CRISPR-Cas-amplified urine biomarkers for multiplexed and portable cancer diagnostics

Liangliang Hao
Massachusetts Institute of Technology
Renee Zhao
Massachusetts Institute of Technology
Chayanon Ngambenjawong
Massachusetts Institute of Technology https://orcid.org/0000-0002-2342-7977
Henry Ko
Massachusetts Institute of Technology https://orcid.org/0000-0003-3665-5407
Heather Fleming
Massachusetts Institute of Technology
Sangeeta Bhatia (sbhatia@mit.edu)
Massachusetts Institute of Technology https://orcid.org/0000-0002-1293-2097

Article

Keywords: Synthetic Biomarkers, Chemically-stabilized DNAs, Nucleic Acid Barcodes, Unprocessed Biofluids, Murine Cancer Models, Point-of-care Tool

Posted Date: July 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-636557/v1

Version of Record: A version of this preprint was published at Nature Nanotechnology on April 24th, 2023. See the published version at https://doi.org/10.1038/s41565-023-01372-9.

Abstract

Synthetic biomarkers, exogenous probes that generate molecular reporters, represent an emerging paradigm in precision diagnostics with applications across infectious and noncommunicable diseases. These methods use multiplexing strategies to provide tools that are both sensitive and specific. However, the field of synthetic biomarkers has not benefited from molecular strategies such as DNA-barcoding due to the susceptibility of nucleic acids *in vivo.* Herein, we exploit chemically-stabilized DNAs to tag synthetic biomarkers and produce diagnostic signals *via* CRISPR nucleases. Our strategy capitalizes on disease-associated, protease-activated release of nucleic acid barcodes and polymerase-amplification-free, CRISPR-Cas-mediated barcode detection in unprocessed biofluids. In murine cancer models, we show that these DNA-encoded nanosensors can noninvasively detect and monitor disease progression, and demonstrate that nuclease amplification can be harnessed to convert the readout to a point-of-care tool. This technique combines specificity with ease of use to offer a new platform to study human disease and guide therapeutic decisions.

Main

Precision diagnostics that reflect the diversity of disease signals are essential for improving therapeutic outcomes. Whereas disease-derived molecules such as proteins or nucleic acids are often limited by either abundance or stability in circulation, rationally-engineered 'synthetic biomarkers' leverage the catalytic, functional, and/or microenvironmental nature of enzymes to achieve advantages in sensitivity and specificity relative to endogenous biomarkers[1-6]. As the development of synthetic biomarkers continues to push molecular diagnostics forward, further progress will necessitate tools capable of achieving highly multiplexed readouts to classify diverse, dynamic disease states.

Molecular strategies for multiplexing abound in analytical chemistry, with DNA-barcoding proving to be transformative in drug screening, gene expression profiling, and cellular function analysis[7-9]. However, the field of synthetic biomarkers has not yet benefited from DNA multiplexing due to the susceptibility of nucleic acids *in vivo*[10]. Even measurements of endogenous, cell-free nucleic acids are hampered by the degradation of sequences that are not protected by the nucleosome in circulation[11-15]. Chemically-modified nucleic acids which have been primarily developed as therapeutics are less resistant to systemic degradation; however, they have not been incorporated into diagnostic platforms because they cannot be amplified with conventional polymerases, nor inexpensively sequenced[16-18].

To multiplex synthetic biomarkers for use as Point-of-Care (PoC) precision diagnostics, we exploited chemically-stabilized DNA for molecular barcoding and employed CRISPR-Cas nucleases that are insensitive to chemical modifications to read DNA barcodes in a sequence-specific manner [19]. By integrating these enabling technologies, we engineered a panel of high-throughput, programmable *in vivo* nanosensors for noninvasive disease detection and monitoring in urine (Fig. 1). These rationally-designed, DNA-encoded synthetic urine biomarkers (DNA-SUBs) 1) target sites of disease after *in vivo* administration, 2) sense dysregulated proteolysis in the tumor microenvironment[20,21], and 3) emit nucleic

acid barcodes that survive circulation and are size-concentrated by the kidney to 4) activate single-stranded DNase activity of CRISPR-Cas12a *via* sequence complementarity to a CRISPR guide RNA. Shed by pathological proteolytic activities, chemically-stabilized DNA barcodes were readily scanned in unprocessed urine using fluorescence kinetics or on paper strips upon Cas12a activation. To demonstrate the modularity of this approach, we assembled DNA-SUBs on both biologic (nanobody) or synthetic (polymer) scaffolds to demonstrate their use in noninvasive monitoring of two types of invasive cancer.

Chemically-modified DNA enables CRISPR-Cas12a-mediated urinary readout for *in vivo* sensing

CRISPR-Cas nucleases show great promise in DNA and RNA diagnostic applications[19,22-27], though they have yet to be applied to *in vivo* disease sensing due to the unspecific degradation of unmodified nucleic acids in circulation. To prevent the susceptibility and immunostimulatory activity of unmodified nucleic acids *in vivo*, we sought to investigate the tolerability of Cas proteins to chemically-stabilized DNA molecules. We introduced native and FDA-approved phosphorothioate-modified DNA molecules to the enzyme CRISPR-Cas12a[28-30]. Cas12a exhibits non-specific endonuclease activity after binding to a specific target via programmable CRISPR RNAs (crRNAs), leading to a subsequent trans-cleavage of bystander DNA molecules that can be monitored using a fluorophore (F)- quencher (Q)- labeled reporter[19] (Fig. 2a). Specifically, the Cas12a from *Lachnospiraceae bacterium* ND2006 (*Lba*Cas12a) recognizes double-stranded DNA (dsDNA) target through a T nucleotide-rich protospacer-adjacent motif (PAM) and single-stranded DNA (ssDNA) target in a PAM-independent manner. In addition to native dsDNA or ssDNA, *Lba*Cas12a was activated by phosphorothioated ssDNA at a relatively slower speed (Fig. 2b). When intravenously administered *in vivo*, native ssDNA in urine collected from injected animals could not activate *Lba*Cas12a assembled with the corresponding crRNA due to the unspecific DNase activities in circulation (Fig. 2c, Supplementary Fig. 1a). In contrast, different lengths of phosphorothioate-modified ssDNAs in solution or unprocessed urine from injected animals triggered the trans-cleavage activity of *Lba*Cas12a (Fig. 2c). Notably, the 20-mer crRNA-complementary ssDNA optimized kidney filtration into urine, producing the highest reporter cleavage activity, whereas the 24-mer ssDNA containing the PAM sequence produced the highest cleavage signal *in vitro* (Fig. 2c, Supplementary Table 1 & 2). Furthermore, we validated multiple crRNA-modified ssDNA activator pairs with high specificity between different sequences, allowing for parallel readout in multiple well assays (Fig. 2d, Supplementary Fig. 1b-h). The *Lba*Cas12a was activated once it encountered its programmed DNA target in unprocessed urine and cleaved a FAM and biotin dual-tagged ssDNA reporter that rapidly appeared on a lateral flow paper strip (Fig. 2e). The presence of the ‘sample band’ at the front of the strip indicated that freed FAM existed after reporter cleavage upon the activation of *Lba*Cas12a by the urinary DNA activator. In addition to the binary DNA activator detection, quantification of the intensity of the sample bands on paper strips allowed assessment of enzymatic kinetics (Fig. 2f). By adjusting the concentrations of assay components, the working linear ranges fell within sub-nanomolar DNA activator concentrations for both fluorescent (Supplementary Fig. 2) and paper-based readouts (Supplementary Fig. 3).

Tumor targeting, singleplex DNA-encoded synthetic urine biomarker specifically detects tumor-bearing animals

To develop DNA-encoded synthetic biomarker platforms, we leveraged deregulated proteolytic activities in the disease microenvironment to cleave and release the phosphorothioated DNA barcodes that were size-specifically concentrated in the urine to produce a noninvasive readout of the target disease. We first evaluated a singleplex synthetic urine biomarker *in vivo* in a human prostate cancer (PCa) xenograft model[31]. To maximize the on-target protease cleavage, we engineered the DNA-SUB on a biological scaffold that enables tumor-targeting abilities. To utilize the robust stability and tissue affinity of single domain antibody fragments (nanobodies), we constructed DNA-encoded, protease-activatable nanobodies by inserting a peptide substrate sequence with an unpaired cysteine for one-step site-specific labeling of cargos via a thio-ether bond[32-35] (Fig. 3a, Supplementary Table 3, Supplementary Fig. 4a, b). The peptide substrate specifically responded to the PCa-associated protease PLAU[31] (Supplementary Fig.7e). To prevent possible misfolding caused by the internal disulfide bond present in the nanobody, the peptide substrate with cysteine was spaced from the nanobody scaffold by a rigid linker (Supplementary Table 3). For *in vivo* validation, we tested the PLAU-activated, cMET-targeting nanobody in the cMET- and PLAU-expressing PC-3 cell-derived tumor model. In the subcutaneous PC-3 tumors, the cMET nanobody mediated active tumor trafficking upon systemic administration (Fig. 3b, Supplementary Fig. 4c, d). PLAU-activated nanobodies covalently conjugated with the 20-mer DNA barcode were efficiently separated via size-exclusion chromatography. The DNA-barcoded, PLAU-activated cMET nanobody (cMET-Nb-DNA) exhibited enhanced tumor accumulation compared with the DNA-barcoded, PLAU-activated non-targeting GFP nanobody (GFP-Nb-DNA)[36] (Fig. 3c). We systemically administered cMET-Nb-DNA to tumor-bearing and healthy control mice and quantified urinary DNA barcodes that were freed from the nanobody scaffold 1 h after injection. Urine samples were incubated with LbaCas12a-coupled with the complementary crRNA, and the trans-cleavage activity triggered by the DNA barcode was analyzed by tracking the kinetics of cleavage of a fluorescence-quencher labeled poly(T) reporter. Administration of cMET-Nb-DNA significantly increased the trans-cleavage rate of LbaCas12a activated by urine samples collected from tumor-bearing mice, relative to that of the healthy controls (Fig. 3d).

