Preprint of

Chen, J., Qiu, T., Mauk, M. G., Su, Z., Fan, Y., Yuan, D. J., Zhou, Q., Qiao, Y., Bau, H. H., Ying, J., & Song, J. (2022). Programmable endonuclease combined with isothermal polymerase amplification to selectively enrich for rare mutant allele fractions. Chinese Chemical Letters, 33(8), 4126–4132.

https://doi.org/10.1016/j.cclet.2021.11.065

Programmable Endonuclease Combined with Isothermal Polymerase Amplification to Selectively Enrich for Rare Mutant Allele Fractions

Junman Chen[1,2,3†], Tian Qiu[4†], Michael G. Mauk[3], Zheng Su[5], Yaguang Fan[6], Dennis J. Yuan[2], Qinghua Zhou[6], Youlin Qiao[5], Haim H. Bau[3], Jianming Ying[4*], Jinzhao Song[2,3*]

[1] Key Laboratory of Clinical Laboratory Diagnostics (Ministry of Education), College of Laboratory Medicine, Chongqing Medical University, Chongqing, 400016, China
[2] The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, Zhejiang 310022, China
[3] Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia PA, 19104, USA
[4] Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, China.
[5] Center for Global Health, School of Population Medicine and Public Health, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100730, China
[6] Tianjin Key Laboratory of Lung Cancer Metastasis and Tumor Microenvironment, Tianjin Lung Cancer Institute, Tianjin Medical University General Hospital, Tianjin, 300052, China.

* To whom correspondence should be addressed. Tel: 215-898-9380; Fax: 215-573-6334; Email: songjinz@seas.upenn.edu, Correspondence may also be addressed to jmying@cicams.ac.cn.

† The first two authors contributed equally.

Abstract

Liquid biopsy is a highly promising method for non-invasive detection of tumor-associated nucleic acid fragments in body fluids but is challenged by the low abundance of nucleic acids of clinical interest and their sequence homology with the vast background of nucleic acids from healthy cells. Recently, programmable endonucleases such as CRISPR-Cas and prokaryotic Argonautes have been successfully used to remove background nucleic acids and enrich mutant allele fractions, enabling their detection with deep NGS. However, the enrichment level achievable with these assays is limited by futile binding events and off-target cleavage. To overcome these shortcomings, we conceived a new assay (PASEA) that combines the cleavage of wild type alleles with concurrent polymerase amplification. While PASEA increases the numbers of both wild type and mutant alleles, the numbers of mutant alleles increase at much greater rates, allowing PASEA to achieve an unprecedented level of selective enrichment of targeted alleles. By combining CRISPR-Cas9 based cleavage with Recombinase Polymerase Amplification, we converted samples with 0.01% somatic mutant allele fractions (MAFs) to products with 70% MAFs in a single step within 20 min, enabling inexpensive, rapid genotyping with such as Sanger sequencers. Furthermore, PASEA's extraordinary efficiency facilitates sensitive real-time detection of somatic mutant alleles at the point of care with custom designed Exo-RPA probes. Real-time PASEA' performance was proved equivalent to clinical ARMS-PCR and NGS when testing over hundred cancer patients' samples. This strategy has the potential to reduce the cost and time of cancer screening and genotyping, and to enable targeted therapies in resource-limited settings.

Keywords

Mutant allele enrichment; Programmable endonuclease; Liquid biopsy; Mutation detection; Point-of-care testing; CRISPR-Cas9; Recombinase Polymerase Amplification; Nucleic acid diagnostics

Introduction

Somatic mutations are implicated in carcinogenesis, cancer progression, and therapeutic resistance. The detection of rare somatic mutant alleles (MAs) in cancer biopsy and liquid biopsy is challenged by their low abundance and sequence homology with a vast background of wild-type (WT) alleles from healthy cells. These challenges are greatly aggravated at early disease stages and during the evolution of drug resistant mutations. To detect low abundance somatic mutations, it is necessary to improve the signal to noise ratio by enriching the rare somatic mutant allele fractions (MAFs). This is accomplished by hybridizing nucleic acids of interest to nucleic acid probes in the presence of nucleic-acid guided endonucleases lacking catalytic activity (e.g., dCas9) (*1, 2*); by preferential enzymatic amplification of mutant alleles with specifically designed primers and DNA blockers (*3-5*); by suppression of the amplification of wild-type alleles with capping nucleic acids (*6*); and by selective depletion of wild-type nucleic acids with RNase H at post-transcriptional level (*7*); with restriction endonucleases (*8, 9*); and programmable endonuclease such as CRISPR Cas 9 (*10, 11*) and Argnoautes (*12-14*), wherein unwanted background sequences are selectively removed from the sample. These various methods can be used independently or in combination. The targeted nucleic acids can then be detected either directly with probes during amplification, hybridization arrays, and sequencing.

Among the aforelisted methods, selective depletion methods based on programmable endonucleases (*10-13*) provide greater flexibility than those based on restriction endonuclease (*8, 9*) since they do not rely on specific target sequences. Guided by synthetic oligonucleotides or silencing RNA, programmable endonucleases (*15-18*) remove the dominant, interfering (background) wild-type sequences to facilitate detection of scarce mutant alleles. While greatly improving the sensitivity of downstream genotyping methods, the efficacy of existing programmable Cas9-based enrichment assays is compromised by a significant fraction of unfertile binding events between enzyme and target, which due to the slow dissociation rate of Cas9 leave targets protected from cleavage (*19*) and by non-specific off-target cleavage that depletes precious biomarkers (*11*). To partially overcome these shortcomings, researchers have employed rounds of selective depletion of WT followed by polymerase chain reaction (PCR) amplification, enabling tens-fold enrichment of the fractions of mutant alleles (*11*). Despite these improvements, high sensitivity of mutant allele detection still requires the use of deep NGS (*11*), rendering these methods laborious, time-consuming, and expensive. A single-step method that

enriches the MAF to enable its detection by inexpensive and rapid means is highly desirable.

To address this need, we devised a new assay dubbed Programmable Enzyme-Assisted Selective Exponential Amplification (PASEA, **Figure 1**) that concurrently amplifies both wild type and mutant alleles in the presence of guided endonuclease that targets only the wild type allele. Given time, the variant that exhibits a superior trait (the mutant allele being less susceptible to cleavage) will dominate. PASEA requires temperature-matched polymerase and endonuclease. Herein, we use CRISPR-Cas9 programmed to cleave wild type alleles in combination with isothermal recombinase polymerase amplification (RPA). We converted samples with 0.01% somatic MAFs to products with 70% MAFs in a single step (single pot) within 20 min, enabling inexpensive, rapid genotyping with such as Sanger sequencers. Previously, we reported the broad outlines of our approach (*20*). In this paper, we expound yet unpublished experimental data that demonstrates PASEA's capabilities and its suitability for resource poor settings. Furthermore, we used PASEA to test 108 clinical tissue samples and 10 blood samples from cancer patients and compared PASEA with next generation sequencing (NGS) and amplification refractory mutation system (ARMS)-PCR.

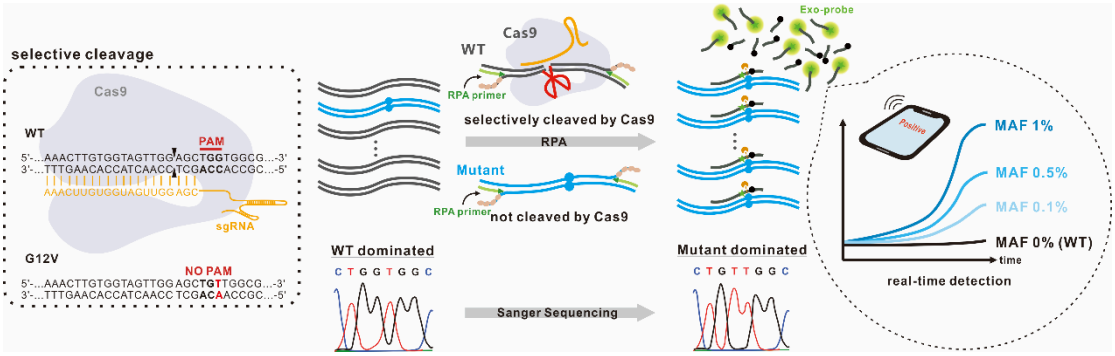


Figure 1. Principle of real-time PASEA. Directed by a single-stranded guided RNA (sgRNA), Cas9 selectively cleaves WT alleles with protospacer adjacent motif (PAM) site while sparing oncogenic mutation lacking PAM. The dotted frame illustrates WT and mutant allele sequences, the location of PAM site in the WT *KRAS* and its absence in *KRAS* G12. The sequences of the *KRAS* gene and mutants are from SOMIC online database (https://cancer.sanger.ac.uk/cosmic). Cleavage takes place between the third nucleotide and the fourth nucleotide upstream from the PAM site. While PASEA amplifies both WT and mutant alleles, the rate of amplification of mutant alleles far exceeds that of the WT, resulting in a product dominanted by

mutant alleles. Exo-probe indicates the number of amplicons and enables quantification in real time.

Results

PASEA relies on selective cleavage to obtain much greater amplification rates of mutant alleles than ofwild-type alleles. The contrast between the PASEA and RPA (in the absence of RNP) amplification rates of WT genomic DNA is striking (**Fig. 2a**). Within 10 minutes, RPA produced about 10^9 amplicons while PASEA produced less than 10^5 - four orders of magnitude less. When PASEA acts on a standard KRAS G12V (MAF 5%) sample (**Fig. 2a**), the numbers of both KRAS G12V and WT KRAS amplicons increase as time increases but the KRAS G12V amplifies at a much greater rate than WT KRAS. The number of amplicons of WT KRAS in the blend is nearly comparable to the number of amplicons when PASEA is applied to pure WT KRAS (blue bars). After ~3 min, the products of the standard sample (5% MAF) are dominated by the mutant allele (green bars).

To verify that the amplicons are, indeed, *KRAS*-G12V, we subjected the PASEA products of 60 ng gDNA, initially with 5% KRAS G12V to Sanger sequencing (**Figure 2b-i**) and estimated the MAFs with the Mutation Surveyor Software (https://softgenetics.com/mutationSurveyor.php) (**Figure 2b-ii**). Consistent with our qPCR results, the Sanger sequencing data shows that the MAF has increased from 5% to 70% after 3 min PASEA incubation and to nearly 100% after 5 min or longer incubation. PASEA provides highly efficient enrichment with RNP concentrations ranging from 0.1 μM to 1 μM RNP (**Figure S2**).

The minimal required incubation time depends on the sample's MAF (**Figure S3**). When MAF = 1%, incubation time of 10 min suffices to deplete the WT to undetectable level in the Sanger sensorgram. When MAF = 0.1%, 10 min PASEA enriches the MAF to approximately 60%, and 20 min PASEA makes mutant allele fraction nearly 100%. Hence, in our subsequent experiments, we used 20 min incubation time. Longer incubation times such as 30 min result in noisy sensorgrams possibly due to the presence of spurious amplicons.

To evaluate PASEA's sensitivity, we subjected a standard genomic DNA panel with 0%, 0.01%, 0.1%, 1%, and 5% *KRAS* G12V MAFs to 20 min PASEA incubation and examined the incubation products with Sanger sequencing (**Fig. 2c**). Sanger sequencing identified the presence of *KRAS* G12V in all the PASEA products of the samples with MAF = 0.1%, 1%, or 5% (N = 3). According to Sanger results, PASEA increased the MAFs to nearly 100% in the 5% (20-fold enrichment) and 1% (100-fold enrichment) samples; to 80% (800-fold enrichment) in the 0.1% samples. Seven out of 10 PASEA products of samples with initial MAF = 0.01% (N = 10), which equals to 2 copies mutant alleles in 60 ng total gDNA, were identified positive by Sanger, with the products' MAF averaging at 40% (4000-fold enrichment). The 3/10 negatives are

attributable to a sampling error. Statistically, it is reasonable to expect that mutant alleles were completely absent in the “false negative” samples. Importantly, PASEA did not produce any false positives.

