Ultrasensitive detection of circulating LINE-1 ORF1p as a specific multi-cancer biomarker

Martin S. Taylor*[,1,‡], Connie Wu*[,1,2,a ‡], Peter C. Fridy[3], Stephanie J. Zhang[1,2], Yasmeen Senussi[1,2], Justina C. Wolters[4], Tatiana Cajuso[5], Wen-Chih Cheng[5], John D. Heaps[5], Bryant D. Miller[5], Kei Mori[1,6], Limor Cohen[1,2,7], Hua Jiang[3], Kelly R. Molloy[8], Brian T. Chait[8], Michael G. Goggins[9], Irun Bhan[10], Joseph W. Franses[10], Xiaoyu Yang[11], Mary-Ellen Taplin[11], Xinan Wang[12], David C. Christiani[10,12], Bruce E. Johnson[11], Matthew Meyerson[11], Ravindra Uppaluri[13], Ann Marie Egloff[13], Elyssa N. Denault[10], Laura M. Spring[10], Tian-Li Wang[9], Ie-Ming Shih[9], Jennifer E. Fairman[9], Euihye Jung[14], Kshitij S. Arora[1,b], Osman H. Yilmaz[15], Sonia Cohen[16], Tatyana Sharova[16], Gary Chi[10], Bryanna L. Norden[10], Yuhui Song[10], Linda T. Nieman[10], Leontios Pappas[10], Aparna R. Parikh[10,] Matthew R. Strickland[10], Ryan B. Corcoran[10],Tomas Mustelin[17], George Eng[1,18], Ömer H. Yilmaz[1,18], Ursula A. Matulonis[11], Steven J. Skates[19], Bo R. Rueda[20], Ronny Drapkin[14], Samuel J. Klempner[10], Vikram Deshpande[1], David T. Ting[10], Michael P. Rout[3], John LaCava[3,21], David R. Walt[1,2,‡], and Kathleen H. Burns[1,5,‡]

1. Department of Pathology, Mass General Brigham and Harvard Medical School, Boston, MA, USA.
2. Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, MA, USA.
3. Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY, USA.
4. Department of Pediatrics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
5. Department of Pathology, Dana Farber Cancer Institute and Harvard Medical School, Boston, MA, USA.
6. Healthcare Optics Research Laboratory, Canon U.S.A., Inc., Cambridge, MA, USA
7. Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA
8. Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, New York, USA
9. Johns Hopkins University School of Medicine, Baltimore, MD, USA.
10. Mass General Cancer Center and Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.
11. Department of Medical Oncology, Dana Farber Cancer Institute and Harvard Medical School, Boston, MA, USA.
12. Department of Environmental Health, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA
13. Department of Surgery, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
14. University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.
15. Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA.
16. Department of Surgery, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
17. Division of Rheumatology, Department of Medicine, University of Washington, Seattle, WA, USA.
18. The David H. Koch Institute for Integrative Cancer Research at MIT, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

19. MGH Biostatistics, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.
20. Department of Obstetrics and Gynecology, Massachusetts General Hospital, and Harvard Medical School, Boston, MA, USA.
21. European Research Institute for the Biology of Ageing, University Medical Center Groningen, The Netherlands.

* Equal Contribution
‡ Correspondence : mstaylor@mgh.harvard.edu, conniewu@umich.edu,dwalt@bwh.harvard.edu, kathleenh_burns@dfci.harvard.edu
Martin S. Taylor (mstaylor@mgh.harvard.edu)
55 Fruit Street, Department of Pathology, Mass General Brigham, Boston, MA 02114
617-643-0800

Connie Wu (conniewu@umich.edu)
210 Washtenaw Avenue, Life Sciences Institute, University of Michigan, Ann Arbor, MI 48109
734-647-2360

David R. Walt (dwalt@bwh.harvard.edu)
60 Fenwood Road, Department of Pathology, Mass General Brigham, Boston, MA 02115
857-307-1112

Kathleen H. Burns (kathleenh_burns@dfci.harvard.edu)
450 Brookline Avenue, Mayer Building 641, Dana-Farber Cancer Institute, Boston, MA 02215
857-215-0216

Present addresses: [a] University of Michigan Life Sciences Institute, Department of Biomedical Engineering, Ann Arbor, MI, USA. [b]Department of Pathology, Louisiana State University Health, Shreveport.

Running Title: LINE-1 Protein Assays for Cancer Detection & Monitoring

Competing Interests

MST has received consulting fees from ROME Therapeutics and Tessera Therapeutics that are not related to this work. MST and JL have equity in ROME therapeutics. DTT has received consulting fees from ROME Therapeutics, Tekla Capital, Ikena Oncology, Foundation Medicine, Inc., NanoString Technologies, and Pfizer that are not related to this work. DTT is a founder and has equity in ROME Therapeutics, PanTher Therapeutics and TellBio, Inc., which is not related to this work. DTT receives research support from ACD-Biotechne, PureTech Health LLC, Ribon Therapeutics, and Incyte, which was not used in this work. LMS declares the following relationships: Consultant/advisory board: Novartis, Puma, G1 therapeutics, Daiichi Pharma, Astra Zeneca; Institutional research support: Phillips, Merck, Genentech, Gilead, Eli Lilly. SJK declares Consulting/advisory: Eli Lilly, Merck, BMS, Novartis, Astellas, AstraZeneca, Daiichi-Sankyo, Novartis, Sanofi-Aventis, Natera, Exact Sciences, Mersana. Stock/Equity: Turning

Point Therapeutics, Nuvalent. BRR serves on SAB for VincenTech and receives research support from Novartis Institutes for Biomedical Research that are not related to this work. DRW has a financial interest in Quanterix Corporation, a company that develops an ultra-sensitive digital immunoassay platform. He is an inventor of the Simoa technology, a founder of the company and also serves on its Board of Directors. KHB declares relationships with Alamar Biosciences, Genscript, Oncolinea/PrimeFour Therapeutics, ROME Therapeutics, Scaffold Therapeutics, Tessera Therapeutics, and Transposon Therapeutics. MST and KHB receive royalties from sales of ORF1p antibodies and MST, CW, PCF, KRM, BTC, MPR, JL, DRW, and KHB are inventors on a patent related to this work. MST, LMS, SJK, BRR, and DTT's interests were reviewed and are managed by Massachusetts General Hospital and Mass General Brigham in accordance with their conflict-of-interest policies. Dr. Walt's interests were reviewed and are managed by Mass General Brigham and Harvard University in accordance with their conflict-of-interest policies. KHB's interests are managed by Dana-Farber Cancer Institute.

