detection.**

- (A) Schematic of the NASBA reaction.
- (B) Detection of nucleic acid target ssRNA 1 amplified by NASBA with three different primer sets and then subjected to Cas13a collateral detection using a quenched fluorescent probe.
- (C) Comparison of detection of ssRNA 1 by NASBA with primer set 2 and SHERLOCK.  $(n=2$  technical replicates; bars represent mean  $\pm$  s.e.m.)



#### **Figure S4. Nucleic acid amplification with RPA and single-reaction SHERLOCK.**

- (A) Schematic of the RPA reaction, showing the participating components in the reaction.
- (B) Digital-droplet PCR quantitation of ssRNA 1 for dilutions used in Fig. 1C. Adjusted concentrations for the dilutions based on the ddPCR results are shown above bar graphs.
- (C) Digital-droplet PCR quantitation of ssDNA 1 for dilutions used in Fig. 1D. Adjusted concentrations for the dilutions based on the ddPCR results are shown above bar graphs.
- (D) The RPA, T7 transcription, and Cas13a detection reactions are compatible and achieve single molecule detection of DNA 2 when incubated simultaneously. (n=3 technical replicates, two-tailed Student t-test; n.s., not significant; \*\*,  $p < 0.01$ ; \*\*\*\*,  $p < 0.0001$ ; bars represent mean  $\pm$  s.e.m.)



#### **Figure S5. Comparison of SHERLOCK to other sensitive nucleic acid detection tools.**

- (A) Detection analysis of ssDNA 1 dilution series with digital-droplet PCR. (n=4 technical replicates, two-tailed Student t-test; n.s., not significant;  $*, p < 0.05; **$ ,  $p < 0.01; ***$ , p  $< 0.0001$ ; red lines represent mean, bars represent mean  $\pm$  s.e.m. Samples with measured copy/ $\mu$ L below 10<sup>-1</sup> not shown.)
- (B) Detection analysis of ssDNA 1 dilution series with quantitative PCR. (n=16 technical replicates, two-tailed Student t-test; n.s., not significant; \*\*,  $p < 0.01$ ; \*\*\*\*,  $p < 0.0001$ ; red lines represent mean, bars represent mean  $\pm$  s.e.m. Samples with relative signal below  $10^{-10}$  not shown.)
- (C) Detection analysis of ssDNA 1 dilution series with RPA with SYBR Green II. (n=4

technical replicates, two-tailed Student t-test;  $*, p < 0.05; **$ ,  $p < 0.01;$  red lines represent mean, bars represent mean  $\pm$  s.e.m. Samples with relative signal below 10<sup>0</sup> not shown.)

- (D) Detection analysis of ssDNA 1 dilution series with SHERLOCK. (n=4 technical replicates, two-tailed Student t-test; \*\*,  $p < 0.01$ ; \*\*\*\*,  $p < 0.0001$ ; red lines represent mean, bars represent mean  $\pm$  s.e.m. Samples with relative signal below 10<sup>0</sup> not shown.)
- (E) Percent coefficient of variation for a series of ssDNA 1 dilutions for four types of detection methods.
- (F) Mean percent coefficient of variation for the 6e2, 6e1, 6e0, and 6e-1 ssDNA 1 dilutions for four types of detection methods. (bars represent mean  $\pm$  s.e.m.)



# **Figure S6. Development of SHERLOCK as a point-of-care diagnostic. Figure S6**

(A) Freeze-dried Cas13a is capable of sensitive detection of ssRNA 1 in the low femtomolar range. (n=2 technical replicates; bars represent mean ± s.e.m.)

(B, C) Cas13a is capable of rapid detection of a 200 pM ssRNA 1 target on paper as spotted as liquid (B) or freeze-dried form (C).  $(n=3 \text{ technical replicates}; \text{bars represent mean } \pm \text{s.d.})$ 

(D, E) The SHERLOCK reaction is capable of sensitive detection of synthesized ZIKV RNA fragments in solution (D) and in freeze-dried form (E) (n=3 technical replicates; bars represent mean  $\pm$  s.e.m.)

(F) Quantitative curve for human ZIKV cDNA detection with SHERLOCK showing significant correlation between input concentration and detected fluorescence.