To translate the fluorescent readout into a PoC detection tool, we incubated LbaCas12a activated by mouse urine samples with the FAM-poly(T)-biotin reporter and ran the cleavage products on lateral flow paper strips. An enhanced sample band appeared in samples collected from tumor-bearing mice injected with cMET-Nb-DNA (Fig. 3e). The high sensitivity and specificity of the sensor to track disease was reflected in a Receiver operating characteristic (ROC) curve analysis (area under the curve (AUC)= 0.89 ± 0.10, P = 0.03). In contrast, urine samples collected from tumor-bearing mice injected with GFP-Nb-DNA activated *Lba*Cas12a at a rate that was almost identical to the samples from healthy controls, indicating that the release of the DNA barcodes was triggered on-target (Fig. 3d, e, Supplementary Fig. 4e). This urinary diagnostic exhibited robustness in detecting PC-3 cell-derived tumors that are prostate-specific antigen (PSA) negative (Supplementary Fig. 4f), suggesting a broad application of the synthetic urine biomarkers in tumor types without available clinical blood biomarkers.

DNA-encoded multiplex synthetic urine biomarkers longitudinally monitor disease progression in a portable manner

It is increasingly appreciated that analysis of multiple cancer hallmarks may optimize diagnostic sensitivity and specificity in heterogenous diseases. Whereas active targeting is limited to diseases that express specific ligands, multiplexing of an untargeted scaffold has the potential to be more generalizable. Therefore, we constructed a multiplexed panel of DNA-SUBs on a polymer-based scaffold and administered them as a single pool to mice (Fig. 4a). Each DNA-SUB was comprised of a 20-mer phosphorothioated DNA-tagged, protease-activated peptide (PAP) covalently conjugated to 40 kDa eight-arm poly(ethylene glycol) (PEG) nanocarriers which prolonged the blood half-life of cargos (Fig. 4a, Supplementary Fig. 5 & Table 3). Upon purification, the completed nanosensors showed increased hydrodynamic diameter, monodispersity, and uniform spherical shape (Supplementary Fig. 5 a-d).

To monitor nanosensor trafficking to tissue contact with tumor-immune compositions, we established a syngeneic mouse model by intravenously injecting a metastatic murine cell line (MC26-LucF) of colorectal cancer (CRC) into immunocompetent Balb/c mice[37] (Fig. 4a, Supplementary Fig. 6).We first identified a panel of CRC-specific proteases through transcriptomic analysis and found multiple proteases expressed in tumors at >1.5-fold levels over normal samples, including members of the matrix metalloproteinase (MMPs), aspartic, and serine protease families (i.e. cathepsins, kallikrein-related peptidases) (Supplementary Fig. 7a). From a matrisome proteomic analysis, we confirmed proteases present in primary CRCs and their distant metastases (e.g. MMP-7, -9, Cathepsin D, PLAU)[38,39] (Supplementary Fig. 7c). We confirmed that these identified proteases were overexpressed in tumor-bearing lung tissue of the MC26 transplantation model compared to normal lung tissue (Supplementary Fig. 7b, d). To identify peptide substrates specific to the selected proteases, we screened 16 peptide sequences against purified recombinant proteases and identified the top five substrates using a fluorogenic activity assay[31,40] (Supplementary Fig. 7e). These protease-activated peptides (PAP7, PAP9, PAP11, PAP13, PAP15) broadly cover metallo, serine, and aspartic protease activities (Supplementary Fig. 7e), and were specifically cleaved by tumor tissue homogenates *ex vivo* with highly predicted disease classification power, and thus were incorporated into the panel of DNA-SUBs for *in vivo* validation (Fig. 4b, Supplementary Fig. 8, 9a). We first showed that a DNA-barcoded, MMP-responsive SUB (DNA-PAP7-SUB) accumulated in the CRC lung tumor nodules following intravenous injection (Fig. 4c, Supplementary Fig. 9b). We then tested the entire 5-plex of DNA-barcoded SUBs *in vivo*, with an emphasis on identifying reporters that differentiated mice bearing lung tumor nodules from the healthy control animals. We systemically administered the multiplexed DNA-SUBs to the two mouse cohorts over the course of tumor development, and quantified urinary DNA barcodes that were freed from the nanosensors one hour after injection. Urine samples were incubated with *Lba*Cas12a-coupled with five different complementary crRNAs in multiple wells, and the trans-cleavage activity triggered by each DNA barcode was analyzed by tracking the cleavage kinetics of a fluorescence-quencher labeled reporter. We found that the MMP-responsive sensor (DNA-PAP7-SUB) from this multiplexed panel succeeded in distinguishing tumor-bearing mice from healthy mice only 11 days after tumor inoculation when the tumor nodules were 1-2 mm (Supplementary Fig. 6a). Some sensor (DNA-PAP9-SUB, DNA-PAP15-SUB) differences were amplified over time (Fig. 4d, Supplementary Fig. 9c), and these sensors (PAP9 and PAP15) corresponded to peptides cleaved by serine and metalloproteases *in vitro,* and also produced distinct cleavage patterns

when incubated with homogenates from either tumor-bearing or healthy lung tissues (Fig. 4b). Based on the ROC curve analysis, the sum of the metallo- (PAP7, PAP15) and serine (PAP9) protease substrate signals significantly increased the classification power of the DNA-SUBs (Fig. 4f). We then combined the 5-plex sensor panel with lateral flow detection for a visual readout that could enable PoC diagnostics. Using the same urine samples assayed in the aforementioned Cas12a kinetic cleavage reactions (Fig. 4d), we designed the lateral flow assay to read the cleavage of the FAM-poly(T)-biotin reporter at the optimized end timepoint. After the activation of *Lba*Cas12a by incubating the enzyme with a specific crRNA and its complementary DNA barcodes in urine, the enzyme complex and FAM-poly(T)-biotin reporter were mixed and added onto an assigned location in a 96-well plate. A series of lateral flow strips were loaded onto the plates and the multiple-pot test paper results appeared in 5 min at room temperature (Fig. 4e). Consistent with the results in the fluorescent readout, the test paper 'fingerprints' revealed distinctions in the intensity of sample bands resulting from Cas12a activation of tumor-bearing mice and healthy mice (Fig. 4e, Supplementary Fig. 9d). Notably, quantification of the sample band intensities exhibited disease classification power with multiple sensors (Fig. 4f), enabling a platform that is amenable to clinical translation due to its well-understood chemical composition and use of DNA multiplexing to overcome relatively low tumor accumulation, relative to ligand-targeted scaffolds.

Discussion

In this work, we present an integrated synthetic biomarker platform that enables *in vivo* sensing of endogenous proteolytic activities in the disease microenvironment and releases DNA fingerprints that are detectable in the urine. These size-specifically concentrated (< 10 kDa), chemically-stabilized DNA molecules triggered a second signal amplification step through activation of a CRISPR-Cas nuclease in a PAM-independent manner. The pathologically-released synthetic urinary DNAs fulfill several required characteristics of an ideal biomarker: high signal-to-noise ratio, stability in circulation until detection, easy accessibility from host biofluids, and high sensitivity and specificity in disease discrimination.

In the synthetic urinary biomarker platform, release of DNA barcodes is triggered by proteases known to respond during tumorigenesis. We identified proteases that were differentially expressed in two types of invasive cancer and selected peptide substrates with broad protease susceptibility. To increase the specificity of *in vivo* sensing, we genetically encoded protease-activated sequences fused to a nanobody to selectively bind an overexpressed growth factor receptor in prostate cancer. In addition to the diagnostic demonstration in a flank xenograft model, we showed disease classification by multiplexing DNA-SUBs in a syngeneic model to embrace proteolytic contributions from immune cells. We observed the expected preferential cleavage of metalloprotease-responsive substrates in CRC-bearing mice and the activation of a serine protease-sensitive substrate (PAP9) that was reported in the lung[40]. This pattern could point to an imbalance of serine protease/protease inhibitors in the pulmonary tumor microenvironment[41-43]. Due to the differences in proteolytic networks present in mouse models and human diseases, additional work to be completed before clinical translation may include an evaluation of

the sensor performance in the context of additional animal models with comorbidities, and also in human samples.

Urinary detection of cancers in mice *via* multiplexed synthetic biomarkers showed improved sensitivity relative to existing and emerging blood-based diagnostics[1,40]. In the LS174T CRC xenograft model, mass-encoded synthetic biomarkers detect the tumor that is 130 mm^3 in size, whereas circulating tumor DNA (ctDNA) is detectable when tumor sizes are >1000 mm^3 and carcinoembryonic antigen can be detected when tumors reach 135 to 330 mm^3 in this model[1,6,44]. Despite the advantages of multiplexing with mass encoding, one challenge of mass spectrometry is that it requires rigorous assessment of instrumentation and data interpretation. Here, through DNA-barcoding, our platform expanded the application of cost-effective CRISPR-based detection from viral or endogenous nucleic acids to synthetic biomarkers[19,22,25,27,45]. This study represents a step towards cancer diagnostics through a lateral flow readout of the synthetic DNA in urine, allowing for rapid, portable detection on paper strips similar to a home pregnancy test. Owing to the inherent programmability of Cas proteins, Cas12a exhibited collateral activity upon multiple 20-mer barcodes tested, which can be extensible to hundreds of orthogonal codes, providing a level of multiplexing that is challenging to attain with commercial isobaric tags for mass spectrometry[1]. We thus envision that the DNA-encoded synthetic biomarkers may be used to sense and report the entire activity landscape of proteases in order to differentiate between health and disease, in combination with spatial transcriptomic and multiplexed immunohistological techniques. To simultaneously monitor the activity of multiple proteases, in the current study we read their urinary signatures using an array of individual paper strips. Future PoC diagnostic systems can be designed to incorporate multiplexed sensing channels or surface-bound spots, and will benefit from computational tools that can be used to quantify more complex sensing results and improve accuracy of diagnoses[46].