Abstract

Improved biomarkers are needed for early cancer detection, risk stratification, treatment selection, and monitoring treatment response. While proteins can be useful blood-based biomarkers, many have limited sensitivity or specificity for these applications. Long INterspersed Element-1 (LINE-1) open reading frame 1 protein (ORF1p) is a transposable element protein overexpressed in carcinomas and high-risk precursors during carcinogenesis with negligible expression in normal tissues, suggesting ORF1p could be a highly specific cancer biomarker. To explore ORF1p as a blood-based biomarker, we engineered ultrasensitive digital immunoassays that detect mid-attomolar (10^{-17} M) ORF1p concentrations in plasma across multiple cancers with high specificity. Plasma ORF1p shows promise for early detection of ovarian cancer, improves diagnostic performance in a multi-analyte panel, provides early therapeutic response monitoring in gastroesophageal cancers, and is prognostic for overall survival in gastroesophageal and colorectal cancers. Together, these observations nominate ORF1p as a multi-cancer biomarker with potential utility for disease detection and monitoring.

Statement of Significance (50 word)

The LINE-1 ORF1p transposon protein is pervasively expressed in many cancers and is a highly specific biomarker of multiple common, lethal carcinomas and their high-risk precursors in

tissue and blood. Ultrasensitive ORF1p assays from as little as 25 µL plasma are novel, rapid, cost-effective tools in cancer detection and monitoring.

Introduction

There is significant clinical need for non-invasive methods to detect, risk stratify, and monitor cancers over time. Many malignancies are diagnosed at late stages when disease is widespread, contributing significantly to cancer morbidity and mortality(1). In contrast, there is a likely window in early-stage disease when patients are typically asymptomatic, in which treatments can be much more effective. Biomarkers are also needed to assess likelihood of progression in patients with precursor lesions, to provide prognostic information, and to predict and monitor responses or resistance to treatment(2). Considerable advances have been made towards detecting circulating tumor DNA, circulating tumor cells, microRNAs, and extracellular vesicles as non-invasive cancer biomarkers(3). However, achieving high sensitivities and specificities, particularly in affordable, scalable, clinical grade screening assays for early cancer detection, remains a major challenge. The plasma proteome provides a rich reservoir of potential biomarkers(4), which may be used individually or in combination for Multi-Cancer Early Detection (MCED) assays(5). However, most readily detectable proteins, including CA125 and HE4(6), FDA-cleared markers for the differential diagnosis of pelvic masses, are not sufficiently sensitive at the required high specificity(7) for cancer screening and/or are expressed in normal tissues and therefore lack the requisite specificity.

We have previously shown that expression of long interspersed element-1 (L1, LINE-1)-encoded open reading frame 1 protein (ORF1p) is a hallmark of many cancers(8), particularly p53-deficient epithelial cancers. These encompass many of the most commonly occurring and lethal human cancers, including esophageal, colorectal, lung, breast, prostate, ovarian, uterine, pancreatic, and head and neck cancers. L1 is the only active protein-coding transposon in

humans. We each inherit, dispersed throughout our genomes, a complement of active L1 loci encoding two proteins: ORF1p, the highly expressed RNA binding protein(8), and ORF2p, an endonuclease and reverse transcriptase with limited expression(9) that generates L1 insertions in cancer genomes(10-13). L1 expression is repressed in normal somatic tissues, resulting in either very low or undetectable levels of L1 RNA and protein that appear to originate from epithelium(9,14). Epigenetic dysregulation of L1 and L1 ORF1p overexpression begins early in carcinogenesis, and histologic precursors of ovarian, esophageal, colorectal, and pancreatic cancers studied all express ORF1p at varying levels(8,15). ORF1p is thus a promising highly specific cancer biomarker.

Although elevated expression of ORF1p is readily detected by immunostaining in tumor tissue, ORF1p is found in plasma at low concentrations, well below detection limits of conventional clinical laboratory methods. We therefore applied the much more sensitive Single Molecule Arrays (Simoa), a digital bead-based ELISA technology, and in preliminary studies detected ORF1p in plasma at femtomolar levels in subsets of patients with advanced breast (33%, n=6)(16) and colorectal (90%, n=32)(17) cancers, respectively. Here, we assess the landscape of ORF1p plasma levels across multiple cancers, iteratively develop highly sensitive assays for potential applications in early or minimal residual disease detection, and provide evidence that plasma ORF1p may be an early indicator of therapeutic response.

Results

Because our preliminary survey of plasma ORF1p levels by Simoa in patients with advanced stage colorectal cancer (CRC) indicated detectable ORF1p levels in 90% of cases(17), higher than the proportion of CRCs we previously reported to express ORF1p by immunohistochemistry (50%, n=18)(8), we first sought to benchmark ORF1p in tissues. Using a re-optimized protocol(8), we stained 211 CRCs [178 sequential cases included on a tissue