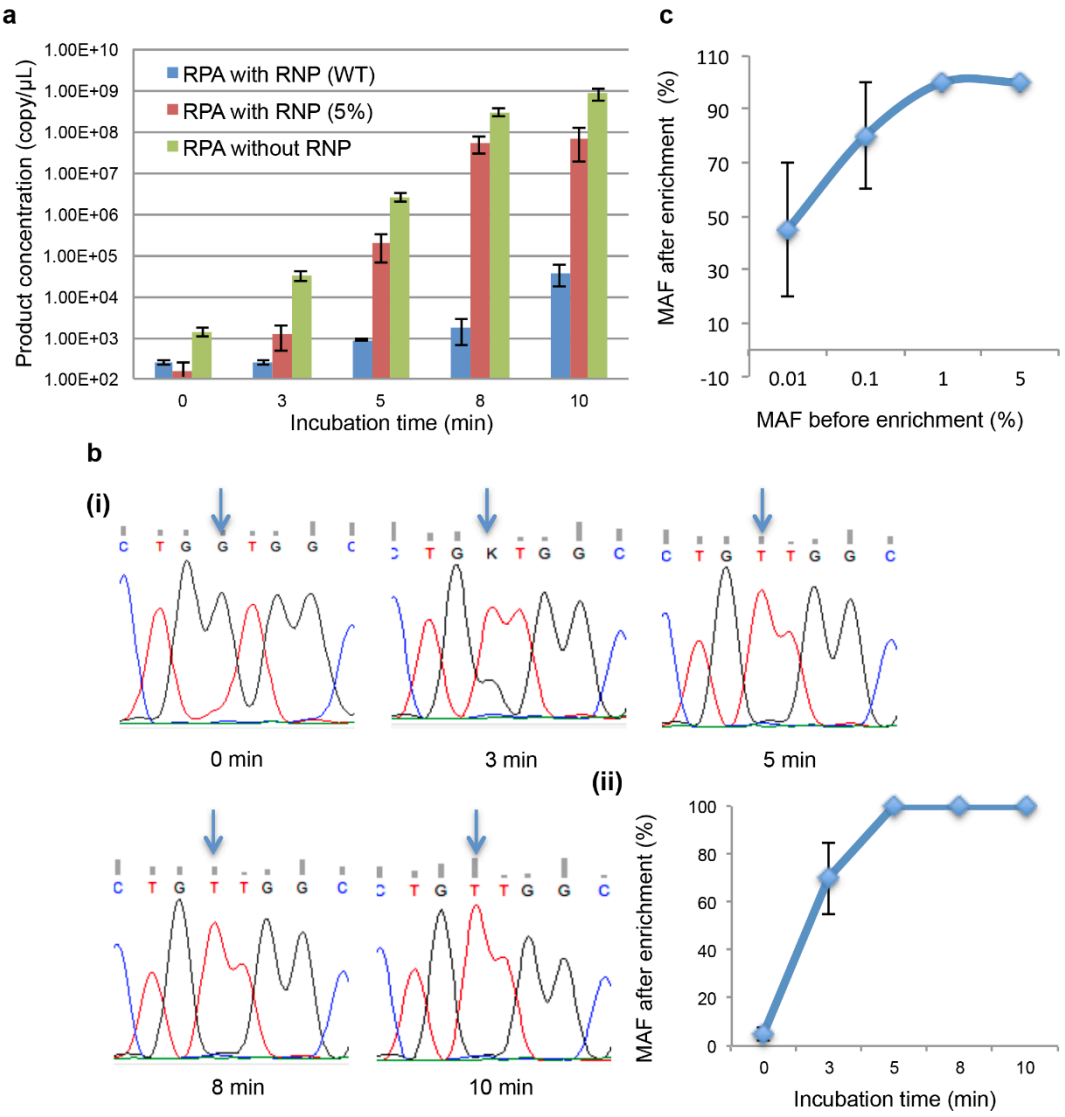


Figure 2. PASEA exponentially enriches MAF. (a) Pure WT-KRAS dsDNA and blends containing *KRAS* G12V dsDNA (MAF 5%) were subjected to PASEA (0.1 µM RNP) for various time spans along with a control group subjected to RPA without RNP. The number of amplicons is inferred from the threshold time of pre-calibrated qPCR **(Figure S1)**. N = 3. **(b)** Blends sample containing *KRAS* G12V dsDNA (MAF 5%) were subjected to PASEA (0.1 µM) for various time spans and then evaluated by Sanger sequencing: **(i)** Sequencing results of PASEA products; **(ii)** Estimated MAF as a function of PASEA incubation time. **(c)** Average MAF following PASEA incubation (enrichment). The MAF of each sample (total gDNA, 60 ng) before

enrichment is from 0.01% to 5%. The PASEA incubation time is 20 minutes with 0.1 µM RNP.

[Jinzaho: in (c), change the vertical axis to range from 0 to 100 – not 110]

Previous researchers (*9-10*) used Cas9-based cleaving assay (CUT) without concurrent amplification. How does PASEA compare with DASH? A 20 min incubation with DASH enabled detection of only MAF>1% in our hands (**Figure S4)** and **MAF >** 0.1% **under** optimal conditions (*10*) **(Table S2**) while PASEA facilitated detection of 0.01% with downstream Sanger sequencing. To improve DASH performance, researchers (6) carried out multiple DASH steps in a process dubbed CUT. [**Jinzhao, say something about CUT**.] PASEA has about two orders of magnitude better performance than DASH in terms of detectable MAF.

PASEA followed with Sanger sequencing and/or qPCR provides sensitive, two-stage detection of rare alleles. To meet the needs of resource poor settings, we designed a single stage, closed pot assay. Our real-time assay uses an Exo-RPA probe (**Figure 3a**) comprised of an abasic nucleotide analogue (tetrahydrofuran residue, THF) with a flanking dT-fluorophore at one side, a dT-quencher on the other, and a C3-spacer to block polymerase extension. When free in solution, the probe's fluorophore is quenched by the quencher located 2-5 bases away from the fluorophore. The THF provides a substrate for the Exonuclease III enzyme (included with the TwistAmp® exo kit) when the probe hybridizes with the amplicon to form a double-strand context. Enzymatic digestion of the THF separates the fluorophore from the quencher, resulting in fluorescent emission.

The short template sequence (~160 bp) (*21, 22*) of the cell-free DNA challenges probe design. It is difficult, if not impossible, to avoid an overlap between the probe's hybridization site and the sgRNA protospacer sequence. We evaluated several Exo-RPA probe sequences (**Figure S5**) and selected the best performer (**Figure 3b)** that hybridizes to the amplicon's middle region. Since the THF localizes to the *KRAS* G12V/D location, the Exo-RPA probe does not discriminate between the WT and mutant alleles, but instead reports on the total number of amplicons (**Figure 3c**).

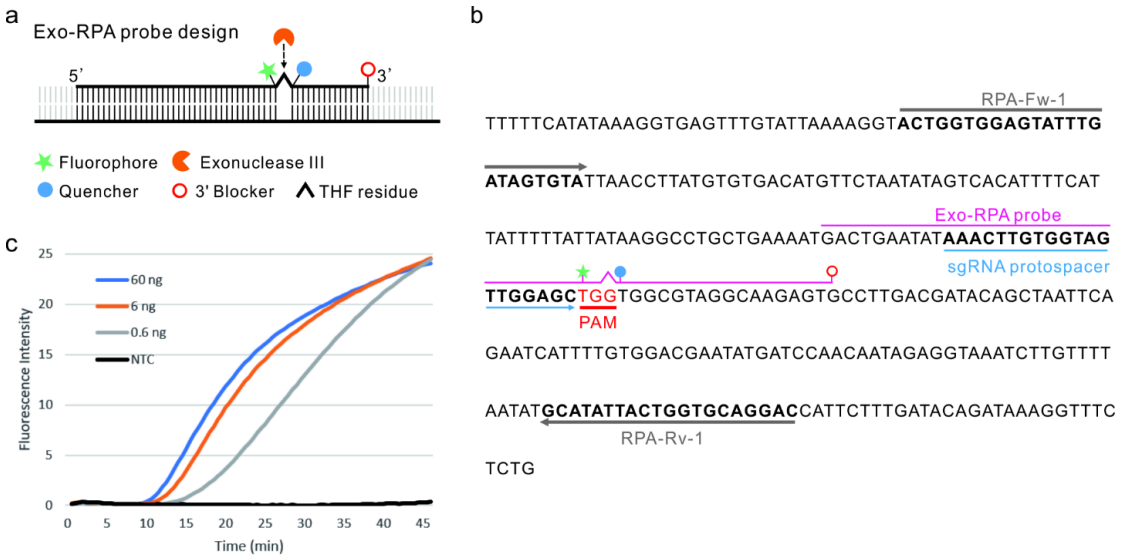


Figure 3. RPA probe enables real-time detection. (a) Schematics of the Exo-RPA probe. **(b)** *KRAS* gene with the positions of the primers, sgRNA protospacer, and Exo-RPA probe indicated. **(c)** Real-time RPA monitoring of serially diluted, wild-type alleles in the absence of RNP. Probe concentration: 240 nM.

Since the probe and the sgRNA have overlapping sequences, the probe affects Cas9 cleaving efficiency (**Figure S6**). At probe concentrations of 120 and 240 nM, real-time PASEA discriminates well between 5% *KRAS* G12V and WT alleles while at higher probe concentrations (e.g., 600 nM), there is little contrast between 5% *KRAS* G12V and WT alleles; likely because of probe interference with the sgRNA hybridization. In all our subsequent experiments, we used 240 nM probe concentration that provided a brighter signal than the 120 nM concentration, therefore reducing demands on the signal detector.

To address the interference between the probe and the RNP, we examined the effect of RNP concentration on the real-time amplification curve in the presence of 240 µM probe (**Figure S7**). Reaction mixes with 0.1 (a) and 0.08 µM (b) RNP discriminated well between samples of 0% and 5% *KRAS* G12V gDNA while 0.05 µM RNP (c) provided less satisfactory contrast. To further finetune our assay for cfDNA detection, we compared the effect of RNP concentrations on real-time PASEA acting on a standard *KRAS* G12D cfDNA control (0%, 0.1%, 1%, and 5% MAF). The assay with 0.08 µM RNP rendered 0.1% MAF (f) detectable while the same MAF was not detectable with 0.1 µM RNP.

At early disease stages, cell free mutant alleles are present in body fluids at very low concentrations. Cell-free RNA, predominantly comprised of small RNAs and

mRNAs, is present in peripheral blood, partly protected from degradation by its packaging into exosomes (*23, 24*). To increase the number of templates (biomarkers) for our assay, we target cell free, tumor derived fragments of both DNA and RNA (*25*). Since the cfDNA (~160 nt) and the *KRAS* exon 2 (122 nt) share a short common sequence (125 nt, **Figure S8a**), we designed and tested various primers for short templates to concurrently amplify both ctDNA and complementary DNA (**Figure S8** and **Table S1**). We then selected the primer set that provides the shortest threshold time. We targeted *KRAS* G12D, which like *KRAS* G12V, lacks the NGG PAM. Real-time PASEA with primers RPA-ct-Fw-1/ RPA-ct-Rv-2 detects *KRAS* G12D in 20 ng of standard cfDNA in the absence of reverse transcriptase (RT) (**Figure 4a, d**); in 400 ng of purified mRNA (~120,000 copies) in the presence of RT (**Figure 4b, e**); and in a mixture of 10 ng cfDNA and 200 ng mRNA in the presence of RT (**Figure 4c, f**) at various MAFs. The quantities of nucleic acid used in our experiments are comparable with those in patient samples. As expected, the threshold time (defined as the time until signal intensity exceeds 10% of signal saturation level) increases as the MAF decreases (**Figure 4d-f**). PASEA readily detects 0.1% MAF cfDNA (~6 copies in 20 ng, **Figure 4a, d**), 0.05% G12D mRNA (~60 copies in 400 ng, **Figure 4b, e**) and 0.05% mixture of cfDNA and mRNA (**Figure 4c, f**).

Interestingly, the fluorescence intensity of the amplification curve's plateau increases as the MAF increases. The fluorescence intensity of the plateau at 45 min (F_{45}) correlates well with the MAF **(Figure 4g, h)** and with the threshold time (**Figure S9**), providing yet another metric for semi-quantitative estimation of the MAF.