Despite the lower melting temperature of duplexes, the phosphorothioate internucleotide linkages endowed the DNA barcodes with nuclease resistance that was essential for *in vivo* sensing. crRNA complementarity was critical for catalytic activation of *Lba*Cas12a by a DNA activator in solution or injection *in vivo* (Fig. 2d, Supplementary Fig. 1b-h). Additionally, the molecular weight of the DNA activators played an important role *in vivo* since they underwent characteristic single-exponential concentration decay after intravenous injection followed by size-dependent renal filtration from the blood (< 10 kDa, peaking at 1 hour after delivery)[47]. Therefore, the ~7 kDa 20-mer DNA activator that effectively concentrated in urine outperformed longer PAM-containing sequences in trans-ssDNA cutting. In localized tumors, naturally occurring ctDNAs are often present at a low abundance and require extensive enrichment steps prior to sequencing. In CRISPR-based nucleic acid detection approaches, the inputs are samples amplified by PCR or Recombinase Polymerase Amplification (RPA). At a tolerable diagnostic dose (<0.3 mg/kg), the synthetic DNA barcodes are readily accessible from compositionally simple, unprocessed urine without *in vivo* toxicity as reflected by the normal histological staining and constant level of diverse cytokines and chemokines upon sensor administration (Supplementary Fig. 10). The DNA detection sensitivity can be further improved by optimizing the ssDNA trans-cleavage activity of *Lba*Cas12a through enhanced base pairing between the DNA activator and modified crRNA. The

specialized substrate recognition and signal amplification of enzymes allows the DNA-encoded SUBs to be a fully modular system with both a designable input *via* a protease-responsive linker and a programmable output *via* CRISPR-Cas activity. We improved the on-target rate of diagnostics by using functional biologics that can be extended to scaffolds with therapeutic efficacy, providing an opportunity to combine therapy and diagnostics to monitor treatment responses in real-time[48]. Future expansion to additional enzymatic families will provide a new noninvasive way to elucidate multi-enzymatic networks at a systemic level[49]. Further refinement of the nanosensor formulation could enhance the performance of SUBs and eventually benefit patient outcomes. In addition to a systemic injection that is widely available in clinical settings, synthetic biomarker formulations that can be delivered through the oral, inhaled, and topical route, or in wearable devices, would allow for sensor administration without trained phlebotomists. This way, patients would achieve greater accessibility to constantly self-monitor their own disease progression to enable early detection and access to effective treatment. Through tailored applications, these programmable sensors may monitor other infectious and noncommunicable diseases and guide treatment decisions to improve disease management in resource-limited settings.

Methods

Peptide, oligonucleotides and peptide-oligonucleotide conjugates synthesis and characterization. All peptides were chemically synthesized and HPLC purified by CPC Scientific, Inc. All oligonucleotides were synthesized and HPLC purified by Integrated DNA Technologies, Inc. (IDT). Peptide- oligonucleotides conjugates were generated by copper-free click chemistry. The conjugates were purified on an Agilent 1100 HPLC. Mass spectral analysis of the conjugates was performed on a Bruker model MicroFlex MALDI-TOF (matrix-absorption laser desorption instrument time-of-flight). Sequences of all molecules utilized in this study were list in Supplementary Table 1 & 3.

Cas12a fluorescent cleavage assay. *Lba*Cas12a (final concentration 100 nM, New England Biolabs) was incubated with 1x NEB Buffer^{TM} 2.1, crRNA (250 nM, IDT), and complementary DNA activators (4 nM unless specifically described, IDT, in solution or spiked in urine) or urine samples collected from experimental animals, in a 50 µL reaction at 37 °C for 30 minutes. Reactions were diluted by a factor of 4 into 1x NEB Buffer ^{TM} 2.1 and ssDNA T_{10} F-Q reporter substrate (30 pmol, IDT) into a reaction volume of 60 µL per well and run in triplicate. *Lba*Cas12a activation was detected at 37 °C every 2 minutes for 3 hours by measuring fluorescence with plate reader Tecan Infinite Pro M200 (λex: 485 nm and λem: 535 nm). Sequences of all oligonucleotides are listed in Supplementary Table 1. Fluorescence for background conditions (either no DNA activator input or no crRNA conditions) were run with each assay to generate background fluorescence as negative controls. Cas12a ssDNase activity was calculated from the kinetics curve generated on the plate reader and reflected by the initial reaction velocity (V_0), which refers to the slope of the curve at the beginning of a reaction.

Cas12a cleavage assay with lateral flow readout. Samples were prepared similarly and incubated for 30 minutes at 37 °C as in Cas12a activation assay described above. Reactions were then diluted by a factor of 4 into 1 x NEB Buffer™ 2.1 and ssDNA T_{10} FAM/Biotin reporter substrate (1 pmol, IDT) into reaction volume of 100 µl. Reactions were allowed to proceed at 37 °C for 1-3 hours unless otherwise indicated, and then 20 µl was added to 80 µl of HybriDetect 1 assay buffer (Milenia). HybriDetect 1 lateral flow strips were dipped into solution and intensity of bands was quantified in ImageJ.

Characterization of DNA activator concentration or length for Cas12a ssDNase activity. To identify the optimal length for detection with Cas12a, we tested truncated native and modified DNA activator lengths from 15-34 nt and found that in the Cas12a fluorescent cleavage assay described above, Cas12a had a peak sensitivity at a native DNA activator length of 24-mer, in which contains PAM sequence and complementary sequence of crRNA. To further explore the robustness of modified DNA activator *in vivo*, phosphorothioate-modified DNA activators with different lengths were injected at 1 nmol in Balb/c mice, respectively, and urine samples were collected after 1 h of injection. Urine samples were applied as DNA activators in the Cas12a fluorescent cleavage assay, Cas12a ssDNase activity triggered by each DNA activator was normalized to that of the 24-mer modified DNA activator.

Cloning and expression of recombinant nanobodies. Double-stranded gBlocks® gene fragments encoding nanobody of interest with flanking NcoI and BlpI restriction sites, as listed below, were ordered from Integrated DNA Technologies (IA, USA). The gene fragments were cloned into Novogen pET-28a(+) expression vector at NcoI and BlpI restriction sites and transformed into SHuffle® T7 competent *E. coli*. (New England Biolabs Inc., MA, USA). Bacteria colonies encoding the correct gene inserts were confirmed with Sanger sequencing. For subsequent recombinant protein production, a 500 mL secondary culture of SHuffle® T7 competent *E. coli*. encoding nanobody gene of interest was grown in kanamycin-supplemented LB broth at 37 °C from an overnight 3-mL primary culture until optical density at 600 nm (OD600) reached about 0.6-0.8. Nanobody expression was then induced with an addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.4 mM final concentration). The culture was incubated at 27 °C for 24 h after which bacteria were pelleted and stored at -80 °C. Subsequently, the bacteria pellet was thawed on a water bath at 37 °C and lysed with B-PER™ complete bacteria protein extraction reagent (ThermoFisher Scientific, MA, USA). The released nanobody was purified via standard immobilized metal affinity chromatography (IMAC) with Ni-NTA agarose (Qiagen, MD, USA). The product was confirmed *via* SDS-PAGE analysis.). Sequences of nanobodies utilized in this study were list in Supplementary Table 3.

Synthesis of DNA-encoded synthetic urine biomarker with a nanobody core. Nanobody (2 mg) was incubated at room temperature overnight in Pierce™ immobilized TCEP disulfide reducing gel (7.5 v/v %)

(ThermoFisher Scientific, MA, USA) to selectively reduce C-terminal cysteine following a previously established protocol (Massa et al., 2014). The reduced C-terminal cysteine (1 eq.) was reacted with sulfo DBCO-maleimide crosslinker (4 eq.) (Click Chemistry Tools, AZ, USA) in PBS (pH 6.5, 1 mM EDTA) at room temperature for 6 h after which the excess crosslinker was removed with a disposable PD-10 desalting column (GE Healthcare Bio-Sciences, PA, USA). DBCO-functionalized nanobody was further refined on the fast-protein liquid chromatography (FPLC, GE Healthcare). DNA reporter conjugation was performed by incubating DBCO-functionalized nanobody (1 eq.) with azide-functionalized DNA reporter (1.1 eq.) in PBS (pH 7.4) at room temperature for 24 h. Excess DNA reporter was removed via size exclusion chromatography as described above. The product was confirmed via SDS-PAGE analysis and quantified with Quant-iT™ OliGreen™ ssDNA Assay Kit. Sequences of DNA-barcoded synthetic urine biomarkers utilized in this study were list in Supplementary Table 4.