71 microarray (TMA) as well as an additional 33 with matched plasma] and found 91% of CRC
72 cases were immunoreactive for ORF1p (**Fig. 1a**). This result is consistent with genetic studies
73 demonstrating somatic L1 retrotransposition in most CRCs(18), including activity in
74 precancerous lesions antedating *APC* tumor suppressor loss(19-21). Similarly, genetic evidence
75 shows esophageal adenocarcinoma (EAC) has high L1 activity(12), and L1 insertions occur in
76 the highly prevalent Barrett's esophagus (BE) precursor early in carcinogenesis(22,23). We
77 therefore assembled a cross-sectional cohort of 72 BE cases with consensus diagnosis reached
78 by three expert gastrointestinal pathologists from two institutions. L1 RNA and ORF1p
79 expression were pervasive in dysplastic BE and present in 100% of 51 esophageal carcinomas
80 (**Fig. 1b,c**); all five BE cases indefinite for dysplasia and positive for ORF1p and/or L1 RNA
81 developed high grade dysplasia on subsequent biopsies. Overall, this picture is similar to high
82 grade serous ovarian cancers (HGSOC), where ORF1p is expressed in 90% of cases and 90%
83 of fallopian tube precursor lesions (serous tubal intraepithelial carcinomas, STICs)(8,15,24). The
84 cumulative ORF1p staining data to date across carcinomas are summarized in **Fig. 1d**. Taken
85 together, ORF1p tissue expression is highly prevalent in gastrointestinal and gynecologic
86 carcinomas and high-risk precursor lesions.
87
88 We next sought to extend our tissue findings and explore plasma ORF1p. We optimized our
89 previously reported ORF1p Simoa assay and assessed the landscape of ORF1p levels in
90 pretreatment plasma from patients with advanced cancers. This "first-generation" assay uses a
91 recombinant, single-domain camelid nanobody (Nb5) as the capture reagent and a monoclonal
92 antibody (Ab6) as the detector reagent and has a limit of detection of 0.056 pg/mL (~470 aM
93 trimeric ORF1p), corresponding to 1.9 fM in plasma after correcting for sample dilution (**Fig. 2a,**
94 **Supplementary Table S1**). With this assay, we surveyed multiple cancer types and >400
95 'healthy' control individuals, who were without known cancer at the time blood was donated to
96 the biobank. Plasma ORF1p appears to be a highly specific cancer biomarker, with

undetectable levels in ~99% of controls (ages 20-90, **Fig. 2b, Supplementary Fig. S1A-C**). Of the five control patients with detectable ORF1p, the one with the highest ORF1p was found six months later to have advanced prostate cancer and 19 months later a cutaneous T cell lymphoma; limited clinical information is available for the other four positive 'healthy' individuals. With a cutoff set at 98% specificity in healthy controls, the highest proportions of ORF1p(+) cases were observed in colorectal (58%, n=101) and ovarian cancers (71%, n=145). While most of these patients had advanced-stage disease, plasma ORF1p remained detectable in several early-stage patients in the cohort, including in those with ovarian and lung cancers and in 5/18 with intraductal papillary mucinous neoplasms in the pancreas (IPMN, **Supplementary Fig. S2-S4**). Notably, four of eight stage I ovarian cancers in the cohort were positive (**Supplementary Fig. S2**), suggesting that plasma ORF1p may be an indicator of early-stage disease. As L1 expression is also dysregulated in autoimmune disease and autoantibodies against ORF1p are prevalent in patients with systemic lupus erythematosus (SLE), we measured plasma ORF1p in 30 SLE patients and observed no detectable levels (**Supplementary Fig. S5**)(25). Detectable ORF1p was seen in 1 of 30 patients with chronic liver disease; the one positive patient was subsequently diagnosed with hepatocellular carcinoma (**Supplementary Fig. S5**). Size exclusion chromatography analysis of patient plasma further showed that the majority of ORF1p resides outside extracellular vesicles (**Supplementary Fig. S6A-B**). Genomics analysis was available for a subset of patients in the lung cancer patient cohort (n=32); interestingly, detectable plasma ORF1p was associated with more genomic amplifications, higher tumor mutational burden (p=0.02 and 0.007, respectively, Wilcoxon test, **Supplementary Fig. S7A**), and tended to have more *TP53* mutations and fewer *KRAS* mutations (**Supplementary Fig. S7B**). ORF1p did not correlate with PSA levels in prostate cancer patients (**Supplementary Fig. S8**). Together, these findings support the hypothesis that tumor-derived ORF1p can be found in the peripheral blood of cancer patients and may act as a cancer-specific biomarker.

Given the gap between proportions of ORF1p(+) cancers by tumor immunohistochemistry (~90% for CRC and HGSOC) versus by blood testing (~60-70%), we evaluated the possibility of increasing plasma assay sensitivity by decreasing the assay's lower limit of detection. To this end, we developed a panel of ORF1p affinity reagents, including new recombinant rabbit monoclonal antibodies (RabMAbs) and engineered camelid nanobodies raised against recombinant human ORF1p. Because ORF1p is homotrimeric, we engineered multimeric nanobody reagents with the goal of enhancing binding affinity via increased avidity. These parallel development efforts ultimately yielded both improved nanobody and rabbit monoclonal antibody reagents with at least low-picomolar equilibrium dissociation constants (K_D) (**Supplementary Fig. S9-S14, Supplementary Tables S2-S4**). Monoclonal antibodies were further validated by western blotting (**Supplementary Fig. S15**). Iterative screening of these reagents with Simoa using recombinant antigen and select patient plasma samples yielded three best-performing capture::detection pairs, termed "second-generation," which use rabbit monoclonal antibodies 34H7 and 62H12 as capture reagents and either Ab6 or homodimeric form of Nb5 (Nb5-5LL) as detector (**Fig. 3a-c, Supplementary Figs. S16-S19**). Adding detergent further improved performance by limiting bead aggregation and improving bead loading into microwells. These second-generation assays achieve detection limits of 0.016-0.029 pg/mL (130-240 aM trimeric ORF1p), and the four different reagents have predominantly non-overlapping epitopes in binning experiments (34H7 and 62H12 partially overlap, **Fig. 3a-b, Supplementary Tables S1, S5-S6**). Somewhat unexpectedly, analytical sensitivity of the assay (for detecting recombinant ORF1p in buffer) did not perfectly correspond to clinical sensitivity (for detecting ORF1p in cancer patient plasma). While the second-generation assays demonstrated less than an order-of-magnitude improvement in analytical sensitivity over the first-generation assay, they showed considerable improvement in circulating ORF1p detectability over background in buffer in re-measured samples across a large cohort of healthy and cancer patients (**Fig 3b, Supplementary Fig. S20A-B**). This difference may be due to

differing accessibilities of circulating ORF1p epitopes or to different nonspecific binding patterns in plasma.