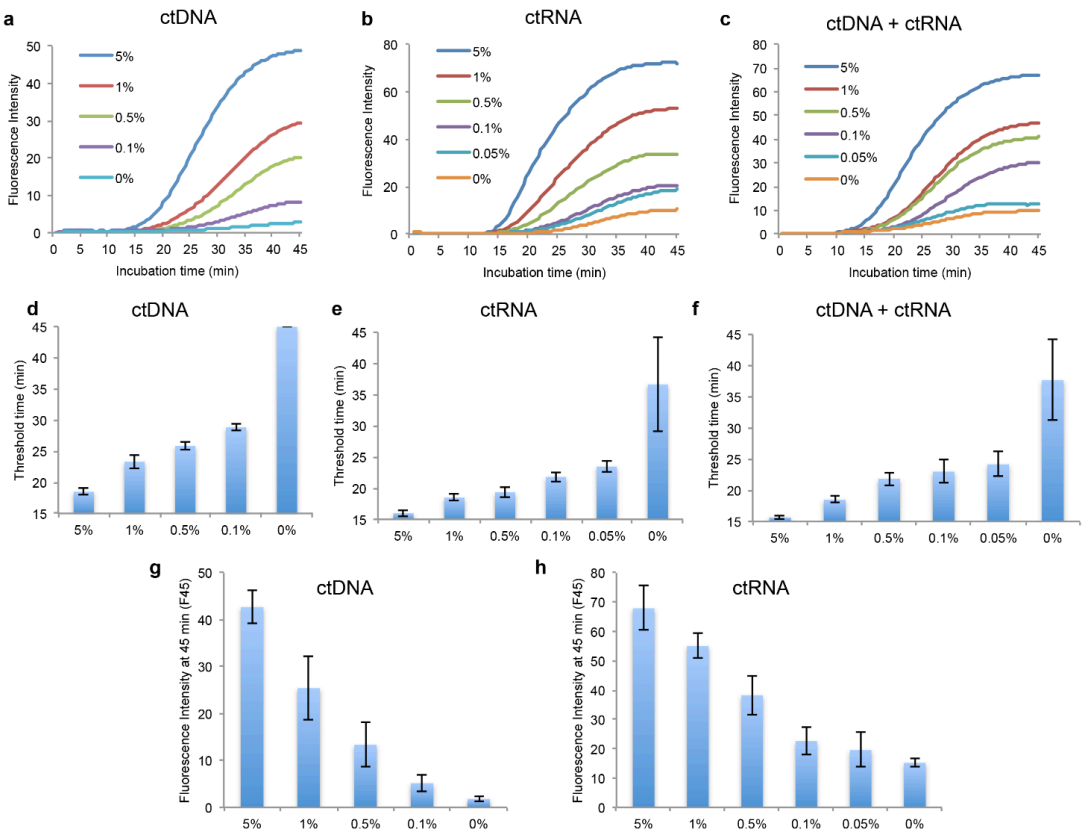
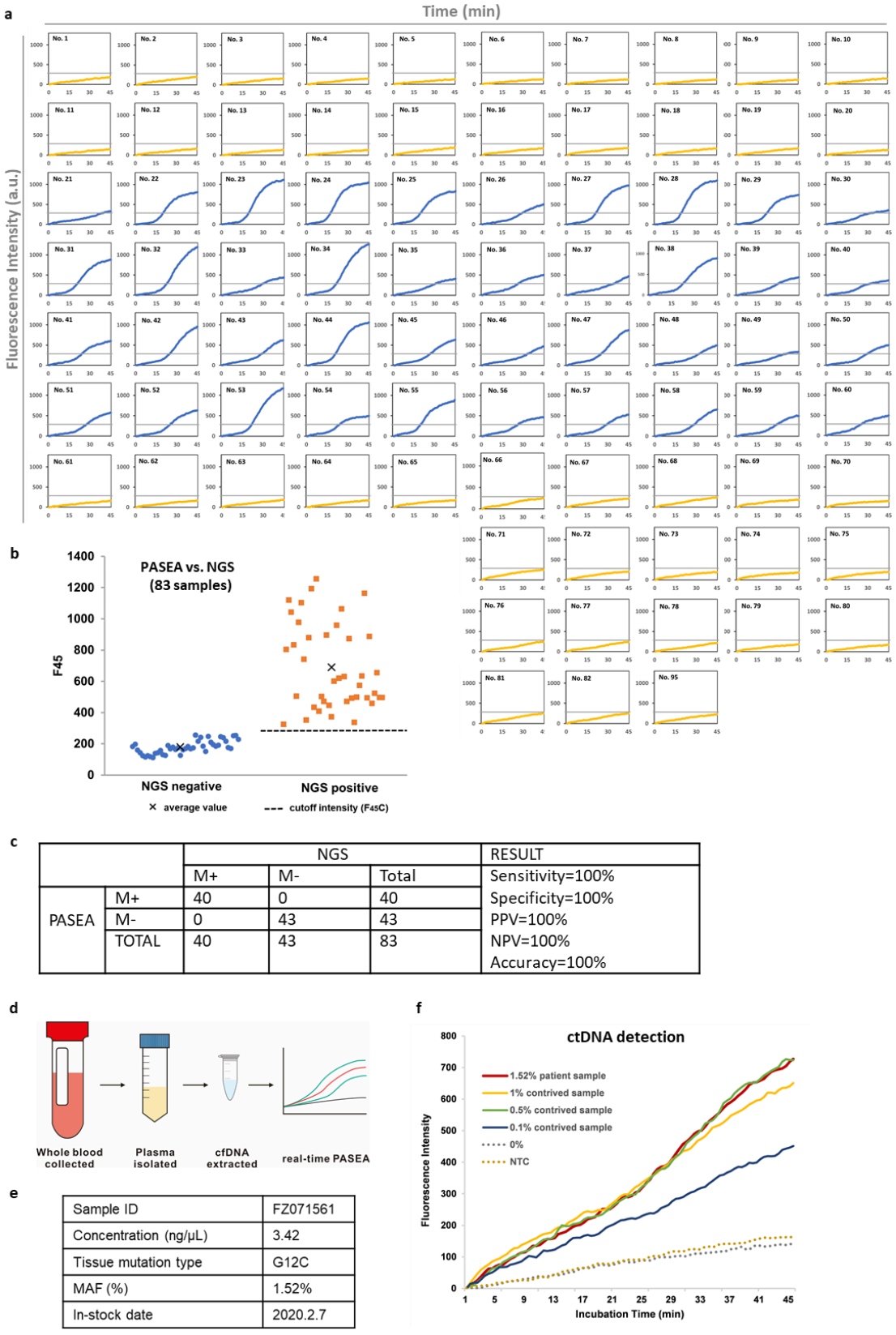


Figure 4. Real-time PASEA performance. Amplification curves of cfDNA (a), RNA (b), mixture of cfDNA and RNA (c) with different MAF. Threshold time as a function of initial MAF when detecting cfDNA (d); RNA (e); and a mixture of cfDNA and RNA (f). Fluorescence intensity at 45 min as a function of MAF when detecting cfDNA (g) and RNA (h).

Is PASEA applicable to clinical samples? We collected 108 tissue samples from colorectal cancer (62/108), pancreatic cancer (1/108), and lung cancer patients (45/108) by either resection or biopsy at the Cancer Hospital of the Chinese Academy of Medical Sciences (Beijing, China). Genomic DNA (gDNA) was extracted from these tissue samples and tested with clinical NGS (83 samples) and ARMS-PCR (97 samples). 40 samples were positive to *KRAS* mutations (G12V, D, S, C, and R) with MAF ranging from 1% to 39%. 68 samples were negative for *KRAS* mutations (**Table S3**). The extracted gDNAs were diluted to 10 ng/μL and then tested with our real-time PASEA. Real-time PASEA was deemed positive when F_{45} exceeded the cutoff ($F_{45}C$), defined as the average F_{45} plus 3 SD (95% confidence level) for standard WT gDNA at 20 ng/μL, which is greater than the DNA concentration in our clinical samples (~10 ng/μL) (**Table S3**). Real-time PASEA correctly identified 40/40 samples as positive and 68/68 samples as negative,

exhibiting 100% sensitivity, specificity, positive predictive value, negative predictive value, and concordance with NGS **(Figure 5abc)** and/or ARMS-PCR **(Figure S10)** genotyping for *KRAS* G12 mutations. In all cases, the emission intensity (F_{45}) correlated well with the threshold time (**Figure S11**).

Plasma samples were collected from 10 lung cancer patients and banked at -80°C (**Figure 5d)**. NGS identified only one sample as positive for *KRAS* mutation (*KRAS* G12C, MAF = 1.52%, **Figure 5e**). To compensate for the positive clinical samples' scarcity, we spiked Horizon standard cfDNA (WT) controls into the positive sample to form contrived samples with MAFs of 1%, 0.5%, and 0.1%. The 10 patient samples and the 3 contrived samples were subjected to real-time PASEA. The amplification curves of positive samples clearly differentiated from the curve of the WT control (0%) (**Figure 5f**), indicating that real-time PASEA detects ctDNA in blood samples with high sensitivity.



		NGS			RESULT
		M+	M-	Total	Sensitivity=100%
PASEA	M+	40	0	40	Specificity=100%
	M-	0	43	43	PPV=100%
	TOTAL	40	43	83	NPV=100%
					Accuracy=100%

Sample ID	FZ071561
Concentration (ng/μL)	3.42
Tissue mutation type	G12C
MAF (%)	1.52%
In-stock date	2020.2.7

Figure 5. Real-time PASEA successfully detects gDNA and cfDNA in clinical samples. (a) Real-time PASEA amplification curves of 83 samples that were also tested with NGS. The amplification curves of the 40 positive samples (No. 21-60) are

in blue and the 43 negative samples (No. 1-20, 61-82, 95) are in yellow. **(b, c)** F_{45} values of patient samples (83) compared with tissue NGS genotyping ("gold standard"). The dashed horizontal line and the symbol "x" denote, respectively, the F_{45} cutoff value and the average F_{45} value. **(d)** Workflow of blood testing with real-time PASEA. **(e)** Details of the NGS *KRAS* positive blood sample. **(f)** Real-time PASEA amplification curves of cfDNA samples with various mutant allele fractions.

Point-of-Care analysis of nucleic acids for cancer profiling and infectious disease diagnosis is highly desirable (*26, 27*). Real-time PASEA is relatively simple to carry out, does not require strict temperature control, and can be used at the point of care (*28*). We implemented real-time PASEA in our multifunctional isothermal amplification microfluidic (MIAR) chip (*28*) that extracts and concentrates nucleic acids from a sample and mates with a homemade portable isothermal amplification processor (**Figure 6a**) (*29*). We carried out real-time PASEA on samples comprised of various concentrations of standard *KRAS* WT cfDNA and G12D ctDNA spiked in PBS. Samples with MAF of 0.5% of *KRAS* G12D were readily detected within 40 min (**Figure 6b, c**). The lengthy incubation time is partially due to the slow temperature ramp of our homemade incubator. Real-time PASEA exhibited detection limit of about 87 copies when operating with samples comprised of 60 ng cfDNA. This can likely be improved further with additional optimization.

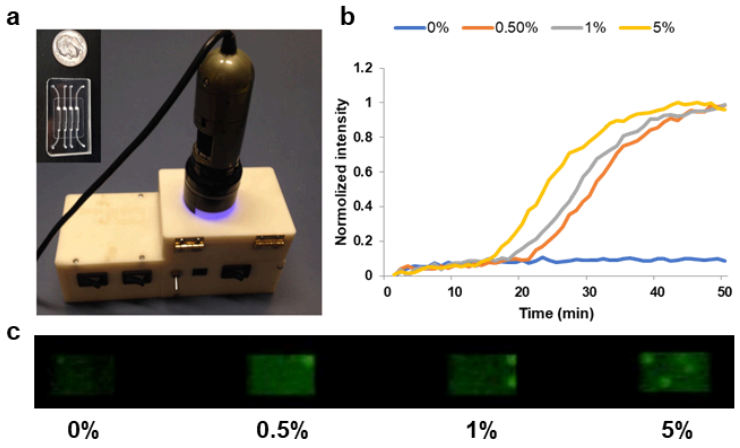


Figure 6. On Chip, Real-time PASEA. (a) Custom-made, multifunctional microfluidic (MIAR) chip (inset) (*28*) and portable heating & imaging platform for isothermal amplification and fluorescence detection (*29*); **(b)** Real-time amplification curve of standard cfDNA with various MAF as detected with portable USB camera; **(c)** Fluorescence emission images detected with a camera from the microfluidic chip.