Synthesis of DNA-encoded synthetic urine biomarkers with polymeric cores. Multivalent PEG (40 kDa, eight-arm) containing maleimide-reactive handles (JenKem Technology) was dissolved in 100 mM phosphate buffer (pH 7.0) and filtered (pore size: 0.2 μm). After filtration, the cysteine terminated peptide-DNA conjugates were added at 2-fold molar excess to the PEG and reacted for at least 4 h at room temperature. Unconjugated molecules were separated using size exclusion chromatography with Superdex 200 Increase 10/300 GL column on ÄKTA fast protein liquid chromatography (FPLC, GE Healthcare). The purified nanosensors were concentrated by spin filters (MWCO = 10 kDa, Millipore). Concentration of the nanosensor was quantified Quant-iT™ OliGreen™ ssDNA Assay Kit (ThermoFisher), fluorescence was read on a Tecan Infinite Pro M200 plate reader Quant-iT Oligreen ssDNA Reagent at λex: 485 nm and λem: 535 nm). Particles were stored at 4 °C in PBS. Dynamic light scattering (Zeta Sizer Nanoseries, Malvern Instruments, Ltd) was used to characterize the hydrodynamic diameter of the nanoparticles. Sequences of DNA-barcoded synthetic urine biomarkers utilized in this study were list in Supplementary Table 4.

Cryogenic transmission electron microscopy. The 5-plex polymeric core-based DNA-SUBs were pooled and concentrated to 0.5 mg/mL by DNA concentration. Samples were loaded on a lacey copper grid coated with a continuous carbon film. The grid was then mounted on a Gatan 626 single tilt cryo-holder equipped in the TEM column. The samples were cooled down by liquid-nitrogen and kept cold during transfer into the microscope and subsequent imaging. Images were taken on a JEOL 2100 FEG microscope that was operated at 200 kV and with a magnification in the ranges of 10,000~60,000 for assessing particle size and distribution. All images were recorded using a Gatan 2kx2k UltraScan CCD camera.

Transcriptomic and proteomic analysis. RNA-Seq data of human colon adenocarcinoma were generated by the TCGA Research Network (http://cancergenome.nih.gov). Differential expression analyses were carried out by DESeq2 1.10.1. Proteomic data on the composition of extracellular matrix in human colon

cancers and normal colon tissues were obtained by mass spectrometry analysis of ECM components and available from Matrisome (http://matrisomeproject.mit.edu/).

Cell culture. Mouse cell lines MC26-LucF (carrying firefly luciferase, from Kenneth K. Tanabe Laboratory, Massachusetts General Hospital) was cultured in DMEM (Gibco) medium supplemented with 10% (v/v) fetal bovine serum (FBS) (Gibco), 1% (v/v) penicillin/streptomycin (CellGro) at 37 ^{o}C and in 5% CO_2. Human cell lines PC-3 (ATCC® CRL-1435™, male) were grown in RPMI1640 (Gibco) supplemented with 10% (v/v) FBS and 1% (v/v) penicillin/streptomycin. RWPE1 (male) cells were cultured in Keratinocyte serum-free medium (Gibco) supplemented with 2.5 µg Human Recombinant EGF (rhEGF) and 25 mg Bovine Pituitary Extract (BPE). All cell lines tested negative for mycoplasma contamination. Cell authentication was not performed.

Animal models. All animal studies were approved by the Massachusetts Institute of Technology (MIT) committee on animal care (MIT protocol 0417-025-20 & 0217-014-20). All experiments were conducted in compliance with institutional and national guidelines and supervised by Division of Comparative Medicine (DCM) of MIT staff.

Female Balb/c (BALB/cAnNTac, 6-8 wks of age, Taconic Biosciences) and female NCr nude (CrTac:NCr-Foxn1^{nu}, 4-5 wks of age, Taconic Biosciences) were used for experiments. Mice were maintained in the Koch Cancer Institute animal facility, with a 12-h light/12-h dark cycles (7am to 7pm), at ~18-23°C and ~50% humidity. Autoclaved water and standard chow diet were accessible at all times. NCr nude mice were housed in autoclaved cages and paper bedding in a room with other immunodeficient-only animals. Cages for nude mice were only opened in a sterile hood and all tools related to experimental procedures were bleached before and after use. Health status of mice was checked weekly and daily when tumors were larger. Mice were euthanized as guided by the veterinarians when their poor body conditions reached the euthanizing criteria. We used a sample size of minimum three mice per group for *in vivo* studies, and numbers of animals per group were specified in the figure legends. Littermates of the same sex were randomly assigned to experimental and control groups.

To establishment of the nurine tumor models, Balb/c female mice were inoculated by intravenous (IV) injection with murine cell lines (100k cells/mouse, MC26-Fluc) expressing firefly luciferase. Tumor progression was monitored weekly using IVIS Imaging Systems (IVIS, PerkinElmer). To establish the prostate cancer xenograft model, NCr nude female mice were inoculated with human PC-3 cell lines (5 million cells per flank, 2 flanks per mouse) while under isoflurane anesthesia. Cells were prepared in 30%

Corning™ Matrigel™ Membrane Matrix (Thermo Fisher Scientific) and low-serum media (Opti-MEM®, Gibco). Tumors were measured weekly and experiments were conducted once flank tumors reached adequate size, which was approximately 5 mm in length or width (~200 mm^3) or three weeks after inoculation. Tumor volume was calculated by caliper measurement of the length and width of the flank; volume calculation followed the equation f_x= (width^2*length)/2, where length is the longer segment. For urine analysis, after injection with synthetic biomarkers, mice were placed into custom housing with a 96-well plate base for urine collection around noon each day. The bladders were voided to collect between 100-200 µL of urine at 1 h post injection. By the end time point of each study, mice were sacrificed using CO2 asphyxiation and cervical dislocation, and tumor tissues were collected for further analysis.

Analysis of urinary DNA barcode activated Cas12a cleavage assay. ssDNAs (1 nmol), 5-plex DNA-barcoded PEG sensors (0.2 nmol each by DNA barcode concentration, 1 nmol by DNA barcode concentration in total), or DNA-barcoded nanobody sensors (1 nmol by DNA barcode concentration) were injected into experimental mice via intravenous injection. Urine samples were collected after 1 h and used as DNA activator in Cas12a fluorescent cleavage assay described above. The initial reaction velocity (V_0) is determined from the slope of the curve at the beginning of a reaction. Mean normalization was performed on V_0 values to account for animal-to-animal variation in urine concentration. In the Cas12a cleavage assay with fluorescent reporter, Y axis represents $\text{MeanNormV}_{0\ \text{Tumor-bearing animals}}/\text{MeanNormV}_{0\ \text{control animals.}}$ Then the same urine sample were utilized to perform the Cas12a cleavage assay with LFA readout. Resulting paper strips were aligned and scanned simultaneously, intensity of control and sample bands were quantified from the scanned images in ImageJ.

Biodistribution and pharmacokinetics studies. Studies were performed in experimental animals with near-infrared dye labeled agents to minimize interference from autofluorescent background. Balb/c mice were intravenously injected with Cy5-labeled modified or native DNA molecules at 1 nmol per mouse, n=3 per condition. Urine samples from each mouse was collected at 30 min, 1, 2, 3, 4 hours after injection. C-met nanobodies were coupled with Sulfo-Cyanine7 NHS ester (Lumiprobe), reacted overnight, purified by spin filtration and injected into PC-3 tumor-bearing nude mice (1.5 nmol dye eq. of protein) *via* i.v. injection. After 24 hours, mice were euthanized and necropsy was performed to remove the tumors, lungs, heart, kidneys, liver, and spleen. Urine, blood and organs were scanned using IVIS Imaging Systems and Odyssey® CLx (Li-COR). Organ fluorescence was quantified in ImageStudio of Odyssey® CLx. Blood circulatory kinetics were monitored in Balb/c mice by serial blood draws at 10 min, 30 min, 2 hr and 3 hrs after i.v. injection of Cy5-labeled DNA or PEG at 1 nmol dye per mouse. Blood for pharmacokinetics measurements was collected using tail vain bleeds. Blood was diluted in PBS with 5 mM EDTA to prevent clotting, centrifuged for 5 min at 5,000 x g, and fluorescent reporter concentration was quantified in 384-well plates relative to standards (LI-COR Odyssey® CLx).

Histology, Immunohistochemistry (IHC) and Immunofluorescence (IF) studies. Paraffin-embedded tissues were preserved in 4% paraformaldehyde (PFA) overnight and stored in 70% ethanol prior to embedding into paraffin. Snap-frozen tissues were preserved in 2% PFA for two hours, stored in 30% sucrose overnight and frozen in optimum cutting temperature (OCT) compound at -80°C. Snap-frozen lungs were processed through intratracheal injection of 50: 50 OCT in PBS immediately after animal euthanasia via isoflurane overdose. The lungs were slowly frozen with OCT embedding in isopentane liquid nitrogen bath. Samples were sectioned into 6 µm slices and stained for further analysis. For IHC studies, slides were stained with primary antibodies in accordance with manufacturer instructions, followed by HRP secondaries. For IF studies, after blocking with 5% goat serum, 2% BSA, 0.1% Triton-X 100 in PBS for 1 h, sections were stained with a primary antibody in1% BSA in PBS overnight at 4 °C. AlexaFluor conjugated secondary antibodies were incubated at 1 µg/mL in 1% BSA in PBS for 30 min at RT. Slides were sealed with ProLong Antifade Mountants (Thermo Scientific). Slides were digitized and analyzed using an 3D Histech P250 High Capacity Slide Scanner (Perkin Elmer). Antibodies and dilutions used were listed in Supplementary Table 5.