Undetectable or extremely low ORF1p levels in healthy individuals could readily be discriminated from measured ORF1p levels in ovarian cancer patients, resulting in a strong discriminatory ability with single-marker models (area under the receiver operating characteristic curve, AUCs of 0.93 to 0.948, sensitivity of 41% to 81% at 98% specificity, **Fig. 3d top panel, Supplementary Table S7**). This large cohort included pre-treatment plasma samples from a sub-cohort of ovarian cancer patients (mostly high-grade serous ovarian carcinoma, "Penn cohort") with age-matched controls (n=51-53 women, **Fig 3c**); again, second-generation assays showed higher sensitivities while maintaining high specificities, notably achieving detection of five out of six Stage I/II patients at >98% specificity. Furthermore, multivariate models combining ORF1p (34H7::Nb5-5LL assay) with ovarian cancer biomarkers CA125 and HE4 yielded improved diagnostic performance over these existing markers (CA125 and HE4 alone, AUC = 0.94, 59% sensitivity at 98% specificity; ORF1p, CA125, and HE4, AUC = 0.98, 91% sensitivity at 98% specificity; **Fig 3d bottom panel, Supplementary Fig. S21; Supplementary Table S8**). While it is not clear whether the low ORF1p levels detected in several healthy individuals is due to nonspecific binding, true background levels of ORF1p, or an unappreciated pre-malignant state, several positive healthy controls were positive by only one of the three second-generation assays (n=4 positive by only 62H12::Nb5-5LL and n=75 positive by only 62H12:Ab6), suggesting nonspecific binding in at least some of these cases and the potential to improve specificity by combining data from multiple assays. Our results indicate that by developing improved affinity reagents, we achieved improved clinical sensitivity in detecting circulating ORF1p in cancer patients, with 83% sensitivity at >98% specificity towards early detection of ovarian cancer.

Receptor subtypes were available for the breast cancer cohort, which includes 30 patients each with metastatic and localized disease (**Supplementary Fig. S22A-B**). Across all assays, triple negative tended to have higher positivity rates, but the most sensitive 2nd generation assay (62H12::Ab6) detected 96% of triple negative cases and 91% of the remaining cases (**Supplementary Fig. S22**) with 93% sensitivity for both localized and metastatic disease. Overall, metastatic disease was detected more commonly than localized disease (43% vs 6.7% for 1st generation assay, 67-93% vs. 23-93% for 2nd generation assays, depending on the assay), and all three 2nd generation assays had higher sensitivity than the 1st generation assay (**Supplementary Fig. S22**).

To further validate our results, we developed a targeted proteomics approach to measure ORF1p following affinity capture, with two distinct peptides measured vs. internal isotopically labeled control peptides (**Fig. 4a**). With this assay, we applied much larger volumes of plasma (3-6 ml, 120-240 fold more than the 25 µL used in Simoa assays) from a cohort of 10 patients, including 2 gastroesophageal (GE) cancer patients and one healthy control with very high ORF1p (230-1230 pg/ml), two healthy controls with high ORF1p, (3-5 pg/ml), and 5 healthy controls with low ORF1p (undetectable – 0.2 pg/ml). The results (**Fig. 4a-b, Supplementary Fig. S23A-D**) show strong correlation with Simoa, providing further confidence in our results ($r=0.97-0.99$, $p<0.0001$, t test).

Building on the improvements made through nanobody engineering in our second-generation assays, we developed an expanded set of homodimeric, heterodimeric, and heterotrimeric anti-ORF1p nanobodies and screened them in combination with 34H7 and 62H12 capture antibodies, resulting in “third-generation” assays (**Supplementary Figs. S11, S14, S24-25**). We noticed that reagents containing Nb2 performed very well in SPR but poorly in Simoa detection, and we hypothesized this was because Nb2 contains a lysine in the CDR, which would be

biotinylated in the procedure, reducing affinity. We therefore engineered the new reagents to be C-terminally biotinylated on cysteine residues and varied linker sequence. Five of these assays, which utilize Nb2- and Nb9- containing constructs, outperform our second-generation assays in a cohort of 25 GE cancer patients with ORF1p measurements that were mostly undetectable previously, while maintaining high specificity versus healthy individuals (**Fig. 5a, Supplementary Fig. S25**).

To leverage more sensitive assays for ORF1p detection, we next tested ORF1p affinity reagents from one of the second-generation Simoa assays on our recently developed Molecular On-bead Signal Amplification for Individual Counting platform (MOSAIC, **Fig. 5b**). MOSAIC develops localized on-bead signal from single captured molecules, in contrast to the microwell array format in Simoa, and improves analytical sensitivity by an order of magnitude over Simoa via increasing the number of beads counted(26). Furthermore, as the developed Simoa assays used only 25 µL plasma, we hypothesized that using larger plasma volumes would enhance ORF1p detectability by increasing the number of analyte molecules present. By using a 20-fold higher sample volume (500 µL plasma) and the MOSAIC platform, we achieved ten-fold higher analytical sensitivity, with a limit of detection of 0.002 pg/ml ORF1p (17 aM trimer, **Supplementary Fig. S26**). Indeed, in a pilot cohort of gastroesophageal cancer and healthy patients, ORF1p levels in nine of ten previously undetectable cancer patients were readily discriminated from healthy individuals (**Fig. 5c**). Thus, in addition to improved affinity reagents, using larger sample volumes and more analytically sensitive technologies can further enhance both sensitivity and discrimination of circulating ORF1p levels between healthy controls and patients with cancer. The relative contributions of increased volume and the improved assay platform to the increased sensitivity remain to be explored; assay background seen in patient plasma (blue dashed line) but not in buffer (analytical limit of detection, red dashed line) will also require further optimization.