Discussion and Conclusions

Researchers have identified various oncogenic mutations responsible for the initiation and maintenance of cancer and the mechanisms of resistance to targeted therapeutics (*30*) Methods for cost effective, non-invasive cancer genotyping are needed to enable targeted therapies. An attractive genotyping method relies on identifying tumor derived, aberrant nucleic acids in body fluids (liquid biopsy). However, the identification of tumor-associated nucleic acid fragments in body fluids is challenged by their low abundance and sequence homology with the vast background of nucleic acids from healthy cells.

Among the current programmable endonuclease-based mutant allele enrichment methods (*10-13*), the CRISPR-mediated, ultrasensitive detection of target DNA by PCR (CUT-PCR) (*11*) is a promising assay to enrich mutant alleles' fraction by 2-3 rounds of selective depletion of WT and subsequent PCR. CUT-PCR has successfully increased MAF by 27-fold in most samples that contained over 0.1% MAF before cleavage (*11*). In combination with deep sequencing, CUT-PCR enables detection of 0.01% mutant alleles.

There are, however, a few factors that limit the enrichment level achievable with assays such as CUT-PCR. Although Cas 9 preferentially cleaves WT alleles, it also cleaves, albeit, to a lesser degree, off-target mutant alleles. While off-target and target cleavage rates depend on the guide RNA design, Cas9 variant, and assay conditions, samples with low MAF (e.g., 0.01%) contain just a handful of molecules of clinical interest and any loss of critical biomarkers compromises assay's sensitivity. Furthermore, not all guide–protein-target triplexes are productive. By some estimates, fewer than 90% triplexes are cleaved (*31*). Since the dissociation rate of the triplex is very slow (*19*), a significant fraction of WT DNA is protected from cleavage but amenable to PCR amplification. For example, if an assay cleaves only 90% of the wild-type alleles, the MAF can be enriched by less than 10-fold in a single step.

Herein, we propose a simple remedy that combines cleavage with concurrent polymerase amplification to overcome the shortcomings of the cleavage only assays. Our assay amplifies concurrently mutant alleles of interest and WT alleles in the presence of relatively high concentration of guided endonuclease. While the copy numbers of both the WT and mutant allele increase with time, the latter do so at much greater rate, alleviating any concerns of losing valuable biomarkers. Compared to restriction site mutation assay (*8, 9*), PASEA has less limitation of target sequence.

PASEA's very high enrichment capability enables libraries preparation for rapid, low-cost sequencers such as Sanger. It also offers the opportunity for real-time detection of mutant alleles in a closed pot without a sequencer, eliminating the need to open amplicon-rich tubes and risking the contamination of the workspace. Clinical evaluation of real-time PASEA exhibits good concordance with NGS and ARMS-PCR genotyping for *KRAS* G12 mutations when testing tissue samples. Real-time PASEA has successfully identified the presence of mutant alleles in all positive samples and yielded no false positives in liquid biopsy. Real-time PASEA enables unprecedented level of enrichment and detection with relatively simple instruments, providing effective means for cancer screening and targeted therapies in low resource settings.

Although real-time PASEA's reliance on Cas9 limits its use to sequences in which the PAM motif is present in the wild-type allele and absent in the mutant allele, the "NGG" PAM site is shorter and appears more frequently than restriction endonuclease recognition sites giving PASEA and advantage over restriction-based assays. Moreover, Lee et al (*11*) estimate, that with the use of various orthologonal CRISPR endonucleases such as SpCas9 and FnCpf1, Cas9-like proteins can target about 80% of known cancer-linked substitution mutations registered in the Catalogue of Somatic Mutations in the Cancer (COSMIC) database. Furthermore, our two-stage PASEA and our real-time PASEA can be extended to use other families of endonuclease such as Argonautes (*32*) that do not require the presence of PAM and therefore are more versatile. PASEA's simplicity makes it amenable for use in resource-limited settings with potential significant impact on global health and in applications other than cancer.

Appendix A. Supporting Information

Material and Methods, and Supplementary data related to this manuscript can be found in Supporting Information file.

Author Contributions

J.S. and H.H.B. conceived the project and designed the experiments. J.S. and J.C. carried out the experiments. T.Q. and J.Y. processed and quantified patient samples. J.C., T.Q., H.H.B., J.Y., and J.S. analyzed the data and wrote the manuscript. M.G.M., Y.F., Z.S., D.J.Y., Q.Z., and Y.Q. assisted with experiments and biostatistics. All authors read and approved the manuscript.

Conflict of interest

The University of Pennsylvania has filed for a patent for the method described in this paper with JS, JC, and HHB listed as co-inventors.

Acknowledgements

This work was supported by China Scholarship Council, NIH grant K01 1K01TW011190-01A1 to the University of Pennsylvania, NIH grant R21 CA228614-01A1 to the University of Pennsylvania, and Beijing Hope Run Special Fund from the Cancer Foundation of China (LC2019L04, LC2020A36).

References

[1] M. Guha, E. Castellanos-Rizaldos, P. Liu, H. Mamon, G. Makrigiorgos, Nucleic. Acids. Res. **41** (2013) e50.

[2] A. Aalipour, J. C. Dudley, S. Park, S. Murty, J. J. Chabon, E. A. Boyle, M. Diehn, S. S. Gambhir, Clin. Chem. **64** (2018) 307-316.

[3] C. R. Newton, A. Graham, L. E. Heptinstall, S. J. Powell, C. Summers, N. Kalsheker, J. C. Smith, M. A. F Nucleic. Acids. Res. **7** (1989) 2503-2516.

[4] L. R. Wu, S. X. Chen, Y. Wu, A. Patel, A., D. Y. Zhang, Nat. Biomed. Eng. **1** (2017) 714-723.

[5] D. Y. Vargas, S. Marras, S. Tyagi, F. R. Kramer, J. Mol. Diagn. **20** (2018) 415-427.

[6] S. Narumi, K. Matsuo, T. Ishii, Y. Tanahashi, T. Hasegawa, Plos. One. **8** (2013) e60525.

[7] D. Yang, Y. Sun, F. Chang, H. Tian, Z. Li, Chinese Chem. Lett. **31** (2020) 1095-1098.

[8] R. Ward, N. Hawkins, R. O'Grady, C. Sheehan, T. O'Connor, H. Impey, N. Roberts, C. Fuery, A. Todd, Am. J. Pathol. **153** (1998) 373-379.

[9] H. Pincas, M. R. Pingle, J. Huang, K. Lao, P. B. Paty, A. M. Friedman, F. Barany, Nucleic. Acids. Res. **32** (2004) e148.

[10] W. Gu, E. D. Crawford, B. D. O'Donovan, M. R. Wilson, E. D. Chow, H. Retallack, J. L. Derisi, Genome. Biol. **17** (2016) 41.

[11] S. Lee, J. Yu, G.-H. Hwang, S. Kim, H. Kim, S. Ye, K. Kim, J. Park, D. Park, Y.-K. Cho, J.-S. Kim, S. Bae, Oncogene **36** (2017) 6823–6829.

[12] J. Song, J. W. Hegge, M. G. Mauk, J. Chen, J. E. Till, N. Bhagwat, L. T. Azink, J. Peng, M. Sen, J. Mays, E. Carpenter, J. V. Der Oost, H. H. Bau, Nucleic. Acids. Res. **48** (2019) e19.

[13] H. H. Bau, J. Song, M. G. Mauk, J. V. Der Oost, J. W. Hegge, US Pat., US2021010064A1, (2019)

[14] Q. Liu, X. Guo, G. Xun, Z. Li, Y. Chong, L. Yang, H. Wang, F. Zhang, S. Luo, Y. Feng, Nucleic. Acids. Res. **49** (2021) e75.

[15] R. Sorek, V. Kunin, P. Hugenholtz, Nat. Rev. Microbiol. **6** (2008) 181-186.

[16] L. A. Marraffini, E. J. Sontheimer, Nat. Rev. Genet. **11** (2010) 181-190.

[17] D. C. Swarts, M. M. Jore, E. R. Westra, Y. Zhu, J. H. Janssen, A. P. Snijders, Y. Wang, D. J. Patel, J. Berenguer, S. J. J. Brouns, Nature **507** (2014) 258-261.

[18] D. C. Swarts, J. W. Hegge, H. Ismael, S. Masami, M. A. Ellis, D. Justin, R. M. Terns, M. P. Terns, J. V. Der Oost, Nucleic. Acids. Res. **43** (2015) 5120-5129.

[19] S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, Nature **507** (2014) 62-67.

[20] J. Chen, T. Qiu, M. G. Mauk, Y. Fan, Y. Jiang, J. Ying, q. Zhou, Y. QIao, H. H. Bau, J. Song, Clin. Chem. **in press** (2021) https://doi.org/10.1093/clinchem/hvab1163.

[21] H. R. Underhill, J. O. Kitzman, H. Sabine, N. C. Welker, D. Riza, D. N. Baker, K. M. Gligorich, R. C. Rostomily, M. P. Bronner, S. Jay, PLoS Genet. **12** (2016) e1006162.

[22] Sato, C. Nakashima, T. Abe, J. Kato, N. Sueoka-Aragane, Oncotarget **9** (2018) 31904-31914.

[23] E. Heitzer, I. S. Haque, C. E. S. Roberts, M. R. Speicher, Nat. Rev. Genet. **20** (2018) 71-88.

[24] S. Inamdar, R. Nitiyanandan, K. Rege, Bioeng. Transl. Med. **2** (2017) 70-80.

[25] K. Krug, D. Enderle, C. Karlovich, T. Priewasser, S. Bentink, A. Spiel, K. Brinkmann, J. Emenegger, D. G. Grimm, E. Castellanos-Rizaldos, Ann. Oncol. **29** (2017) 700-706.

[26] Y. Xu, T. Wang, Z. Chen, L. Jin, Z. Wu, J. Yan, X. Zhao, L. Cai, Y. Deng, Y. Guo, S. Li, N. He, Chinese Chem. Lett. **in press** (2021) https://doi.org/10.1016/j.cclet.2021.1006.1025.

[27] G. Xiang, W. Zhang, N. Li, Q. Pu, J. Lin, Chinese Chem. Lett. **in press**, (2021) https://doi.org/10.1016/j.cclet.2021.1008.1073.

[28] J. Song, V. Pandian, M. G. Mauk, B. H. H., S. Cherry, L. C. Tisi, C. Liu, Anal. Chem. **90** (2018) 4823-4831.

[29] K. Kadimisetty, J. Song, A. M. Doto, Y. Hwang, J. Peng, M. G. Mauk, F. D. Bushman, R. Gross, J. N. Jarvis, C. Liu, Biosens. Bioelectron. **109** (2018) 156-163.

[30] G. R. Oxnard, C. P. Paweletz, Y. Kuang, S. L. Mach, A. O\"Connell, M. M. Messineo, J. J. Luke, M. Butaney, P. Kirschmeier, D. M. Jackman, P. A. Janne, Clin. Cancer. Res. **20** (2014) 1698-1705.

[31] M. Yang, S. Peng, R. Sun, J. Lin, N. Wang, C. Chen, Cell Reports **22** (2018) 372-382.

[32] J. W. Hegge, D. C. Swarts, J. V. Der Oost, Nat. Rev. Microbiol. **16** (2018) 5-11.