RNA extraction and RT-qPCR. PC-3 and RWPE1 cells were cultured and collected after trypsinization. Tissue samples were collected by necropsy after mice were euthanized and were immediately kept in RNAlater RNA Stabilization Reagent (Qiagen, Inc.). RNA from cell pellets or cryoground tissue samples was extracted using RNeasy Mini Kit (Qiagen, Inc.). RNA was reverse transcribed into cDNA using BioRad iScript Reverse Transcription Supermix on a Bio-Rad iCycler. qPCR amplification of the cDNA was measured after mixing with Taqman gene expression probes and Applied Biosystems TaqMan Fast Advanced Master Mix (Thermo Scientific) according to manufactory's instruction. qPCR was performed on a CFX96 Real Time System C1000 Thermal Cycler from Bio-Rad.

Recombinant protease substrate cleavage assay. Fluorogenic protease substrates with fluorescence (FAM) and quencher (CPQ2) were synthesized by CPC Scientific Inc. Recombinant proteases were purchased from Enzo Life Sciences and R&D Systems. Assays were performed in the 384-well plate in triplicate in enzyme-specific buffer with peptides (1 µM) and proteases (40 nM) in 30 µL at 37 °C. Fluorescence was measured at Ex/Em 485/535 nm using a Tecan Infinite 200pro microplate reader (Tecan). Signal increase at 60 minutes was used across conditions. Enzymes and buffer conditions were listed in Supplementary Table 6.

Protein extraction and tissue lysate proteolytic cleavage assay. Tissue samples were homogenized in PBS and centrifuged at 4 °C for 5 minutes at 6,000 x g. Supernatant was further centrifuged at 14,000 x g

for 25 minutes at 4 °C. Protein concentration was measured using ThermoFisher BCA Protein Assay Kit and prepared at 2 mg/mL prior to assay. Assays were performed in the 384-well plate in triplicate in enzyme-specific buffer with peptides (1 μM) and cell lysates (0.33 mg/mL) in 30 μL at 37 °C. Fluorescence was measured at Ex/Em 485/535 nm using a Tecan Infinite 200pro microplate reader (Tecan). Signal increase at 60 minutes was used across conditions.

PSA enzyme-linked immunosorbent assay (ELISA). Approximately 200 μL of blood was collected from the lateral saphenous vein of experimental animals and blood cells were pelleted immediately by centrifugation at 14,000 x g for 25 minutes at 4 °C. Plasma was stored at -80 ºC prior to PSA quantification. PSA levels were measured using the PSA Quantikine ELISA kit according to manufacturer protocols (R&D Systems).

Statistical analysis and reproducibility. Statistical analyses were conducted in GraphPad Prism 9 for macOS. Data were presented as means with standard error of the mean (SEM). Differences between groups were assessed using parametric and non-parametric group comparisons when appropriate with adjustment for multiple hypothesis testing. Results were tested for statistical significance by Student's unpaired *t*-test (parametric) for two group comparisons and ANOVA for multiple group comparisons. Sample sizes and statistical tests are specified in the figure legends. All experiments were repeated independently with similar results.

Declarations

Acknowledgments

We thank Dr. Kathleen Cormier from the KI histology core for tissue sectioning and staining; Dr. Virginia Spanoudaki and Sarah Elmiligy from the KI Animal Imaging & Preclinical Testing Core for assistance with IVIS imaging; Richard Cook, Heather Amoroso, and Alla Leshinsky from the KI Biopolymers & Proteomics Core for help with the HPLC and mass spectrometry; Dr. DongSoo Yun from the KI Nanotechnology Materials Core for Cyro-TEM; Dr. Duanduan Ma from the KI Bioinformatics & Computing Core for assistance with transcriptome data analysis. We would also like to thank Drs. Qian Zhong and Edward Tan for helpful discussion on the manuscript. This study was supported in part by a Koch Institute Support Grant No. P30-CA14051 from the National Cancer Institute (Swanson Biotechnology Center), and a Core Center Grant P30-ES002109 from the National Institute of Environmental Health Sciences. L.H. is supported by a K99/R00 Pathway to Independence Award from the National Cancer Institute. S.N.B. is a Howard Hughes Medical Institute Investigator.

Data availability

The Cancer Genome Atlas (TCGA) Research Network (http://cancergenome.nih.gov) and Matrisome (http://matrisomeproject.mit.edu/) are open access resources. All data that support the findings of this study are available within the article and Supplementary Information or from the corresponding author upon reasonable request.

Author contributions

L.H., H.E.F., S.N.B. conceived and designed the research. L.H., R.T.Z., C.N. H.K. performed research. L.H., H.E.F., S.N.B. wrote and edited the paper.

Declaration of Interests

S.N.B., L.H. are listed as inventors on patent applications related to the content of this work. S.N.B. holds equity in Glympse Bio, Satellite Bio, Cend Therapeutics and Catalio Capital; is a director at Vertex; consults for Moderna, and receives sponsored research funding from Johnson & Johnson.

References

1 Kwong, G. A. *et al.* Mass-encoded synthetic biomarkers for multiplexed urinary monitoring of disease. *Nat Biotechnol* **31**, 63-70, doi:10.1038/nbt.2464 (2013).

2 Ronald, J. A., Chuang, H. Y., Dragulescu-Andrasi, A., Hori, S. S. & Gambhir, S. S. Detecting cancers through tumor-activatable minicircles that lead to a detectable blood biomarker. *Proc Natl Acad Sci U S A* **112**, 3068-3073, doi:10.1073/pnas.1414156112 (2015).

3 Hori, S. S. & Gambhir, S. S. Mathematical model identifies blood biomarker-based early cancer detection strategies and limitations. *Sci Transl Med* **3**, 109ra116, doi:10.1126/scitranslmed.3003110 (2011).

4 Badeau, B. A., Comerford, M. P., Arakawa, C. K., Shadish, J. A. & DeForest, C. A. Engineered modular biomaterial logic gates for environmentally triggered therapeutic delivery. *Nat Chem* **10**, 251-258, doi:10.1038/nchem.2917 (2018).

5 Koo, K. M., Mainwaring, P. N., Tomlins, S. A. & Trau, M. Merging new-age biomarkers and nanodiagnostics for precision prostate cancer management. *Nat Rev Urol* **16**, 302-317, doi:10.1038/s41585-019-0178-2 (2019).

6 Aalipour, A. *et al.* Engineered immune cells as highly sensitive cancer diagnostics. *Nat Biotechnol* **37**, 531-539, doi:10.1038/s41587-019-0064-8 (2019).

7 Mannocci, L. Z., Y.; Scheuermann, J.; Leimbacher, M., Bellis, G.D.; Rizzi, E.; Dumelin, C.; Melkko, S.; Neri, D. High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proc Natl Acad Sci U S A* **105**, 17670-17675, doi:https://doi.org/10.1073/pnas.0805130105 (2008).

8 Burgess, D. J. Spatial transcriptomics coming of age. *Nat Rev Genet* **20**, 317, doi:10.1038/s41576-019-0129-z (2019).

9 Ullal, A. V. *et al.* Cancer cell profiling by barcoding allows multiplexed protein analysis in fine-needle aspirates. *Sci Transl Med* **6**, 219ra219, doi:10.1126/scitranslmed.3007361 (2014).

10 Dahlman, J. E. *et al.* Barcoded nanoparticles for high throughput in vivo discovery of targeted therapeutics. *Proc Natl Acad Sci U S A* **114**, 2060-2065, doi:10.1073/pnas.1620874114 (2017).

11 Cristiano, S. *et al.* Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385-389, doi:10.1038/s41586-019-1272-6 (2019).

12 Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164**, 57-68, doi:10.1016/j.cell.2015.11.050 (2016).

13 Chabon, J. J. *et al.* Integrating genomic features for non-invasive early lung cancer detection. *Nature* **580**, 245-251, doi:10.1038/s41586-020-2140-0 (2020).

14 Sun, K. *et al.* Orientation-aware plasma cell-free DNA fragmentation analysis in open chromatin regions informs tissue of origin. *Genome Res* **29**, 418-427, doi:10.1101/gr.242719.118 (2019).

15 Larson, M. H., Pan, W., Kim, H.J. et al. . A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. *Nat Commun* **12**, 2357, doi:https://doi.org/10.1038/s41467-021-22444-1 (2021).

16 Smith, C. I. E. & Zain, R. Therapeutic Oligonucleotides: State of the Art. *Annu Rev Pharmacol Toxicol* **59**, 605-630, doi:10.1146/annurev-pharmtox-010818-021050 (2019).

17 Tabrizi, S. J. *et al.* Targeting Huntingtin Expression in Patients with Huntington's Disease. *N Engl J Med* **380**, 2307-2316, doi:10.1056/NEJMoa1900907 (2019).

18 Khvorova, A. & Watts, J. K. The chemical evolution of oligonucleotide therapies of clinical utility. *Nat Biotechnol* **35**, 238-248, doi:10.1038/nbt.3765 (2017).

19 Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436-439, doi:10.1126/science.aar6245 (2018).

20 Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat Med* **19**, 1423-1437, doi:10.1038/nm.3394 (2013).

21 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

22 Gootenberg, J. S. *et al.* Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* **360**, 439-444, doi:10.1126/science.aaq0179 (2018).

23 Liu, J. J. *et al.* CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218-223, doi:10.1038/s41586-019-0908-x (2019).

24 English, M. A. *et al.* Programmable CRISPR-responsive smart materials. *Science* **365**, 780-785, doi:10.1126/science.aaw5122 (2019).

25 Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356**, 438-442, doi:10.1126/science.aam9321 (2017).