To test whether ORF1p might be useful for monitoring therapeutic response, 19 patients with gastroesophageal cancer were identified who had both detectable plasma ORF1p at diagnosis as well as subsequent samples available collected during or after treatment (average 80 days after initiation of therapy, range 26-179 days). Primary tumors were all adenocarcinoma and located in the esophagus (n=7), gastroesophageal junction (n=7) and stomach (n=5). All patients received systemic therapy. A smaller fraction of patients also received radiation and/or surgery (**Supplementary Table S9)**. Clinical response ('Responders' and 'Non-Responders') was determined by review of re-staging CT and PET-CT imaging by clinicians blinded to the assay results. Over an average of 465 days (range 98-1098), 12 patients died, six were alive at last follow-up (all 'Responders'), and one was lost to follow-up. Non-Responders had higher pre-treatment plasma ORF1p (**Fig. 6a**, left panel, p=0.02). All 6 patients with detectable ORF1p at follow-up sampling, as defined by positivity over background in two of three assays, were also Non-Responders by imaging (**Fig. 6a**, right panel, p<0.0001, Fisher's Exact test) and had reduced survival (p = 0.001 log-rank test for overall survival). In contrast, in all 13 Responders, circulating ORF1p dropped to undetectable levels at follow-up sampling. Plasma ORF1p in four Responders and two Non-Responders was measured at an early timepoint of 26-33 days. The timing of sampling was not different between groups (average 93 days for Non-Responders, 74 for Responders, p=0.5). Pre-therapy blood was drawn on an average of 20 days after diagnosis (range -8-48, average 22 for Non-Responders and 19 for Responders, p=0.6). Representative PET and PET-CT images are shown (**Fig. 6b**), both images are taken approximately two months after initiation of therapy, a month after the plasma ORF1p result. Thus, reduction in circulating ORF1p paralleled treatment response and survival, while persistent circulating ORF1p corresponded to patients with refractory disease, indicating the predictive potential of this marker.

Because these results indicated that pre-treatment plasma ORF1p levels might be prognostic, we evaluated the prognostic value of 2nd generation ORF1p Simoa assays in our cohorts of GE, CRC, and ovarian cancer patients. We stratified the patients based on either the median ORF1p value or ORF1p detectability (methods) and found that higher pre-therapy plasma ORF1p was significantly associated with poor survival in GE and CRC (**Fig. 6c and Supplementary Fig. S27**, p=0.0017 and 0.011, log rank test, respectively) but not in ovarian cancer (**Supplementary Fig. S27**). ORF1p remained significantly prognostic in multivariate analysis in GE and CRC (methods, **Supplementary Figs. S28-29**, **Supplementary Tables S10-S11, Supplementary**).

Discussion

Taken together, our data reveal for the first time that circulating ORF1p is a multi-cancer protein biomarker with potential utility across clinical paradigms, including early detection, risk stratification, prognostication, and treatment response. These assays are enabled by ultrasensitive single-molecule detection technologies and high-quality affinity reagents, which are both required due to the attomolar-to-femtomolar circulating levels of ORF1p in cancer patients. Iterative improvements including optimized affinity reagents, buffer, and assay design yield highly sensitive and specific assays. A 20-fold volume scale-up to 500 µL appears promising for improving sensitivity without obviously compromising specificity, and this volume remains much smaller than a typical 5-10 mL blood draw and could be scaled further without limiting clinical applicability, although it remains unclear how much of this improvement was due to the increased volume or the MOSAIC platform itself; future studies are needed to address the relative contributions to sensitivity by sample volume and platform. The data strongly suggest that these assays are measuring *bona fide* tumor-derived circulating ORF1p for the following reasons: (1) four developed assays with predominantly non-overlapping high affinity reagents all measure similar levels across hundreds of samples; (2) levels appear specific to cancer patients, whose tumors overexpress ORF1p; (3) they correlate strongly with measurements

made by targeted proteomics, and (4), plasma levels pre- and on/post treatment correlated with therapeutic response. Nonetheless, the low levels of circulating ORF1p makes orthogonal confirmation in larger cohorts by any other method challenging, as even the most sensitive mass spectrometry assays have limits of detection orders of magnitude higher.

The results expand our understanding that L1 expression is early and pervasive across carcinomas from multiple organs and high-risk precursor lesions, including dysplastic Barrett's esophagus, which is challenging to diagnose and manage. Circulating ORF1p shows promise in early detection applications such as in ovarian cancer and may be more useful as part of a multi-analyte detection test combined with, for example, cfDNA methylation, longitudinal CA125 in ovarian cancer, or CEA in colorectal cancer(3,5,27). We demonstrate that ORF1p is an early indicator of chemotherapeutic response in gastric and esophageal cancers at timepoints as short as 26 days, where other parameters are often ambiguous, opening possibilities for monitoring minimal residual disease or relapse. Shorter time intervals will be needed to understand whether ORF1p can monitor tumor lysis. Importantly, ORF1p appears to provide a level of specificity for cancers not achieved by other protein biomarkers, likely due to the unique biology of the retrotransposon, with repression of L1 in normal somatic tissue(9,13,14). ORF1p is therefore attractive as a putative "binary" cancer biomarker, in which a positive signal is highly specific for disease, with diagnostic utility both in tissue and plasma.

The assays are cost-effective (<$3 in consumables), rapid (<two hours), simple to perform, scalable, and have clinical-grade coefficients of variation (<15%). Flow cytometers for MOSAIC are common in clinical reference laboratories, and the assay could be modified for DNA-based readout by qPCR or sequencing. Limitations of the current work include the relatively small numbers of early-stage samples, a small and heterogeneous gastroesophageal therapeutic cohort, and relatively small, heterogeneous, and mostly late-stage cohorts in survival analysis.

Larger cohorts will be needed for further validation. The results are limited thus far to carcinomas; hematologic, mesenchymal, skin, and central nervous system cancers have not yet been studied. Further optimizations to both assay design and reagents will likely be possible, and larger cohorts are needed to further validate and develop third generation Simoa assays and improved MOSAIC assays, including automation of MOSAIC for scalability. Finally, it is unclear how ORF1p, which is normally cytosolic, enters the blood and what clinicopathologic factors might affect these levels. While senescent and germ cells in humans and mice are known to produce ORF1p (28-30), they may release ORF1p differently than tumor cells or may not release appreciable ORF1p at all. Arguing against significant release from senescent or germ cells, there was no correlation of plasma ORF1p with age or sex in either healthy or cancer patient samples. Future work will also be needed to understand whether there is a normal baseline level of circulating ORF1p, as implied by the trace amounts seen when ORF1p was measured from much larger volumes of plasma using targeted mass spectrometry, and what factors affect this level.