(G) Cas13a detection of ssRNA 1 performed in the presence of varying amounts of human serum.  $(n=2 \text{ technical replicates}; \text{bars represent mean } \pm \text{s.e.m.})$ 



#### **Figure S7. Detection of carbapanem resistance in clinical bacterial isolates.**

Detection of two different carbapenem-resistance genes (KPC and NDM-1) from five clinical isolates of *Klebsiella pneumoniae* and an *E. coli* control. (n=4 technical replicates, two-tailed Student t-test; \*\*\*\*, p < 0.0001; bars represent mean ± s.e.m.; n.d., not detected)



#### **Figure S8. Engineering Cas13a to have single-base specificity.**

- (A) Cas13a is not sensitive to single mismatches, but can distinguish between single nucleotide differences in target when loaded with crRNAs with additional mismatches. ssRNA 1-3 were detected with 11 crRNAs, with 10 spacers containing synthetic mismatches at various positions in the crRNA. Mismatched spacers did not show reduced collateral cleavage of ssRNA 1, but showed inhibited collateral cleavage of mismatched targets ssRNA 2 and ssRNA 3.
- (B) Schematic of the process for rational design of single-base specific spacers with synthetic mismatches. Synthetic mismatches are placed in proximity to the SNP or base of interest.
- (C) Highly specific detection of strain SNPs allows for the differentiation of ZIKV African versus American RNA targets differing by only one nucleotide using Cas13a detection with truncated (23 nt) crRNAs. (n=2 technical replicates, one-tailed Student t-test;  $*$ , p < 0.05; \*\*,  $p < 0.01$ ; \*\*\*\*,  $p < 0.0001$ ; bars represent mean  $\pm$  s.e.m.)



**Figure S9. Characterization of LwCas13a sensitivity to truncated spacers and single mismatches in the target sequence.** 

- (A) Sequences of truncated spacer crRNAs used in (B)-(G). Also shown are sequences of ssRNA 1 and 2, which has a single base-pair difference highlighted in red. crRNAs containing synthetic mismatches are displayed with mismatch positions colored in red.
- (B) Collateral cleavage activity on ssRNA 1 and 2 for 28 nt spacer crRNA with synthetic mismatches at positions 1-7. ( $n=4$  technical replicates; bars represent mean  $\pm$  s.e.m.)
	- (C) Specificity ratios of crRNA tested in (B). Specificity ratios are calculated as the ratio of

the on-target RNA (ssRNA 1) collateral cleavage to the off-target RNA (ssRNA 2) collateral cleavage.  $(n=4$  technical replicates; bars represent mean  $\pm$  s.e.m.)

- (D) Collateral cleavage activity on ssRNA 1 and 2 for 23 nt spacer crRNA with synthetic mismatches at positions 1-7. ( $n=4$  technical replicates; bars represent mean  $\pm$  s.e.m.)
- (E) Specificity ratios of crRNA tested in (D). Specificity ratios are calculated as the ratio of the on-target RNA (ssRNA 1) collateral cleavage to the off-target RNA (ssRNA 2) collateral cleavage.  $(n=4$  technical replicates; bars represent mean  $\pm$  s.e.m.)
- (F) Collateral cleavage activity on ssRNA 1 and 2 for 20 nt spacer crRNA with synthetic mismatches at positions 1-7. ( $n=4$  technical replicates; bars represent mean  $\pm$  s.e.m.)
- (G) Specificity ratios of crRNA tested in (F). Specificity ratios are calculated as the ratio of the on-target RNA (ssRNA 1) collateral cleavage to the off-target RNA (ssRNA 2) collateral cleavage.  $(n=4$  technical replicates; bars represent mean  $\pm$  s.e.m.)



## **Figure S10. Identification of ideal synthetic mismatch position relative to mutations in the target sequence.**

- (A) Sequences for evaluation of the ideal synthetic mismatch position to detect a mutation between ssRNA 1 and ssRNA 2. On each of the targets, crRNAs with synthetic mismatches at the colored (red) locations are tested. Each set of synthetic mismatch crRNAs is designed such that the mutation location is shifted in position relative to the sequence of the spacer. Spacers are designed such that the mutation is evaluated at positions 3, 4, 5, and 6 within the spacer. positions 3, 4, 5, and 6 within the spacer.<br> **(B)** Collateral cleavage activity on ssRNA 1 and 2 for crRNAs with synthetic mismatches at
	- varying positions. There are four sets of crRNAs with the mutation at either position 3, 4, 5, or 6 within the spacer:target duplex region. (n=4 technical replicates; bars represent mean  $\pm$  s.e.m.)