26 Myhrvold, C. *et al.* Field-deployable viral diagnostics using CRISPR-Cas13. *Science* **360**, 444-448, doi:10.1126/science.aas8836 (2018).

27 Harrington, L. B. *et al.* Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* **362**, 839-842, doi:10.1126/science.aav4294 (2018).

28 Stein, C. A. & Castanotto, D. FDA-Approved Oligonucleotide Therapies in 2017. *Mol Ther* **25**, 1069-1075, doi:10.1016/j.ymthe.2017.03.023 (2017).

29 Shen, W. *et al.* Chemical modification of PS-ASO therapeutics reduces cellular protein-binding and improves the therapeutic index. *Nat Biotechnol* **37**, 640-650, doi:10.1038/s41587-019-0106-2 (2019).

30 Kim, J. *et al.* Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. *N Engl J Med* **381**, 1644-1652, doi:10.1056/NEJMoa1813279 (2019).

31 Dudani, J. S., Ibrahim, M., Kirkpatrick, J., Warren, A. D. & Bhatia, S. N. Classification of prostate cancer using a protease activity nanosensor library. *Proc Natl Acad Sci U S A* **115**, 8954-8959, doi:10.1073/pnas.1805337115 (2018).

32 Morrison, C. Nanobody approval gives domain antibodies a boost. *Nat Rev Drug Discov* **18**, 485-487, doi:10.1038/d41573-019-00104-w (2019).

33 Massa, S. *et al.* Site-specific labeling of cysteine-tagged camelid single-domain antibody-fragments for use in molecular imaging. *Bioconjug Chem* **25**, 979-988, doi:10.1021/bc500111t (2014).

34 Kirley, T. L., Greis, K. D. & Norman, A. B. Selective disulfide reduction for labeling and enhancement of Fab antibody fragments. *Biochem Biophys Res Commun* **480**, 752-757, doi:10.1016/j.bbrc.2016.10.128 (2016).

35 Muyldermans, S. Nanobodies: natural single-domain antibodies. *Annu Rev Biochem* **82**, 775-797, doi:10.1146/annurev-biochem-063011-092449 (2013).

36 Fridy, P. C. *et al.* A robust pipeline for rapid production of versatile nanobody repertoires. *Nat Methods* **11**, 1253-1260, doi:10.1038/nmeth.3170 (2014).

37 Danino, T. *et al.* Programmable probiotics for detection of cancer in urine. *Sci Transl Med* **7**, 289ra284, doi:10.1126/scitranslmed.aaa3519 (2015).

38 Hynes, R. O. & Naba, A. Overview of the matrisome--an inventory of extracellular matrix constituents and functions. *Cold Spring Harb Perspect Biol* **4**, a004903, doi:10.1101/cshperspect.a004903 (2012).

39 Naba, A. *et al.* The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol Cell Proteomics* **11**, M111 014647, doi:10.1074/mcp.M111.014647 (2012).

40 Kirkpatrick, J. D. *et al.* Urinary detection of lung cancer in mice via noninvasive pulmonary protease profiling. *Sci Transl Med* **12**, doi:10.1126/scitranslmed.aaw0262 (2020).

41 Devoogdt, N. *et al.* Secretory leukocyte protease inhibitor promotes the tumorigenic and metastatic potential of cancer cells. *Proc Natl Acad Sci U S A* **100**, 5778-5782, doi:10.1073/pnas.1037154100 (2003).

42 Pawar, N. R., Buzza, M. S. & Antalis, T. M. Membrane-Anchored Serine Proteases and Protease-Activated Receptor-2-Mediated Signaling: Co-Conspirators in Cancer Progression. *Cancer Res* **79**, 301-310, doi:10.1158/0008-5472.CAN-18-1745 (2019).

43 Fayard, B. *et al.* The serine protease inhibitor protease nexin-1 controls mammary cancer metastasis through LRP-1-mediated MMP-9 expression. *Cancer Res* **69**, 5690-5698, doi:10.1158/0008-5472.CAN-08-4573 (2009).

44 Kwong, G. A. *et al.* Mathematical framework for activity-based cancer biomarkers. *Proc Natl Acad Sci U S A* **112**, 12627-12632, doi:10.1073/pnas.1506925112 (2015).

45 Kaminski, M. M. *et al.* A CRISPR-based assay for the detection of opportunistic infections post-transplantation and for the monitoring of transplant rejection. *Nat Biomed Eng*, doi:10.1038/s41551-020-0546-5 (2020).

46 Ballard, Z. S. J., H.A., Goncharov, A.; Liang, J.; Nugroho, K.; Carlo, D.D.; Garner O.B.; Ozcan, A. Deep learning-enabled point-of-care sensing using multiplexed paper-based sensors. *npj Digital Medicine* **3**, doi:https://doi.org/10.1038/s41746-020-0274-y (2020).

47 Choi, C. H., Zuckerman, J. E., Webster, P. & Davis, M. E. Targeting kidney mesangium by nanoparticles of defined size. *Proc Natl Acad Sci U S A* **108**, 6656-6661, doi:10.1073/pnas.1103573108 (2011).

48 Tauriello, D. V. F. *et al.* TGFbeta drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature* **554**, 538-543, doi:10.1038/nature25492 (2018).

49 Chakladar, S. *et al.* A mechanism-based inactivator of glycoside hydrolases involving formation of a transient non-classical carbocation. *Nat Commun* **5**, 5590, doi:10.1038/ncomms6590 (2014).

Figures

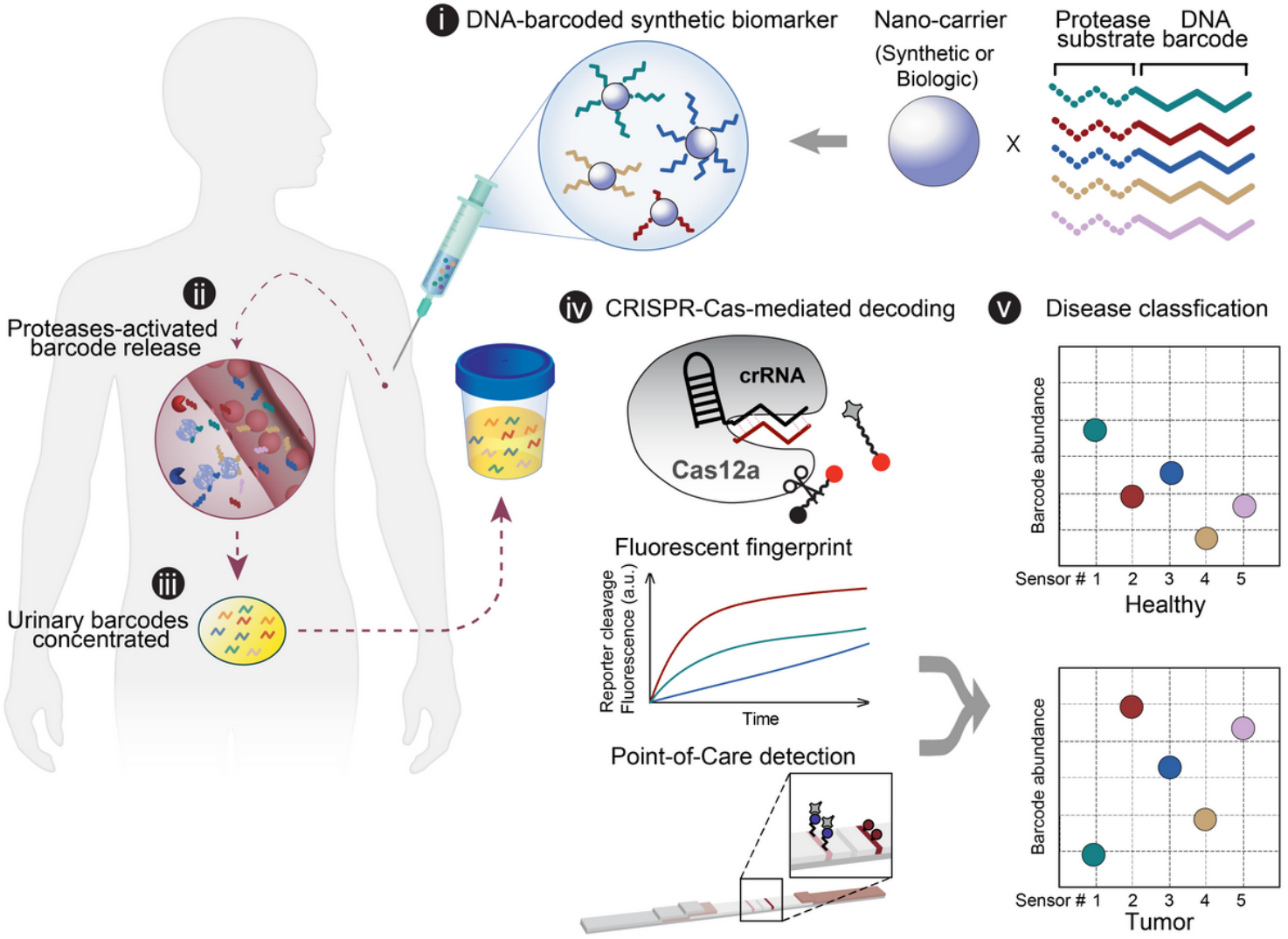

Figure 1

Engineering DNA-encoded synthetic urine biomarkers with CRISPR-Cas-mediated disease detection. DNA-encoded synthetic urine biomarkers are comprised of a nano-carrier (synthetic or biologic) functionalized with protease-activated short peptides barcoded with oligonucleotides (i). After in vivo administration, activation of DNA-encoded sensors by disease-specific protease activity triggers release of synthetic DNA barcodes (ii) that are size-specifically concentrated in the urine for disease monitoring (iii). DNA barcodes in the urine activate programmable CRISPR-Cas nucleases to release the multiplexed reporter signals that

are fluorescent, or detected on paper (iv), enabling in situ disease classification at the point-of-care via the patterns of local proteolytic activities in the disease microenvironment (v).