Materials and Methods

Materials. All affinity reagents used in this work are listed in the Supplementary Information (**Supplementary Table S2**). Conjugation reagents, paramagnetic beads, and assay buffers were obtained from Quanterix Corporation. DNA oligos used in the MOSAIC assay were obtained from Integrated DNA Technologies. Antibodies used in final Simoa and MOSAIC assays (monoclonals Ab6, Ab54, 62H12, 34H7) were additionally validated by western blotting (**Supplementary Fig. S15**).

Preparation of capture and detector reagents. All capture antibodies and nanobodies were obtained in or dialyzed into phosphate buffered saline (PBS). For the first-generation Simoa

assay, 7×10^8 carboxylated paramagnetic 2.7-μm beads (Homebrew Singleplex Beads, Quanterix Corp.) were first washed three times with 400 μL Bead Wash Buffer (Quanterix Corp.) and two times with 400 μL cold Bead Conjugation Buffer (Quanterix Corp.) before being resuspended in 390 μL cold Bead Conjugation Buffer. A 1 mg vial of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) (Thermo Fisher Scientific) was then dissolved to 10 mg/mL in cold Bead Conjugation Buffer, and 10 μL was added to the beads. The beads were shaken for 30 minutes at 4°C to activate the carboxyl groups on the beads, which were then washed once with 400 μL cold Bead Conjugation Buffer and resuspended in the capture nanobody solution (10 μg nanobody total), diluted in Bead Conjugation Buffer to a final volume of 400 μL. The beads were shaken for two hours at 4°C, washed twice with 400 μL Bead Wash Buffer, and resuspended in 400 μL Bead Blocking Buffer (Quanterix Corp.) before shaking at room temperature for 30 minutes to block the beads. After one wash each with 400 μL Bead Wash Buffer and Bead Diluent (Quanterix Corp.), the beads were resuspended in Bead Diluent and stored at 4°C. Beads were counted with a Beckman Counter Z Series Particle Counter before using in assays. For second-generation Simoa assays, the following bead coupling conditions were used: 4.2×10^8 starting beads, 300 μL wash volumes, 6 μL EDC, and 40 μg antibody.

For biotinylation of detector antibodies or nanobodies, a 1 mg vial of Sulfo-NHS-LC-LC-biotin was freshly dissolved in 150 μL water and added at 80-fold molar excess to a 1 mg/mL solution of antibody or nanobody. The reaction mixture was incubated at 30 minutes at room temperature and subsequently purified with an Amicon Ultra-0.5 mL centrifugal filter (50K and 10K cutoffs for antibody and dimeric nanobody, respectively). Five centrifugation cycles of 14,000xg for five minutes were performed, with addition of 450 μL PBS each cycle. The purified biotinylated detector reagent was recovered by inverting the filter into a new tube and

centrifuging at 1000xg for two minutes. Concentration was quantified using a NanoDrop spectrophotometer.

Recombinant ORF1p protein production. ORF1p was prepared as described(25); briefly, codon optimized human ORF1p corresponding to L1RP (L1 insertion in X-linked retinitis pigmentosa locus, GenBank AF148856.1) with N-terminal His6-TEV was expressed in E. Coli, purified by Ni-NTA affinity, eluted, tag cleaved in the presence of RNaseA, and polished by size exclusion in a buffer containing 50 mM HEPES pH 7.8, 500 mM NaCl, 10 mM MgCl2, and 0.5 mM tris(2-carboxyethyl) phosphine (TCEP), resulting in monodisperse trimeric ORF1p bearing an N-terminal glycine scar.

Nanobody generation and screening. Nanobodies were generated essentially as decribed(31,32) using mass spectrometry/lymphocyte cDNA sequencing to identify antigen-specific nanobody candidates. Briefly, a llama was immunized with monodisperse ORF1p, and serum and bone marrow were isolated. The heavy chain only IgG fraction (VHH) was isolated from serum and bound to a column of immobilized ORF1p. Bound protein was eluted in SDS and sequenced by mass spectrometry, utilizing a library derived from sequencing VHH fragments PCR-amplified from bone marrow-derived plasma cells. Candidate sequences were cloned into an E. coli expression vector with C-terminal His6 tag and expressed in 50 ml cultures in E. coli Arctic Express RP (Agilent) with 0.2 mM IPTG induction at 12°C overnight. Periplasmic extract was generated as follows: pellets were resuspended in 10 ml per L culture TES buffer (200 mM Tris-HCl, pH 8.0, 0.5 mM EDTA, and 500 mM sucrose), 20 ml/L hypotonic lysis buffer added (TES buffer diluted 1:4 with ddH2O), supplemented with 1 mM PMSF, 3 µg / ml Pepstatin A, incubated 45 min at 4°C, and centrifuged at 25,000 x g for 30 min. The supernatant (periplasmic extract) was bound to ORF1p-conjugated Sepharose, washed 3 times, eluted with SDS at 70°C for 10 min, and periplasmic extract and elution were analyzed by SDS-

PAGE to assay expression and yield. ORF1p-binding candidates were purified as below and analyzed by ELISA (**Supplementary Fig. S7**).

Nanobody and multimeric nanobody purification. C-terminally His6-tagged nanobody constructs were expressed and purified essentially as described(31). Briefly, protein was expressed in E. coli Arctic Express RP (Agilent) with 0.2 mM IPTG induction at 12°C overnight. Periplasmic extract (generated as above) was supplemented with 5 mM $MgCl_2$, 500 mM NaCl, and 20 mM imidazole, purified by Ni-NTA chromatography, dialyzed into 150 mM NaCl, 10 mM HEPES, pH 7.4, and concentrated to 1-3 mg/ml by ultrafiltration. “5xCys tail” constructs were purified with the addition of 5 mM TCEP-HCl in resuspension, wash, elution, and dialysis buffers.