Supplementary Information

Programmable Endonuclease Combined with Isothermal Polymerase Amplification to Selectively Enrich for Rare Mutant Allele Fractions

Junman Chen[1,2,3†], Tian Qiu[4†], Michael G. Mauk[3], Zheng Su[5], Yaguang Fan[6], Dennis J. Yuan[2], Qinghua Zhou[6], Youlin Qiao[5], Haim H. Bau[3], Jianming Ying[4*], Jinzhao Song[2,3*]

1. Key Laboratory of Clinical Laboratory Diagnostics (Ministry of Education), College of Laboratory Medicine, Chongqing Medical University, Chongqing, 400016, China
2. The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, Zhejiang 310022, China
3. Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia PA, 19104, USA
4. Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, China.
5. Center for Global Health, School of Population Medicine and Public Health, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100730, China
6. Tianjin Key Laboratory of Lung Cancer Metastasis and Tumor Microenvironment, Tianjin Lung Cancer Institute, Tianjin Medical University General Hospital, Tianjin, 300052, China.

* To whom correspondence should be addressed. Tel: 215-898-9380; Fax: 215-573-6334; Email: songjinz@seas.upenn.edu, Correspondence may also be addressed to jmying@cicams.ac.cn.

† The first two authors contributed equally.

Material and Methods

Samples

Standard genomic DNA (gDNA), cell-free DNA (cfDNA), and RNA samples. Standard gDNA and cfDNA were purchased from Horizon Discovery. Total RNA was extracted with RNeasy® mini kit (Qiagen, Valencia, CA, USA) per manufacturer's protocol from Human cancer cell lines U87-MG (WT *KRAS* RNA) and ASPC1 (*KRAS* G12D RNA).

Patient tissue DNA samples. Tissue samples from 62 colorectal cancer patients, 45 lung cancer patients, and 1 pancreatic cancer patient (**Table S3**) were collected at the Cancer Hospital, Chinese Academy of Medical Sciences by either resection or biopsy under the IRB-approved protocol (20/383-2579). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Genomic DNA (gDNA) was extracted with DNeasy® Blood & Tissue Kit (Qiagen, Valencia, CA, USA). Subsequently, the extracted gDNA was quantified with NanoDrop™ spectrophotometer and diluted to 10 ng/μL.

Patient cfDNA samples. Ten lung cancer patient blood samples were obtained at the Cancer Hospital, Chinese Academy of Medical Sciences under the IRB-approved protocol (20/383-2579). cfDNA was extracted with QIAamp® Circulating Nucleic Acid kit (Qiagen, Valencia, CA, USA). Subsequently, the extracted cfDNA was quantified with NanoDrop™ spectrophotometer and qualified with clinical NGS.

Preparation of Cas9-sgRNA ribonucleoprotein complex

S.p. Cas9 Nuclease V3 (Cas9) and sgRNA (protospacer sequence: 5'-AAACTTGTGGTAGTTGGAGC-3') were purchased from Integrated DNA Technologies (Coralville, US). The sgRNA and Cas9 were mixed in Working buffer (30 mM HEPES pH 7.5, KCl 150 mM) in equimolar amounts and incubated at room temperature for 10 min to form the Cas9-sgRNA ribonucleoprotein (RNP) complex.

Programmable enzyme based exponential enrichment assay

Exponential enrichment was carried out in a 10 μL rehydrated (1x) RPA reaction mix (Twistamp® Basic kit) containing extra 0.1 μM RNP, 0.5 μM RPA primers (5'-ACTGGTGGAGTATTTGATAGTGTA-3', 5'- GTCCTGCACCAGTAATATGC-3'), 14 mM Magnesium Acetate (MgOAc), and 60 ng standard genomic DNA (Horizon Discovery) with various MAF (0%, 0.01%, 0.1, 1%, 5%). The reaction mixes were incubated at 37°C for 3~20 minutes and then at 95°C for 10 min to stop the reaction by denaturing enzymes. Subsequently, 1 μL RNase A (10 mg/mL) was added to digest sgRNA with

incubation at room temperature for 10 min. And 1 μL Proteinase K (20 mg/mL) was added to digest the RNase A and Cas9 endonuclease with incubation at 56°C for 30 min. Then, the Proteinase K was deactivated at 95°C for 10 min for the following analysis.

Programmable enzyme based linear enrichment assay

For comparison, linear enrichment assay was carried out with Twistamp® Basic kit (TwistDx, Cambridge, UK) without primers. The 10 μL reaction volume contains well-mixed 0.1 μM RNPs, 14 mM Magnesium Acetate (MgOAc), and 60 ng standard genomic DNA with various MAF (0%, 0.01%, 0.1, 1%, 5%) in rehydrated (1×) Twistamp® Basic RPA reaction mix. The reaction mixes were incubated at 37°C for 20 minutes and then stopped. sgRNA and enzymes were removed or deactivated as described above.

Quantitative PCR of enrichment product

For relative quantitation of the CRISPR-RPA products, we prepared a standard calibration curve with qPCR (DiaCarta, Inc.) following manufacturer's instructions (DiaCarta, Inc.) (**Figure S1)**. Enrichment products were quantified using the same qPCR protocol. The 10 μL reaction volume contained 2 μL of the 10^4-10^6 fold diluted, enriched products, and 1 μL PCR primer/probe in 1× PCR Master Mix. qPCR was carried out with a BioRad Thermal Cycler (BioRad, Model CFX96) with a temperature profile of 95°C for 5 minutes, followed by 45 cycles of amplification (95°C for 20 seconds, 70°C for 40 seconds, 60°C for 30 seconds, and 72°C for 30 seconds).

Sanger Sequencing

The qPCR products were inspected for quality and yield by running 5 μl in 2.2% agarose Lonza FlashGel DNA Cassette. After treating with Exo-CIP Rapid PCR Cleanup Kit to remove any remaining primers, probes, and dNTPs, the products were processed for Sanger sequencing at the Penn Genomic Analysis Core with the reverse *KRAS* PCR primer 5'-TTGGATCATATTCGTCC-3'. Mutation Surveyor Software (https://softgenetics.com/mutationSurveyor.php) was used for mutation quantification.

Real-time PASEA for detection of rare mutant alleles

We firstly tested our real-time detection assay by using gDNAs as targets. Each experiment was carried out in a 10 μL rehydrated (1x) RPA reaction mix (TwistAmp® Exo kit), containing 60 ng standard gDNA, 420 nM each of RPA primers (RPA-g-Fw and RPA-g-Rv, **Table S1**), 14 mM Magnesium Acetate (MgOAc), 0.1 μM RNP, and 240 μM Exo-RPA Probe (**Table S1**). The reaction volume was prepared on ice. After vortex, the reaction mix was placed into a BioRad Thermal Cycler (BioRad, Model

CFX96) with plate-read each 30 sec for isothermal amplification at 37°C. For real time detection of clinical tissue samples, 2 µL of extracted gDNA (10 ng/µL) was added into the PASEA reaction mixture.

RPA primer development for detection of both cell-free DNA and RNA

For real-time detection of both *KRAS* DNA and *KRAS* RNA mutant alleles, we designed various primers (**Table S1**) based on a shared gDNA and cDNA sequence (125 nt) containing the *KRAS* exon 2 that harbors G12 mutations. Since the average size of cfDNA fragments is ~160 bp (*1, 2*), we targeted a short template. We evaluated the performance of our primers with the TwistAmp® Exo kit's standard protocol and various amounts (4 ng, 400 pg, 40 pg, 4 pg, and 0 pg) of standard cfDNA (Horizon Discovery) as the targets.

The primer set with the lowest limit of detection was selected for use in our experiments with the above-described protocol and reduced RNP concentration of 0.08 µM. 20 ng standard cfDNA (Horizon Discovery), 400 ng RNA, or the mixture of 10 ng standard cfDNA and 200 ng RNA with MAF ranging from 0% to 5% were added as templates to the rehydrated (1×) Twistamp® Exo RPA buffer. When testing RNA, 0.2 µL AMV Reverse Transcriptase (10 U/µL) was also included. When testing clinical samples, 2 µL diluted gDNA (10 ng/µL) from tissue sample or 3 µL extracted cfDNA from blood sample were added into real-time PASEA reaction mixture.

Point-of-care detection of rare mutant alleles on a microfluidic chip

We used our custom-made microfluidic chips with four independent multifunctional, isothermal amplification reactors (*3*). For each test, 600-µL mixture of 200-µL plasma, 200-µL Qiagen AL buffer, and 200-µL ethanol was filtered through the nucleic acid isolation membrane of one of the amplification chambers. The nucleic acids bound to the membrane. Subsequent to the sample introduction, 150 µL of Qiagen wash buffer 1 (AW1) was injected into the reactor to remove amplification inhibitors. Then, the silica membrane was washed with 150 µL of Qiagen™ wash buffer 2 (AW2), followed by air-drying for 30 seconds. Next, 25 µL of real-time PASEA reaction mixture prepared as described above was injected into each reactor. The inlet and outlet ports were then sealed with transparent tape. The chip was placed in a portable custom-made device (*4*) that houses a heating system and USB-based microscope (**Figure 6a**) for real time fluorescence excitation and emission imaging.

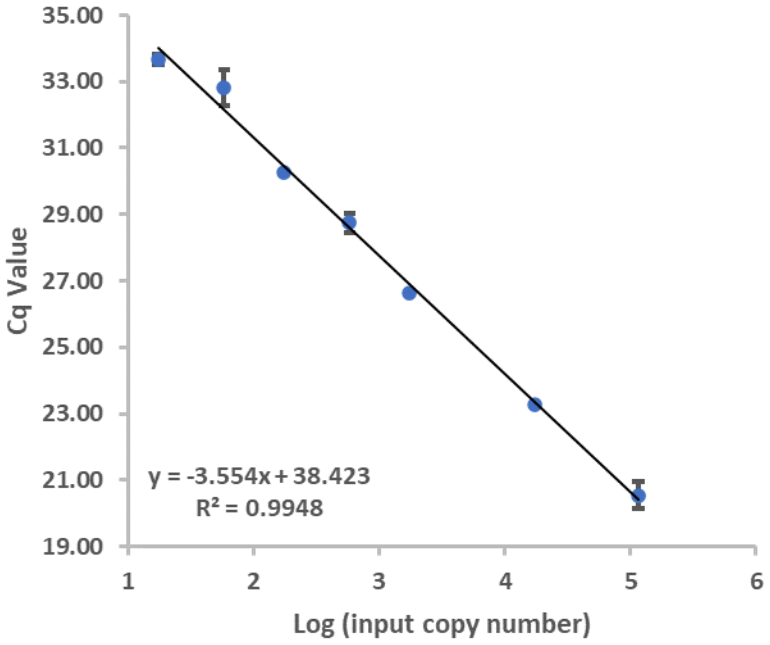


Figure S1: qPCR calibration curve for *KRAS* alleles. The threshold cycle Cq is depicted as a function of logarithmic copy number of templates (N=3). **[Jinzhao: please reduce the number of significant digits in the correlation to 38.4-3.5 x. You are not measuring time with ms precision.]**

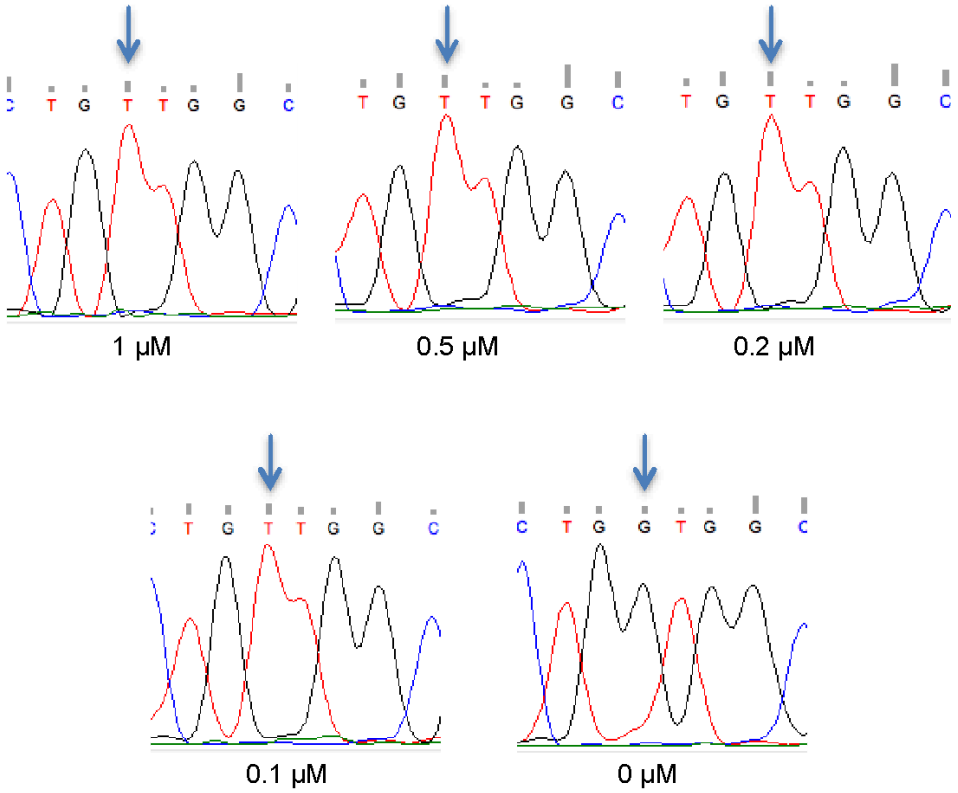


Figure S2. Enrichment efficiency as a function of RNP concentration. Sanger sequencing data. 10 min PASEA incubation time**.** 60 ng genomic DNA initially containing 5% of *KRAS* G12V. With RNP concentration of 0.1 μM or larger, PASEA converted 5% mutant allele in the sample into the dominating allele in the product.