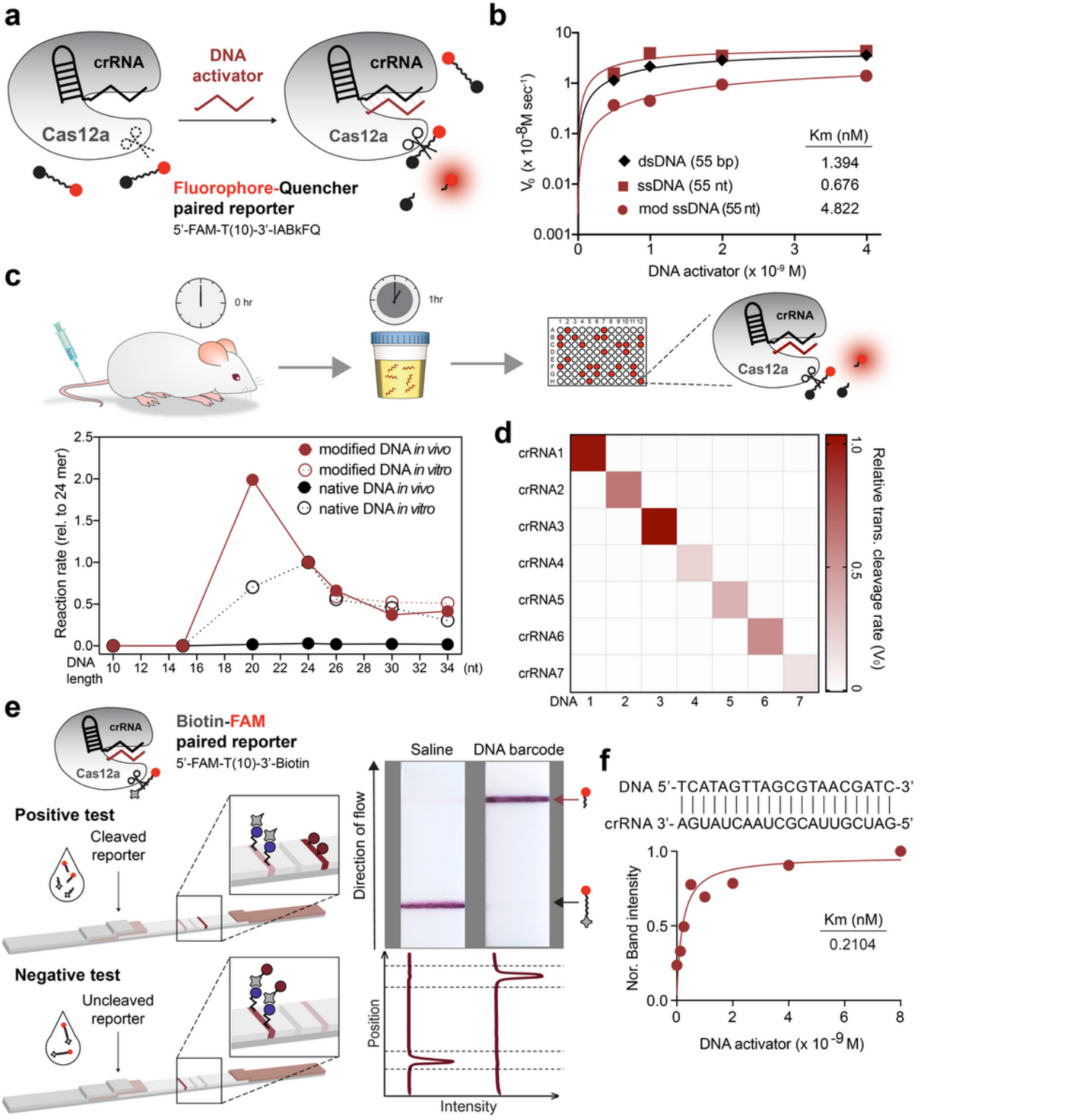

Figure 2

Chemically-modified DNA enables CRISPR-Cas-mediated urinary readout for in vivo sensing. a, DNA fragment activates nonspecific single-stranded DNase cleavage upon binding to crRNA on Cas12a. Such activity can be tracked by the dequenching of fluorescence from a fluorophore (F, 5'-FAM)-quencher (Q, 3'-

IABkFQ)-paired oligonucleotide. b, Representative Michaelis-Menten plot of LbaCas12a-catalyzed ssDNA trans cleavage using a native dsDNA, ssDNA, or fully phosphorothioate-modified ssDNA activator. The initial reaction velocity (V0) is determined from the slope of the curve at the beginning of a reaction. c, Schematic showing urine testing in a mouse model and the study time course (1 h). Plot underneath shows length optimization of ssDNA activator in vitro and in vivo, by quantifying the trans-cleavage rate of Cas12a upon activation of native or modified ssDNA in solution (4 nM) or mouse urine (1 nmol per injection). The trans-cleavage rates in each condition in the format of initial reaction velocity were normalized to that of a 24-mer. d, Heatmap of trans-cleavage rates of different modified ssDNA activator-crRNA pairs. Assays were performed with urine samples collected from mice injected with 1 nmol of modified ssDNA activator after 1 h of i.v. administration. e, Schematic showing set up of paper-based lateral flow assay (LFA) (left). When Cas12a is activated by the DNA activator in mouse urine, it cleaves the fluorescein (FAM)-biotin-paired oligonucleotide reporter and frees the FAM molecule that can be detected on the ‘sample band’. Uncleaved reporters are trapped on the ‘control band’ via binding of biotin to streptavidin. Different bands are visible on paper strips (right). Band intensities were quantified using ImageJ, and each curve was aligned below the corresponding paper strip. The top peak of the curve shows the freed FAM molecule in cleaved reporter samples, and the bottom peak shows the presence of the uncleaved FAM-biotin reporter. f, Michaelis-Menten plot of LbaCas12a-catalyzed ssDNA trans-cleavage upon a representative DNA-crRNA pairing (complementary sequences are shown) on paper. Data were plotted with the quantified band intensity of cleaved reporter on paper strips.

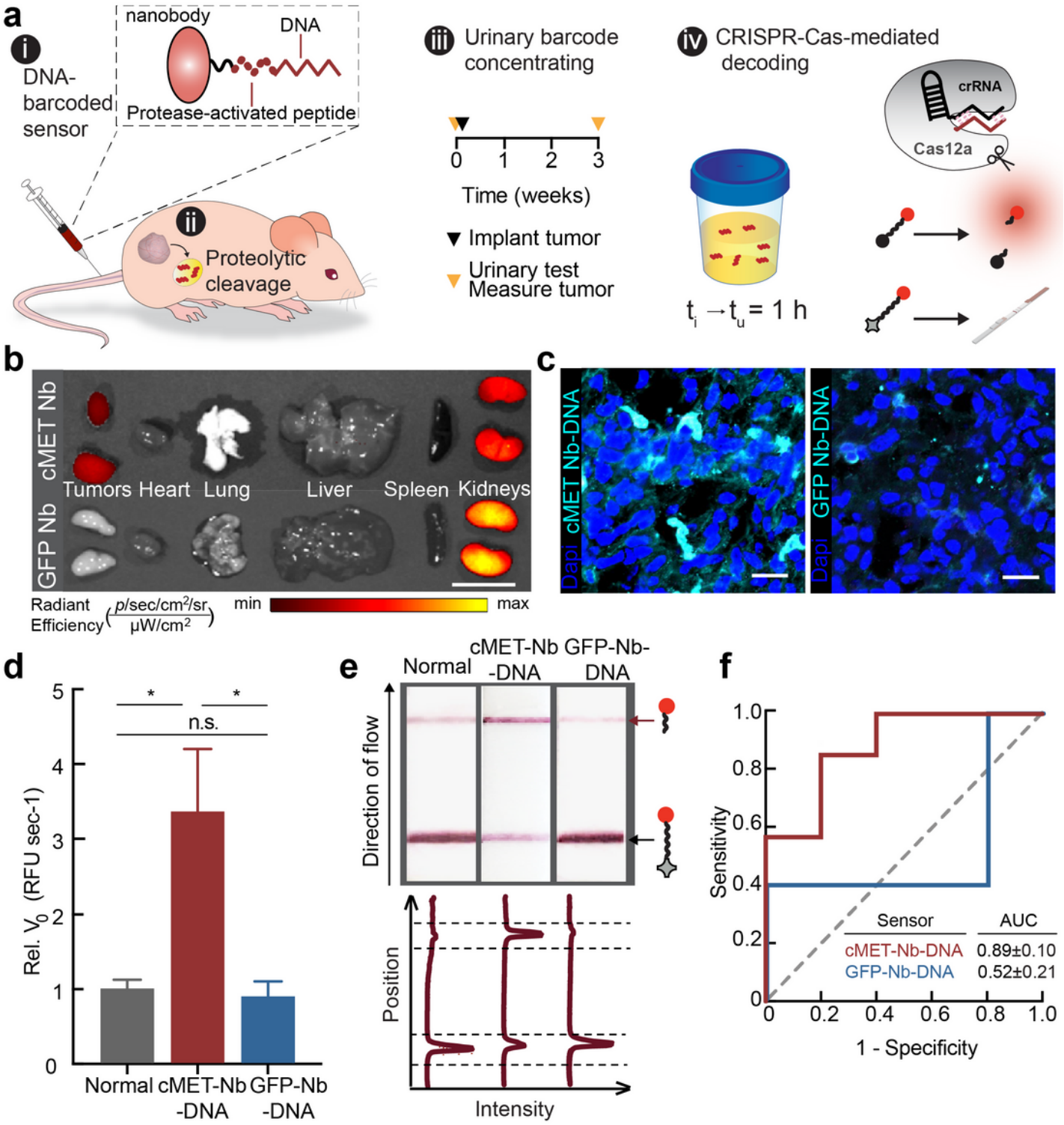

Figure 3

Tumor-targeting, singleplex DNA-encoded synthetic urine biomarker for detection of prostate cancer. a, Schematic showing urine testing in a human prostate cancer xenograft model: generation of the DNA-encoded protease-activatable nanobody with an unpaired cysteine and 1-step conjugation of ssDNA activator (i). Activation of diagnostic by disease-specific protease activity triggers (ii) release of ssDNA activator into urine for disease detection (iii). Study time course of urine testing and detection of the