Surface plasmon resonance (SPR) assays. Binding kinetics (k_a, k_d, and K_D**)** of antibody and nanobody constructs for ORF1p were obtained on a Biacore 8K instrument (Cytiva). Recombinant ORF1p was immobilized on a Series S CM5 sensor chip at 1.5 µg/ml using EDC/NHS coupling chemistry according to the manufacturer’s guidelines. Nanobodies and antibodies were prepared as analytes and run in buffer containing 20 mM HEPES pH 7.4, 150 mM NaCl, and 0.05% Tween-20. Analytes were injected at 30 µl/min in single-cycle kinetics experiments at concentrations of 0.1, 0.3, 1, 3.3, and 10 nM, with association times of 120-180 sec, and a dissociation time of 1200-7200 sec, depending on observed off-rate. Residual bound protein was removed between experiments using 10 mM glycine-HCl pH 3.0. Data were analyzed using Biacore software, fitting a Langmuir 1:1 binding model to sensorgrams to calculate kinetic parameters.

For epitope binning, pairs of antibodies were sequentially flowed over immobilized ORF1p using Biacore tandem dual injections according to the manufacturer’s guidelines. Antibodies were injected at concentrations of 200 nM with a flow rate of 10 µl/min. Contact time for the first

antibody was 120 sec, followed by 150 sec for the second antibody, then a 30 sec dissociation. Response signal for the second antibody was measured in a 10 sec window at the beginning of dissociation. The chip was regenerated between experiments with glycine pH 3.0 as above. Data were analyzed using the Biacore software epitope binning module.

ORF1p Simoa assays. Simoa assays were performed on an HD-X Analyzer (Quanterix Corp.), with all assay reagents and consumables loaded onto the instrument according to the manufacturer's instructions. 250,000 capture beads and 250,000 helper (non-conjugated) beads were used in each Simoa assay. A three-step assay configuration was used for the first- and second-generation assays, consisting of a 15-minute target capture step (incubation of capture beads with 100 μL sample), 5-minute incubation with detector reagent (0.3 μg/mL for both first- and second-generation assays), and 5-minute incubation with streptavidin-β-galactosidase (150 pM for first-generation assay; 300 pM for second-generation assays). The beads were washed with System Wash Buffer 1 (Quanterix Corp.) after each assay step. Upon the final wash cycle, the beads were loaded together with the fluorogenic enzyme substrate resorufin β-D-galactopyranoside into a 216,000-microwell array, which was subsequently sealed with oil. Automated imaging and counting of "on" and "off" wells and calculation of average enzyme per bead (AEB) were performed by the instrument. Calibration curves were fit using a 4PL fit with a $1/y^2$ weighting factor, and the limit of detection (LOD) was determined as three standard deviations above the blank.

All plasma and serum samples were diluted four-fold in Homebrew Sample Diluent (Quanterix Corp.) with 1x Halt Protease Inhibitor Cocktail (ThermoFisher), with an additional 1% Triton-X 100 added in the second-generation assays. All recombinant ORF1p calibrators were run in triplicates, with four replicates for the blank calibrator, and all plasma and serum samples were run in duplicates. The average LOD across all sample runs was determined for each assay and depicted in each figure.

Healthy individual plasma and serum samples were obtained from the Mass General Brigham Biobank, with additional samples from the Penn Ovarian Cancer Research Center and Tomas Mustelin (University of Washington). Additional breakdown of patients within each cancer type, by demographic and clinicopathological variables, where available, is included in **Supplementary Figs. S2,S3, S7, S8, and S22 , and Supplementary Table S12**.

ORF1p large-volume MOSAIC assays. MOSAIC assays were performed as previously described, using 2 ml microcentrifuge tubes for the initial capture step. For each sample, 500 μL plasma was diluted four-fold in Homebrew Sample Diluent with protease inhibitor and 1% Triton-X 100 to a total volume of 2 mL. Briefly, 100,000 capture beads were incubated with sample and mixed for two hours at room temperature, followed by magnetic separation and resuspended in 250 μL System Wash Buffer 1 before transferring to a 96-well plate. The beads were then washed with System Wash Buffer 1 using a Biotek 405 TS Microplate Washer before adding 100 μL nanobody detector reagent (0.3 μg/mL, diluted in Homebrew Sample Diluent) and shaking the plate for 10 minutes at room temperature. After washing with the microplate washer, the beads were incubated with 100 μL streptavidin-DNA (100 pM, diluted in Homebrew Sample Diluent with 5 mM EDTA and 0.02 mg/mL heparin) with shaking for 10 minutes at room temperature, followed by another washing step. The beads were transferred to a new 96-well plate, manually washed with 180 μL System Wash Buffer 1, and resuspended in 50 μL reaction mixture for rolling circle amplification (RCA). The RCA reaction mixture consisted of 0.33 U/uL phi29 polymerase, 1 nM ATTO647N-labeled DNA probe, 0.5 mM deoxyribonucleotide mix, 0.2 mg/mL bovine serum albumin, and 0.1% Tween-20 in 50 mM Tris-HCl (pH 7.5), 10 mM $(NH_4)_2SO_4$, and 10 mM $MgCl_2$. The beads were shaken at 37°C for one hour, followed by addition of 160 μL PBS with 5 mM EDTA and 0.1% Tween-20. After washing the beads once with 200 μL of the same buffer, the beads were resuspended in 140 μL buffer with 0.2% BSA. All samples were analyzed using a NovoCyte flow cytometer (Agilent) equipped with three lasers.

Analysis of average molecule per bead (AMB) values was performed as previously described using FlowJo software (BD Biosciences) and Python. All code used for MOSAIC data analysis can be downloaded as part of the waltlabtools.mosaic Python module, which is available at https://github.com/tylerdougan/waltlabtools.