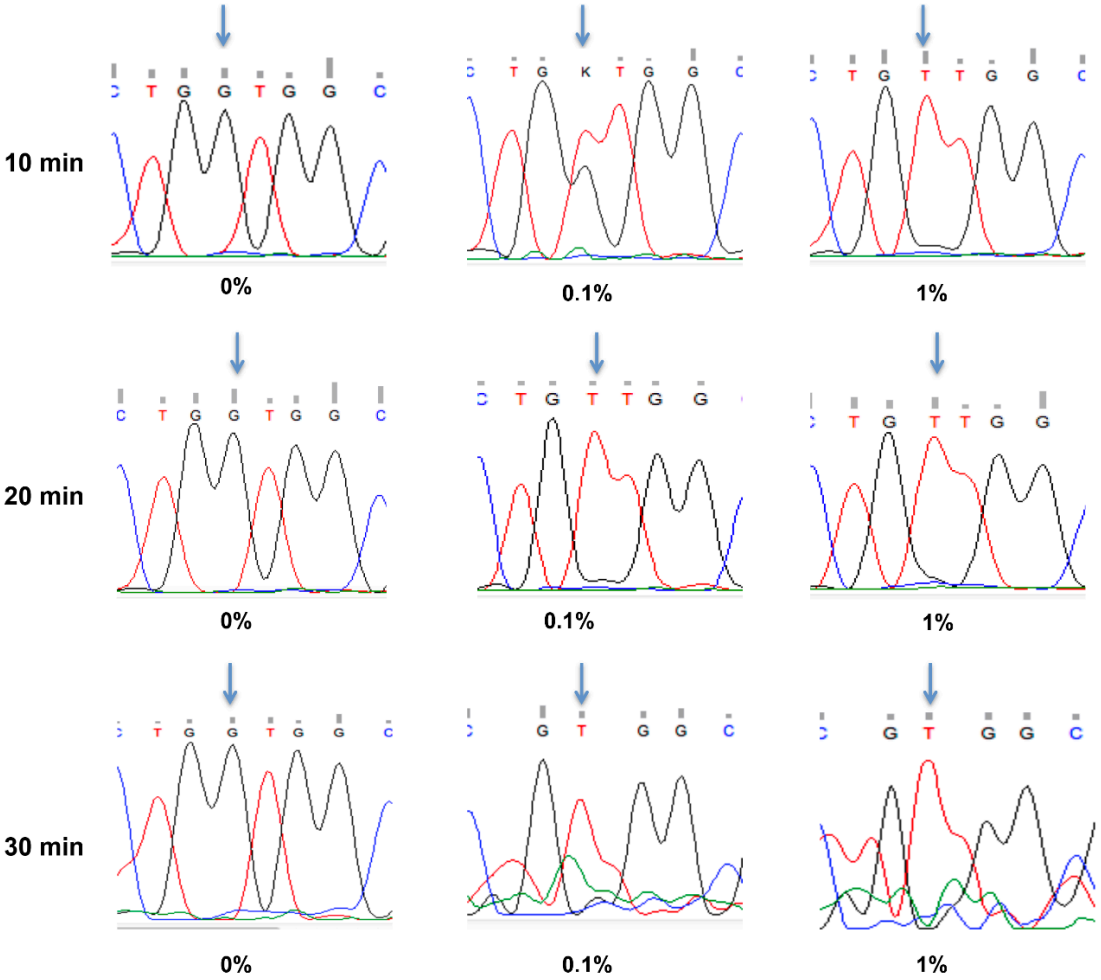


Figure S3. PASEA's performance as a function of mutant allele fraction (MAF) with incubation times of 10, 20, and 30 minutes. Sanger sequencing data. The total gDNA in each sample is 60ng.

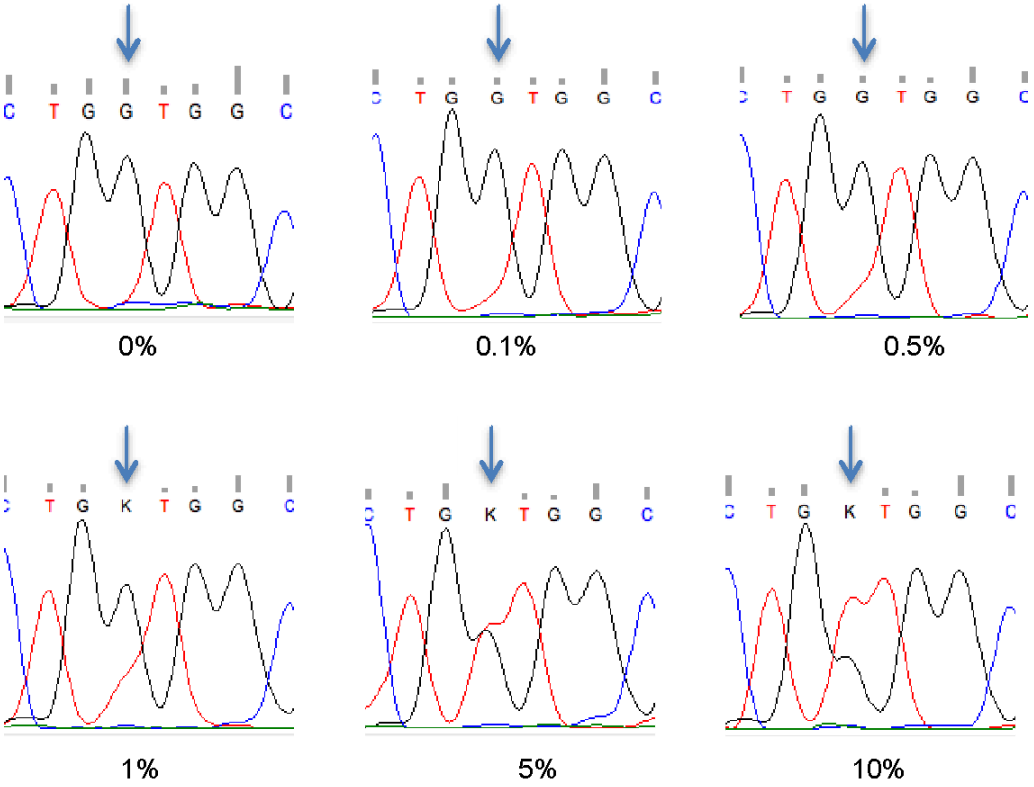


Figure S4. Sanger sequencing sonograms of samples incubated with CRISPR Cas9 (DASH) for 20 min in the absence of RPA amplification. DASH enabled detection of mutant alleles only in samples with a MAF >1%.

[**Jinzhao: in Table S2, you say that DASH can do 0.1%. What makes the difference. You need either to change the statement or explain how conditions are different**.]

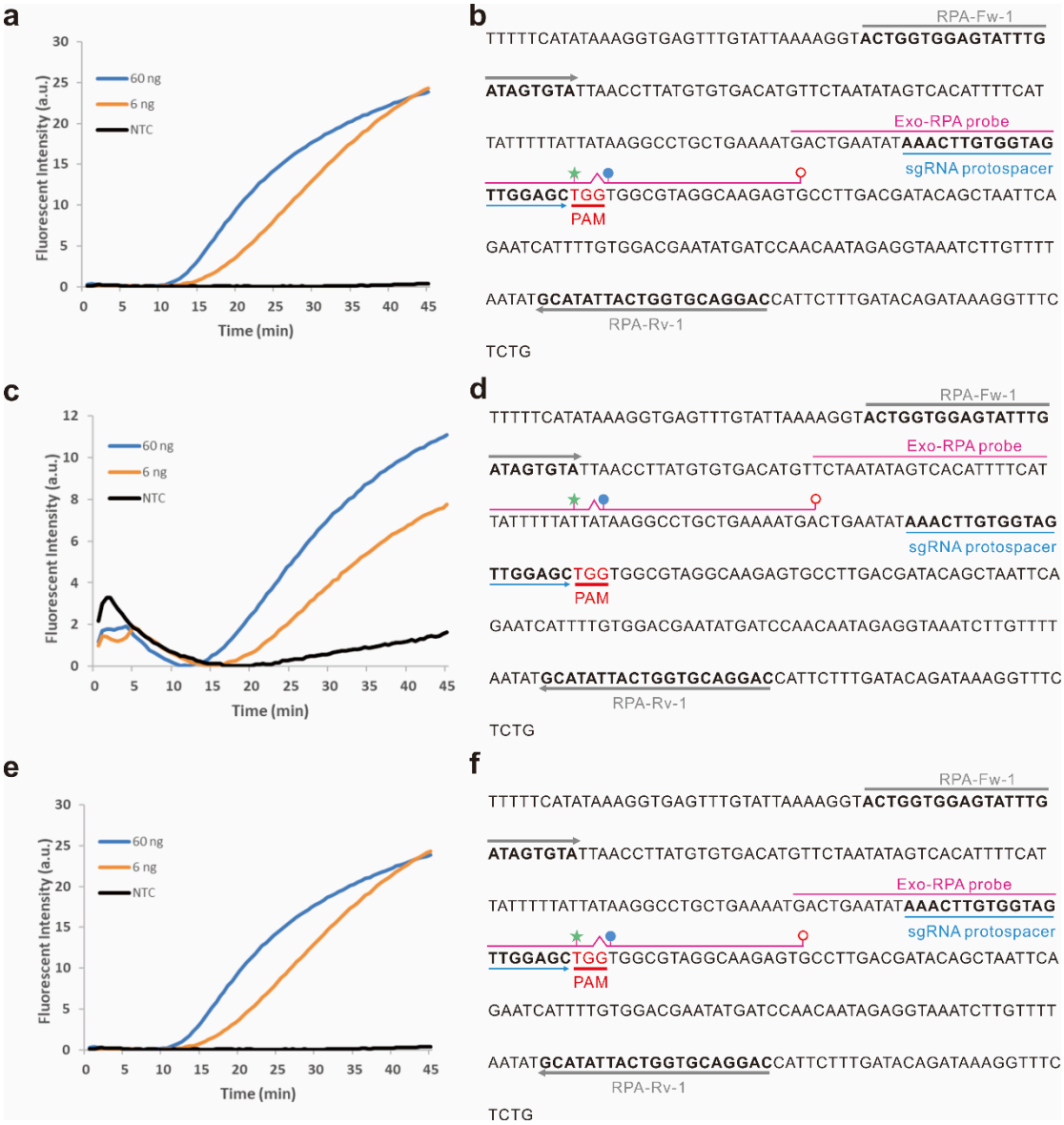


Figure S5. RPA probe design and evaluation for real-time detection. **(a)**, **(c),** and **(e)** real-time RPA monitoring of serially diluted wild-type alleles in the absence of RNP with our Exo-RPA probes. **(b)**, **(d)** and **(f)** *KRAS* gene with the locations of different Exo-RPA probe designs, primers, and sgRNA protospacer. The probe in (f) (**Table S1**) was selected for real-time PASEA. Probe concentration: 240 nM.