**A**

(C) Specificity ratios of crRNA tested in (B). Specificity ratios are calculated as the ratio of the on-target RNA (ssRNA 1) collateral cleavage to the off-target RNA (ssRNA 2) collateral cleavage. ( $n=4$  technical replicates; bars represent mean  $\pm$  s.e.m.)



## **Figure S11. Genotyping with SHERLOCK at an additional locus and direct genotyping from boiled saliva.**

- (A) SHERLOCK can distinguish between genotypes at the rs5082 SNP site. (n=4 technical replicates, one-tailed Student t-test;  $*, p < 0.05;$  \*\*\*,  $p < 0.001;$  \*\*\*\*,  $p < 0.0001;$  bars represent mean  $\pm$  s.e.m.)
- (B) SHERLOCK can distinguish between genotypes at the rs601338 SNP site in genomic DNA directly from centrifuged, denatured, and boiled saliva. (n=4 technical replicates, two-tailed Student t-test; \*\*,  $p < 0.01$ ; \*\*\*\*,  $p < 0.001$ ; bars represent mean  $\pm$  s.e.m.)



## **Figure S12. Development of synthetic genotyping standards to accurately genotype human SNPs.**

- (A) Genotyping with SHERLOCK at the rs601338 SNP site for each of the four individuals compared against PCR-amplified genotype standards. (n=4 technical replicates; bars represent mean  $\pm$  s.e.m.)
- (B) Genotyping with SHERLOCK at the rs4363657 SNP site for each of the four individuals compared against PCR-amplified genotype standards. (n=4 technical replicates; bars represent mean  $\pm$  s.e.m.)
- (C) Heatmaps of computed p-values between the SHERLOCK results for each individual and the synthetic standards at the rs601338 SNP site. A heatmap is shown for each of the allele-sensing crRNAs. The heatmap color map is scaled such that insignificance ( $p >$ 0.05) is red and significance ( $p < 0.05$ ) is blue. (n=4 technical replicates, one-way ANOVA)
- (D) Heatmaps of computed p-values between the SHERLOCK results for each individual and the synthetic standards at the rs4363657 SNP site. A heatmap is shown for each of the allele-sensing crRNAs. The heatmap color map is scaled such that insignificance ( $p >$

0.05) is red and significance ( $p < 0.05$ ) is blue. (n=4 technical replicates, one-way ANOVA)

(E) A guide for understanding the p-value heatmap results of SHERLOCK genotyping. Genotyping can easily be called by choosing the allele that corresponds to a p-value > 0.05 between the individual and allelic synthetic standards. Red blocks correspond to non-significant differences between the synthetic standard and individual's SHERLOCK result and thus a genotype-positive result. Blue blocks correspond to significant differences between the synthetic standard and individual's SHERLOCK result and thus a genotype-negative result.



#### **Figure S13. Detection of ssDNA 1 as a small fraction of mismatched background target.**

- (A)SHERLOCK detection of a dilution series of ssDNA 1 on a background of human genomic DNA. Note that there should be no sequence similarity between the ssDNA 1 target being detected and the background genomic DNA. (n=2 technical replicates; bars represent mean  $\pm$  s.e.m.)
- (B) Schematic of SHERLOCK detection of ssDNA 1 (mutant DNA) on a background of ssDNA 2, which differs from ssDNA 1 by only a single mismatch (wild-type background DNA).
- (C) SHERLOCK achieves single nucleotide specificity detection of ssDNA 1 in the presence of ssDNA 2, which differs by only a single mismatch. Various concentrations of ssDNA 1 were combined with a background excess of ssDNA 2 and detected by SHERLOCK.

## **SUPPLEMENTARY TABLES**



## **Supplementary Table 1: SNP variants tested with SHERLOCK genotyping.**

## **Supplementary Table 2: SHERLOCK cost analysis\*.**



\*Cost analysis of the SHERLOCK detection on paper. Fraction used is determined by calculating how much of each kit or procedure is ultimately used in the  $2 \mu L$  SHERLOCK-POC reaction. Some reagents, such as the buffer or DNA templates for synthesizing the crRNA, are omitted from the calculation because their contribution to the overall cost is negligible (See Supplementary Note 5).

#### **Supplementary Table 3: RPA/NASBA primers used in this study.**







## **Supplementary Table 4: crRNA sequences used in this study.**











# **Supplementary Table 5: RNA and DNA targets used in this study.**







## **Supplementary Table 6: Table of qPCR primer/probe sequences.**

## **Supplementary Table 7: Plasmids used in this study**