ssDNA activator with Cas12a trans-cleavage assay using fluorescent or paper readout (iv). t, time at given testing step; ti, time of sensor injection; tu, time of urine collection. b, IVIS image shows biodistribution of cMET targeting nanobody and on-targeting GFP nanobody when injected intravenously in nude mice bearing PC-3 xenografts. Scale bar = 1 cm. c, Immunofluorescent staining of Cy7-labeled DNA-encoded cMET nanobody and DNA-encoded, non-targeting GFP nanobody on sections of PC-3 tumors. Scale bar = 20 μm. d, Unprocessed urine samples collected from tumor-bearing mice injected with DNA-encoded cMET nanobody or DNA-encoded GFP nanobody, and healthy control mice injected with DNA-encoded cMET nanobody were applied in the Cas12a trans-cleavage assay. Initial reaction velocity (V0) of the Cas12a trans-cleavage assays were calculated and normalized to that of healthy control mice (n=5 for control groups, n= 7 mice for DNA-encoded cMET nanobody injected group; ± SEM; unpaired two-tailed t-test with Welch's correction, *P<0.05). e, Paper-based LFA of Cas12a activated by urine samples collected from tumor-bearing or healthy control mice in (c). Band intensities were quantified using ImageJ and each curve was aligned below the corresponding paper strip. The top peak of the curve shows the presence of the cleaved reporter and the bottom peak shows the presence of the uncleaved reporter. f, ROC curves characterize the predictive power of a biomarker by returning the area under the curve (AUC) as a metric, with a baseline AUC of 0.5 representing a random biomarker classifier. AUC comparison between DNA-encoded cMET nanobody or DNA-encoded GFP nanobody injected tumor cohort against normal cohort in (c). Dashed line represents an AUC of 0.5, and a perfect AUC is 1.0.

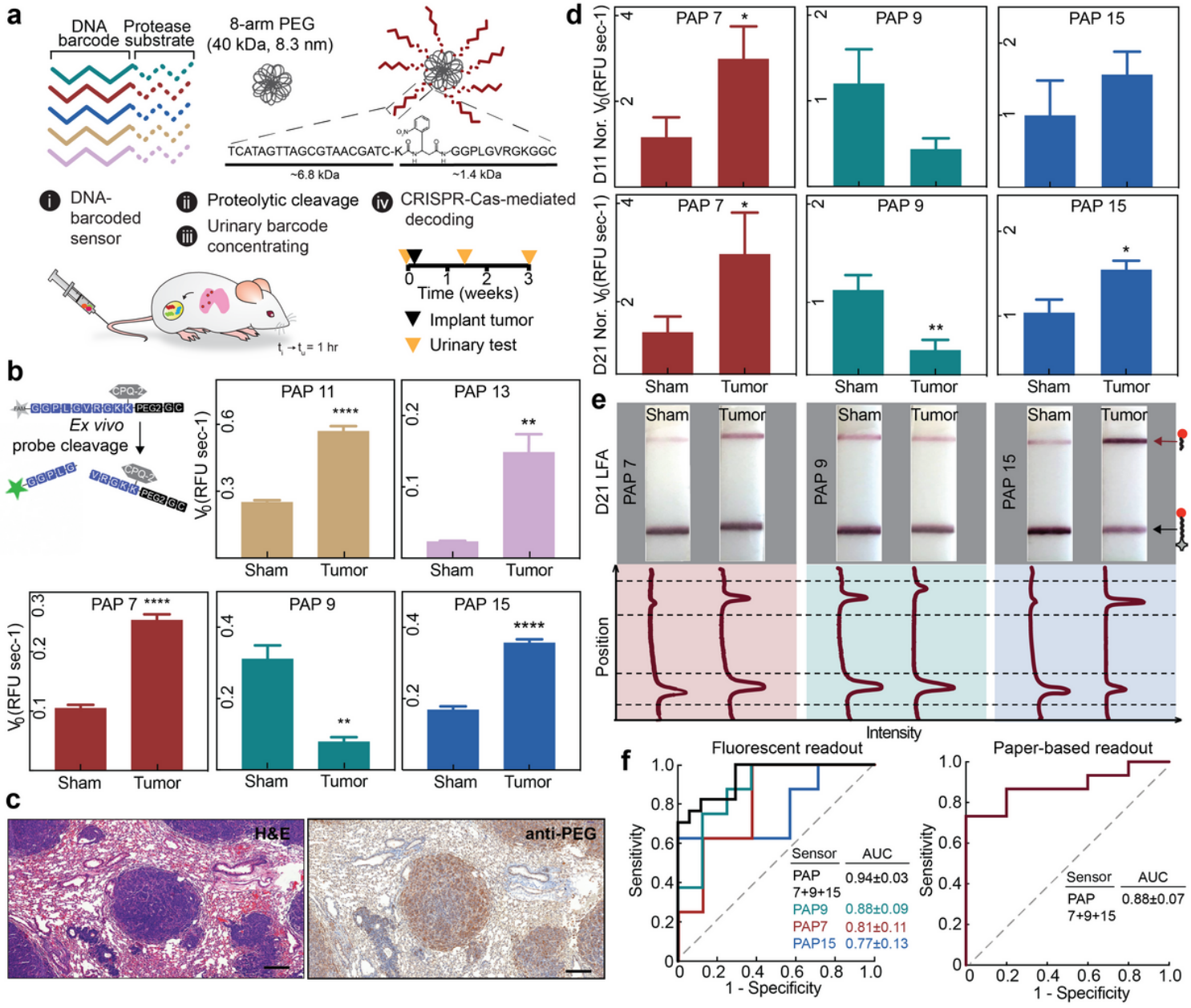

Figure 4

DNA-encoded multiplex synthetic urine biomarkers for portable monitoring of invasive CRC. a, Scheme of the workflow for longitudinal disease monitoring with the multiplexed DNA-encoded synthetic urine biomarkers. b, Forster resonance energy transfer (FRET)-based peptide assay to identify the real-time cleavage of peptide substrates by invasive CRC tissue homogenates collected 21 days after tumor inoculation. Peptide cleavage kinetics were monitored and cleavage rates were plotted (n=5 mice per group; ± SEM; unpaired two-tailed t-test with Welch's correction, **P<0.01, ****P<0.0001). c, Histological staining of lung sections of Balb/c mice bearing CRC lung tumor nodules 21 days after tumor inoculation (left). Mice were injected with 5-plex DNA-SUBs with the PEG core and immunohistochemistry of the same tissue stained with anti-PEG antibody (right). Scale bar = 200 μm. d, Pooled DNA-SUBs were administered to Balb/c mice bearing CRC lung tumor nodules (tumor) and saline-injected control animals

(sham) at day 11 or 21 after tumor initiation. All urine samples were collected at 1 h after sensor administration. Cas12a trans-cleavage assays were performed against each DNA-barcode with the fluorophore-quencher paired reporter. Initial reaction velocity (V0) of the Cas12a trans-cleavage assays were calculated and normalized to that of saline injected control animals (n=8 or 10 mice per group; ± SEM; unpaired two-tailed t-test with Welch's correction, *P<0.05, **P<0.01). The initial reaction velocity (V0) refers to the slope of the curve at the beginning of a reaction. e, Representative paper strips of the paper-based LFA of Cas12a activated by mouse urine samples collected in (d). Band intensities were quantified using ImageJ. The top peak of the curve shows the freed FAM molecule in cleaved reporter and the bottom peak shows the presence of the uncleaved FAM-biotin reporter. f, ROC curves showing the ability of DNA-SUBs to distinguish between diseased mice and healthy controls solely or in combination. The area under the curve was calculated. Left, ROC curve analysis indicates predictive ability of single or combined DNA-SUBs with fluorescent readout in (d). Combined sensors PAP7/9/15 AUC=0.94±0.03, P<0.0001; PAP7 AUC= 0.81±0.11, P=0.03; PAP9 AUC=0.88±0.08, P=0.01; PAP15 AUC=0.77±0.13, P=0.08. Right, ROC curve shows the predictive ability of paper-based urinary readout in (e). Combined sensors PAP7/9/15 AUC=0.88 ± 0.07, P=0.004. This ROC analysis utilized ratio of quantified cleaved reporter band intensity over its corresponding control band intensity. Dashed line represents an AUC of a random biomarker classifier (0.5), and the AUC of a perfect biomarker is 1.0.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 4.SupplementalInformationRSV1.docx