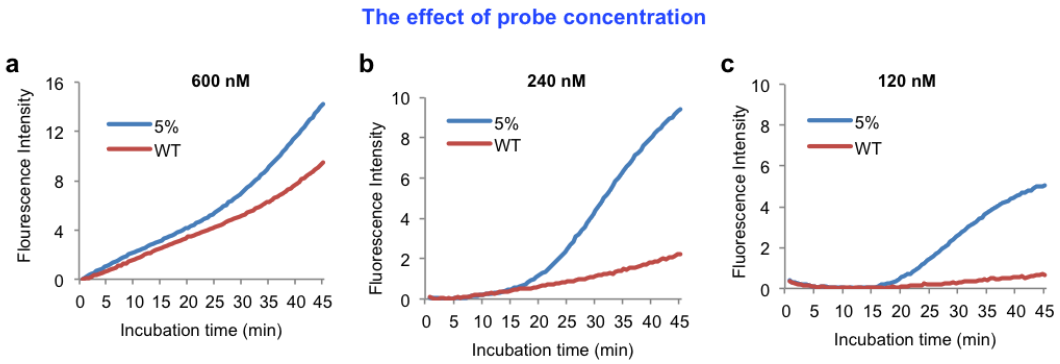


Figure S6. Optimization of probe concentration for real-time PASEA. Real-time PASEA with various Exo-probe concentrations: 600 nM (a), 240 nM (b), 120 nM (c). Genomic DNA samples containing 5% and 0% *KRAS* G12V alleles were used as templates. 0.1 μM RNP.

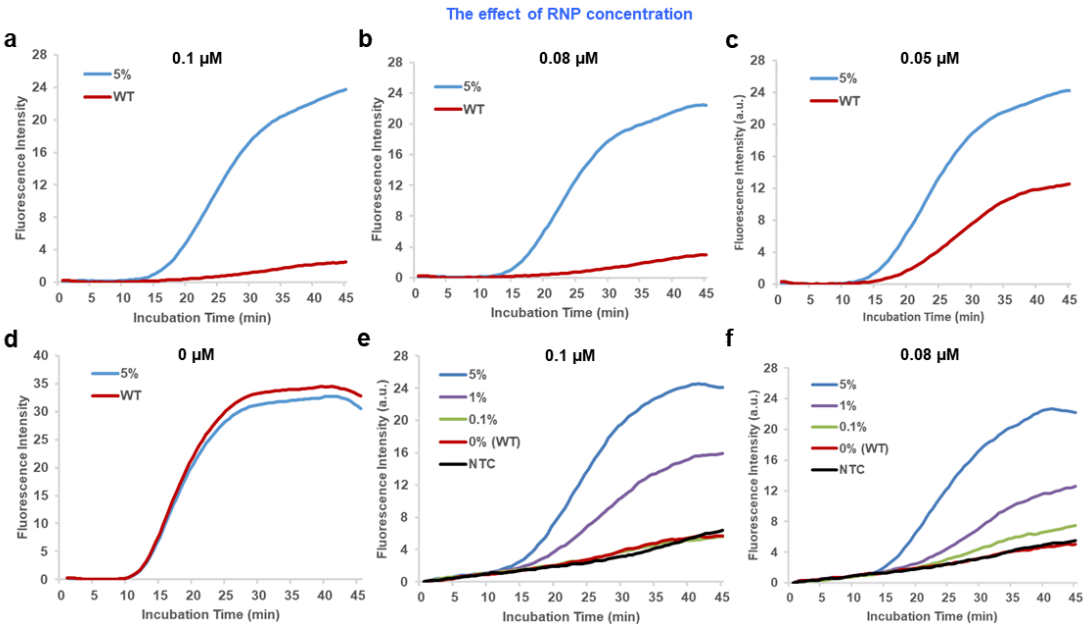


Figure S7. Optimization of RNP concentration for real-time PASEA. Amplification curves of genomic DNA samples containing 5% and 0% *KRAS* G12V with 0.1 μM (a), 0.08 μM (b), 0.05 μM (c), and 0 μM (d) RNP. Amplification curves of cfDNA samples containing 0%, 0.1%, 1%, and 5% *KRAS* G12D with 0.1 μM (e) and 0.08 μM RNP (f). NTC is non-template control. 240 μM Exo-probe.

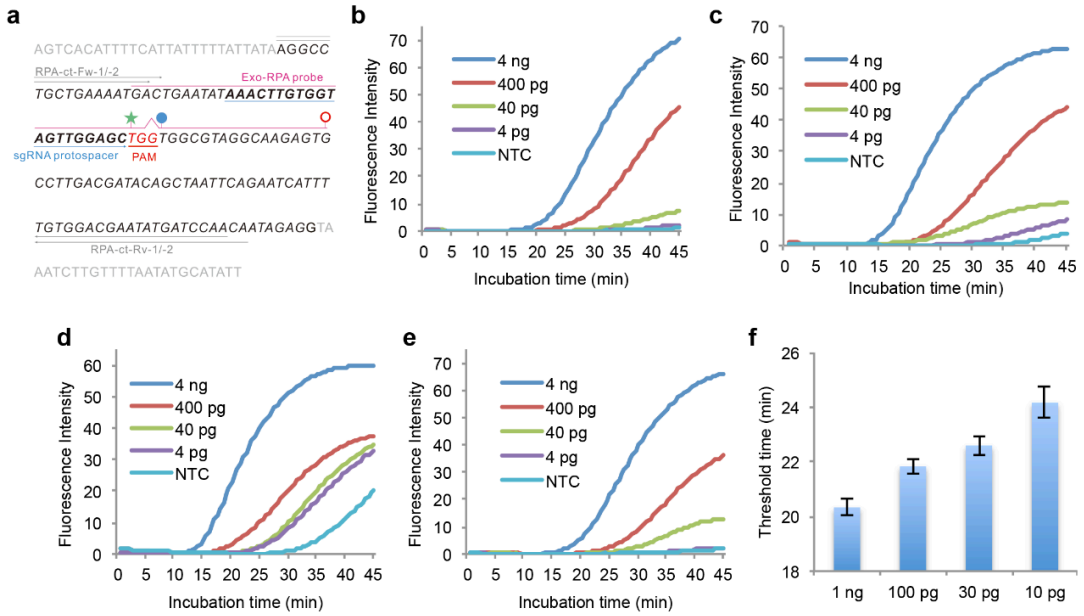


Figure S8. Selection of RPA primers for both ctDNA and ctRNA. (a) *KRAS* gene sequence around exon 2, showing the locations of sgRNA protospacer, Exo-RPA probe, and the primers. The homologous sequence (125 bp) with cDNA is shown in black letters, which contains exon 2 (italic, 122 bp). **(b)-(e)** Real-time RPA curves of serially diluted, wild-type alleles in the absence of RNP with primer sets (**Table S1**) RPA-ct-Fw-1/ RPA-ct-Rv-1 (b), RPA-ct-Fw-1/ RPA-ct-Rv-2 (c), RPA-ct-Fw-2/ RPA-ct-Rv-2 (d), and RPA-ct-Fw-2/ RPA-ct-Rv-1 (e). **(f)** Threshold time as a function of total nucleic acid concentration in the sample (primer set RPA-ct-Fw-1/ RPA-ct-Rv-2). Probe concentration: 240 nM. The threshold time is defined as the time until the fluorescent signal exceeds ~10% of its saturation emission intensity.

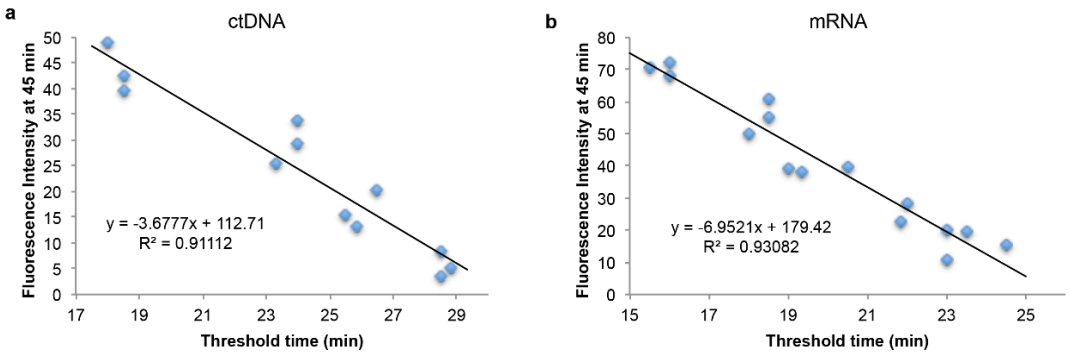
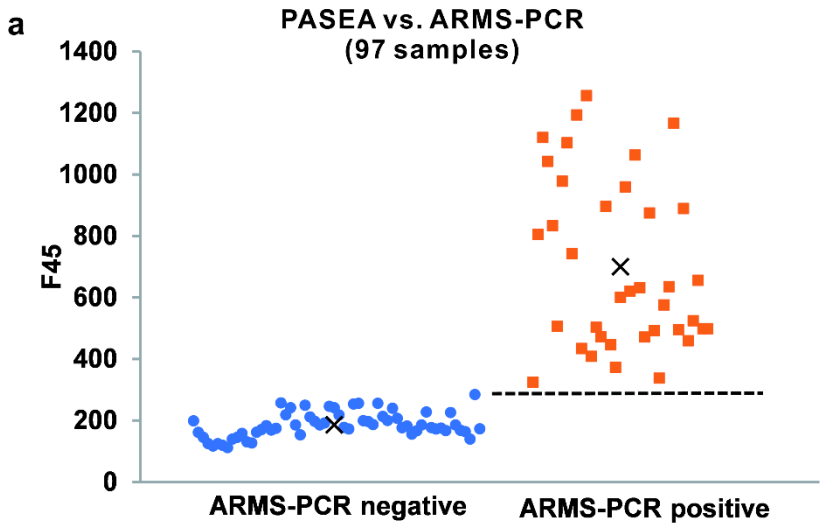


Figure S9. The emission intensity (F_{45}) correlates well with threshold time. (a) ctDNA detection. **(b)** RNA detection. The data of the emission intensity and the threshold time corresponds to **Figure 4d**, **4e**, **4g**, **4h**.

a



b

		ARMS-PCR			RESULT
		M+	M-	Total	Sensitivity=100%
PASEA	M+	37	0	37	Specificity=100%
	M-	0	60	60	PPV=100%
	TOTAL	37	60	97	NPV=100%
					Accuracy=100%

Figure S10. Real-time PASEA detects successfully gDNA in clinical samples. **(a)** F_{45} values of PASEA tests (97) compared with tissue ARMS-PCR genotyping ("gold standard"). Dashed horizontal line and the symbol "x" denote, respectively, F_{45} cutoff value and average F_{45} value **(Table S3)**. **(b)** Real-time PASEA compared with tissue ARMS-PCR genotyping.

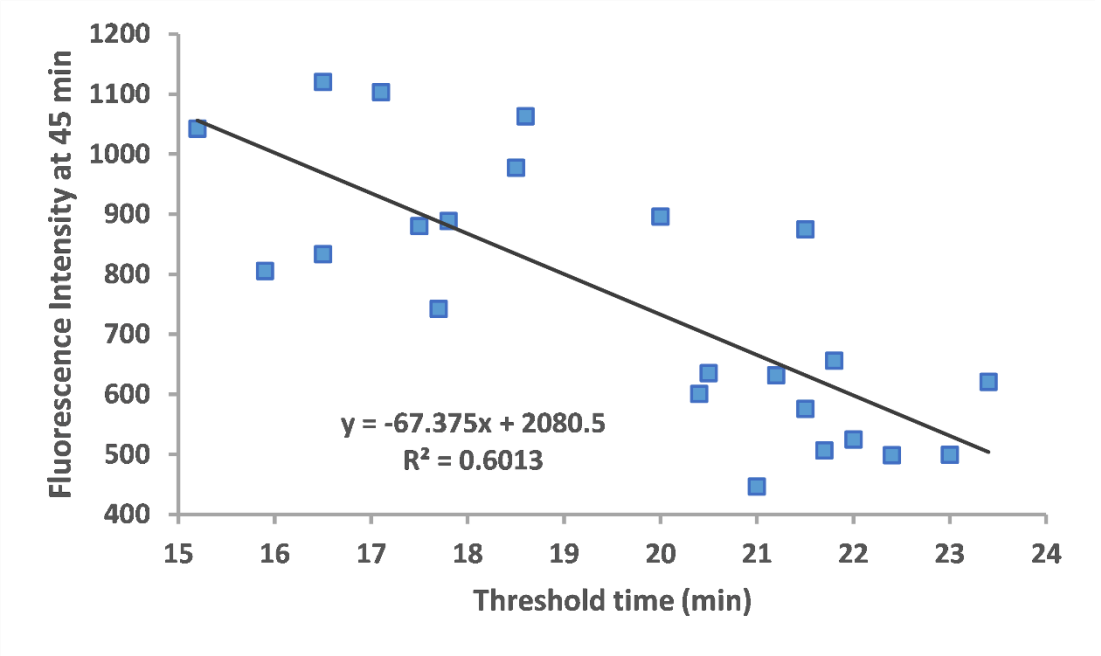


Figure S11. The emission intensity (F_{45}) correlates well with threshold time for clinical sample detection. The data of the emission intensity and the threshold time are from **Figure 5a**.

Table S1: Sequences of RPA primers, RNA guide, and Exo-probe.

Name	Sequence (from 5'to 3')
RPA-g-Fw-1	ACTGGTGGAGTATTTGATAGTGTA
RPA-g-Rv-1	GTCCTGCACCAGTAATATGC
RPA-ct-Fw-1	AGGCCTGCTGAAAATGA
RPA-ct-Fw-2	AGGCCTGCTGAAAATGAC
RPA-ct-Rv-1	TTGGATCATATTCGTCCACA
RPA-ct-Rv-2	TGTTGGATCATATTCGTCCACA
sgRNA protospacer sequence	AAACTTGTGGTAGTTGGAGC
Exo-RPA-Probe	GACTGAATATAAACTTGTGGTAGTTGGAGC[FAM-dT]G[THF][BHQ-dT] GGCGTAGGCAAGAGTG[C3 Spacer]

Table S2: Comparison of various Cas9-assisted mutant allele enrichment methods

MAF before enrichment	Method	Incubation time (min)	MAF after enrichment	Fold enrichment
0.01%	DASH (*5*)	N/A	N/A	N/A
	CUT-PCR (*6*)	60	0.35%	18
	PASEA	20	40%	4000
0.1%	DASH (*5*)	60	6.5%	65
	CUT-PCR (*6*)	60	3%	30
	PASEA	20	80%	800
1%	DASH (*5*)	60	30%	30
	CUT-PCR (*6*)	60	20%	20
	PASEA	20	100%	100

Table S3: Genomic DNA extracted from 108 tissue samples (29 patients diagnosed with colonic adenocarcinoma, 33 patients diagnosed with rectal adenocarcinoma, 45 patients diagnosed with lung cancer, and 1 patient diagnosed with pancreatic cancer).

Blinded sample No.	Sex (1: male; 2: female)	Diagnosis	Tissue resource	Tissue mutation type	NGS*	ARMS-PCR (*KRAS*)*	PASEA
1	1	colonic adenocarcinoma	resection	WT	N		N
2	1	rectal adenocarcinoma	resection	WT	N	N	N
3	1	colonic adenocarcinoma	resection	WT	N	N	N
4	1	rectal adenocarcinoma	biopsy	WT	N	N	N
5	1	rectal adenocarcinoma	resection	WT	N	N	N
6	2	rectal adenocarcinoma	resection	WT	N	N	N
7	1	rectal adenocarcinoma	resection	WT	N	N	N
8	1	rectal adenocarcinoma	resection	WT	N	N	N
9	1	rectal adenocarcinoma	resection	WT	N	N	N
10	1	rectal adenocarcinoma	resection	WT	N	N	N
11	1	colonic adenocarcinoma	resection	WT	N	N	N
12	2	colonic adenocarcinoma	resection	WT	N	N	N
13	1	rectal adenocarcinoma	resection	WT	N	N	N
14	2	colonic adenocarcinoma	resection	WT	N	N	N
15	1	lung cancer	resection	WT	N		N
16	1	lung cancer	resection	WT	N		N
17	1	lung cancer	resection	WT	N		N
18	1	lung cancer	resection	WT	N		N
19	2	lung cancer	resection	WT	N		N
20	1	lung cancer	resection	WT	N		N
21	1	colonic adenocarcinoma	resection	G12C	P	P	P (+)
22	2	colonic adenocarcinoma	resection	G12D	P	P	P (+++)
23	1	colonic adenocarcinoma	resection	G12D	P	P	P (++++)
24	1	colonic adenocarcinoma	resection	G12D	P	P	P (++++)
25	1	colonic adenocarcinoma	biopsy	G12D	P	P	P (+++)
26	1	colonic adenocarcinoma	resection	G12D	P	P	P (+)

27	1	colonic adenocarcinoma	resection	G12D	P	P	P (+++)
28	1	rectal adenocarcinoma	biopsy	G12V	P	P	P (++++)
29	2	rectal adenocarcinoma	resection	G12D	P	P	P (++)
30	2	lung cancer	resection	G12V	P		P (+)
31	1	lung cancer	resection	G12C	P		P (+++)
32	1	rectal adenocarcinoma	resection	G12V	P	P	P (++++)
33	2	rectal adenocarcinoma	resection	G12S	P	P	P (+)
34	1	colonic adenocarcinoma	biopsy	G12V	P	P	P (++++)
35	1	rectal adenocarcinoma	resection	G12V	P	P	P (+)
36	1	rectal adenocarcinoma	resection	G12V	P	P	P (+)
37	1	rectal adenocarcinoma	resection	G12V	P	P	P (+)
38	2	rectal adenocarcinoma	resection	G12C	P	P	P (+++)
39	1	pancreatic cancer	resection	G12V	P	P	P (+)
40	1	rectal adenocarcinoma	resection	G12D	P	P	P (+)
41	1	colonic adenocarcinoma	biopsy	G12V	P	P	P (++)
42	1	colonic adenocarcinoma	biopsy	G12V	P	P	P (+++)
43	2	rectal adenocarcinoma	resection	G12D	P	P	P (++)
44	1	colonic adenocarcinoma	resection	G12D	P	P	P (++++)
45	2	colonic adenocarcinoma	resection	G12D	P	P	P (++)
46	1	rectal adenocarcinoma	biopsy	G12D	P	P	P (+)
47	1	rectal adenocarcinoma	resection	G12C	P	P	P (+++)
48	2	colonic adenocarcinoma	resection	G12C	P	P	P (+)
49	1	rectal adenocarcinoma	resection	G12S	P	P	P (+)
50	1	lung cancer	resection	G12C	P		P (+)
51	2	lung cancer	resection	G12C	P	P	P (++)
52	1	lung cancer	resection	G12V	P	P	P (++)
53	2	lung cancer	resection	G12D	P	P	P (++++)
54	1	lung cancer	resection	G12C	P	P	P (+)
55	2	lung cancer	resection	G12V	P	P	P (+++)
56	1	lung cancer	resection	G12R	P	P	P (+)
57	1	lung cancer	resection	G12C	P	P	P (++)
58	1	lung cancer	resection	G12D	P	P	P (++)

59	1	lung cancer	resection	G12D	P	P	P (+)
60	2	lung cancer	resection	G12V	P	P	P (+)
61	1	colonic adenocarcinoma	resection	WT	N	N	N
62	2	rectal adenocarcinoma	resection	WT	N	N	N
63	2	rectal adenocarcinoma	resection	WT	N	N	N
64	1	colonic adenocarcinoma	resection	WT	N	N	N
65	1	rectal adenocarcinoma	biopsy	WT	N	N	N
66	2	rectal adenocarcinoma	resection	WT	N	N	N
67	1	colonic adenocarcinoma	resection	WT	N	N	N
68	2	rectal adenocarcinoma	resection	WT	N	N	N
69	1	rectal adenocarcinoma	resection	WT	N	N	N
70	1	rectal adenocarcinoma	resection	WT	N	N	N
71	2	rectal adenocarcinoma	resection	WT	N	N	N
72	2	colonic adenocarcinoma	resection	WT	N	N	N
73	1	colonic adenocarcinoma	resection	WT	N	N	N
74	1	colonic adenocarcinoma	biopsy	WT	N	N	N
75	1	rectal adenocarcinoma	biopsy	WT	N	N	N
76	2	colonic adenocarcinoma	resection	WT	N	N	N
77	2	colonic adenocarcinoma	resection	WT	N	N	N
78	1	colonic adenocarcinoma	biopsy	WT	N	N	N
79	1	colonic adenocarcinoma	resection	WT	N	N	N
80	1	rectal adenocarcinoma	resection	WT	N	N	N
81	1	rectal adenocarcinoma	resection	WT	N	N	N
82	2	colonic adenocarcinoma	resection	WT	N	N	N
83	2	lung cancer	resection	WT		N	N
84	2	lung cancer	resection	WT		N	N
85	1	lung cancer	resection	WT		N	N
86	2	lung cancer	resection	WT		N	N
87	1	lung cancer	resection	WT		N	N
88	1	lung cancer	resection	WT		N	N
89	1	lung cancer	resection	WT		N	N
90	2	lung cancer	resection	WT		N	N

91	2	lung cancer	resection	WT		N	N
92	2	lung cancer	resection	WT		N	N
93	2	lung cancer	resection	WT		N	N
94	2	lung cancer	resection	WT		N	N
95	2	lung cancer	resection	WT	N		N
96	1	lung cancer	resection	WT		N	N
97	1	lung cancer	resection	WT		N	N
98	1	lung cancer	resection	WT		N	N
99	2	lung cancer	resection	WT		N	N
100	2	lung cancer	resection	WT		N	N
101	1	lung cancer	resection	WT		N	N
102	1	lung cancer	resection	WT		N	N
103	2	lung cancer	resection	WT		N	N
104	2	lung cancer	resection	WT		N	N
105	1	lung cancer	resection	WT		N	N
106	2	lung cancer	resection	WT		N	N
107	2	lung cancer	resection	WT		N	N
108	2	lung cancer	resection	WT		N	N
Positive control				G12V (33%)			P (++++)
Positive control				G12V (5%)			P (++)
Positive control				G12V (1%)			P (+)

"++++", "+++", "++", "+" indicates, respectively, positive when F_{45} > 1006.8, >756.8, >450.0, >285.7

Cutoff (0%) = F_{45} + 3SD (N=6) = 181.9+3×34.6 = 285.7

*** From the 108 tissue biopsy and resection samples, 83 were tested with NGS, 97 were tested with ARMS-PCR (*KRAS*), and 72 were tested with both methods.**

References

[1] H. R. Underhill, J. O. Kitzman, H. Sabine, N. C. Welker, D. Riza, D. N. Baker, K. M. Gligorich, R. C. Rostomily, M. P. Bronner, S. Jay, PLoS Genet. **12** (2016) e1006162.

[2] A. Sato, C. Nakashima, T. Abe, J. Kato, N. Sueoka-Aragane, Oncotarget **9** (2018) 31904-31914.

[3] J. Song, V. Pandian, M. G. Mauk, B. H. H., S. Cherry, L. C. Tisi, C. Liu, Anal. Chem. **90** (2018) 4823-4831.

[4] K. Kadimisetty, J. Song, A. M. Doto, Y. Hwang, J. Peng, M. G. Mauk, F. D. Bushman, R. Gross, J. N. Jarvis, C. Liu, Biosens. Bioelectron. **109** (2018) 156-163.

[5] W. Gu, E. D. Crawford, B. D. O'Donovan, M. R. Wilson, E. D. Chow, H. Retallack, J. L. Derisi, Genome. Biol. **17** (2016) 41.

[6] S. Lee, J. Yu, G.-H. Hwang, S. Kim, H. Kim, S. Ye, K. Kim, J. Park, D. Park, Y.-K. Cho, J.-S. Kim, S. Bae, Oncogene **36** (2017) 6823–6829.