Targeted proteomics analysis of immunoprecipitated ORF1p. Protein levels the LINE-1 ORF1p (UniProt ID: Q9UN81) were determined with targeted proteomics using isotopically-labeled standard peptides (AQUA QuantProHeavy peptides with $^{13}C^{15}N$-labeled C-terminal lysine or arginine, Thermo Fisher) for accurate quantification. Assays were developed for two quantotypic peptides of ORF1p, namely LSFISEGEIK and cysteine-alkylated NLEECITR (the approach is similar to the assay development described previously for other proteins(33)). Briefly, 3-6 mL patient plasma was diluted with an equal volume of 2x dilution buffer (PBS containing 2% Triton X-100, 10 mM EDTA, and 1 Pierce protease inhibitor tablet per 25 ml (2x concentration, Thermo) for a final concentration of 1% Triton X-100, 5 mM EDTA, and 1x protease inhibitor and bound to 7 million 62H12-conjugated magnetic beads for 1 hour at room temperature. Beads were washed 3 times with 5x PBS containing 0.1% tween 20 and 1x protease inhibitor, then once with the same buffer lacking tween 20, and eluted in 50 µl buffer containing 2% SDS and 50 mM Tris pH 8.5 by heating for 5 minutes at 95°C with agitation. Separated eluates were subjected to in-gel digestion using trypsin (150 ng sequencing grade modified trypsin V5111; Promega) after reduction with 10 mmol/L dithiothreitol and alkylation with 55 mmol/L iodoacetamide proteins, prior to LC-MS analyses of the target peptides(33).

Classification models. Classification models were trained for (1) all healthy and all ovarian cancer patients measured by the second-generation assays; and (2) the subset of 51 ovarian cancer and 50 age-matched healthy female patients, obtained from Ronny Drapkin (University of Pennsylvania). Each dataset contained no missing values, and the measurements in the datasets were log-transformed and normalized beforehand for classification analysis of healthy

and ovarian cancer subjects. Logistic regression was used for the univariate classifier and the k-nearest neighbors (KNN) and light gradient-boosting machine (LightGBM), which had the best performances among the classifiers, were used for the multivariate classifier, and implemented in Python 3.7.15 with scikit-learn version 1.0.2 package. Each classifier was given a weight optimization between classes to deal with data imbalance between healthy and cancer subjects, as well as hyperparameter tuning using grid search.

The performance of each biomarker in differentiating ovarian cancer subjects from healthy subjects was evaluated with fivefold cross validation by calculating accuracy, precision, recall, f1-value, sensitivity, specificity, and area under the receiver-operating characteristic (ROC) curve (AUC). A stratified five-fold cross-validation strategy randomly splits the positive and negative samples into five equally sized subsets. One positive subset and one negative subset were selected as the test dataset each time, and the other samples were used to train a classification model.

In the multivariate analysis, the Variance Inflation Factor (VIF) for the biomarkers was calculated, and any biomarkers with extremely high correlation with VIF greater than 10 were excluded from the classification model in advance.

Barrett's esophagus cases. A cohort of 75 esophageal biopsies with BE and varying degrees of dysplasia were assembled. Negative cases were screened to have no prior history of dysplasia. The mean age of the cohort was 67 years with a male predominance (M:F ratio = 3.7:1). All samples were re-analyzed for histological features of dysplasia by three experienced gastrointestinal pathologists (LRZ, VD, OHY) who were blinded to the original diagnosis. A consensus was reached for 72 cases and the consensus diagnosis was used as the gold standard. There was moderate agreement between pathologists (kappa 0.43-0.51).

Colon cancer tissue microarray. 178 sequential CRCs resected by a single surgeon from 2011-2013 were assembled on a 3 mm core tissue microarray. All cases were independently

scored by two pathologists. The mean age of the cohort was 65 years with 49.8% males. Mean follow-up was 25 months. At resection, 23% were stage I, 33% were stage II, 44% were stage III, and 1% were stage IV.

Ovarian Cancer Samples. Age-matched ovarian cancer (n=53) and healthy control (n=50) patient plasma samples were from University of Pennsylvania Ovarian Cancer Research Center, OCRC Tumor BioTrust Collection, Research Resource Identifier (RRID): SCR_02287.

Gastroesophageal cancer treatment cohort. Nineteen patients received systemic therapy, 3 of which also underwent surgical resection. Patients were treated with concurrent chemotherapy (carboplatin/taxol) and radiation (N=3), fluorouracil/ leucovorin/ oxaliplatin/ docetaxel (FLOT, N=2), fluorouracil/ leucovorin/ irinotecan/ oxaliplatin (FOLFIRINOX, N=2), fluorouracil/ leucovorin/ oxaliplatin (FOLFOX, N=9), FOLFOX + trastuzumab (N=1), pembrolizumab (N=1) or FOLFOX then chemoradiation (1). The mean age of the cohort was 76 years. All patients were male (100%). Fifty-eight percent had locally advanced disease (stage II-III) and 42% had advanced disease (stage IV) at the time of initial diagnosis. Sixty-eight percent (N=13) were deemed Responders to therapy while 32% (N=6) were deemed Non-Responders to standard therapy on review of re-staging imaging (CT and/or PET-CT) by investigators blinded to the assay results. Note that the on/post-treatment blood draw measured by Simoa often preceded these imaging studies.

Patient Consent. All plasma samples were obtained with informed written consent under IRB approved protocols at Mass General Brigham (MGB), University of Pennsylvania, and University of Washington. All experiments with patient samples were conducted under IRB approval and in accordance with ethical guidelines in the Belmont Report. Tissue samples were obtained with consent, or, where appropriate, with waiver of consent under MGB approved protocols.

Histochemistry: ORF1p immunohistochemistry was performed essentially as described using anti-ORF1 4H1 (Millipore)(34) diluted 1:3000 and re-optimized on a Leica Bond

system(17). Cases were scored by three experienced gastrointestinal pathologists (MST, VD, OHY) at two institutions. **LINE-1 in situ hybridization** was performed as described using RNAscope catalog 565098 (Advanced Cell Diagnostics) on a Leica Bond system(17). The probe is complementary to the 5' end of L1RP (L1 insertion in X-linked retinitis pigmentosa locus). Cases were scored by three experienced gastrointestinal pathologists (MST, VD, OHY).

Survival Analysis: Kaplan-Meier (KM) curves(35) were computed to study the association between overall survival and plasma ORF1p concentration in ovarian cancer, colorectal cancer and esophageal cancer. To investigate the association with survival, we classified ORF1p concentrations in two different ways. First, by classifying each of the three assays as positive if the signal was above the limit of detection (LoD)- in at least two out of three assays (majority vote method). Second, we evaluated whether ORF1p concentration measured by the most sensitive assay (62H12::Ab6) alone was associated with survival, classifying patients as ORF1p High and Low based on the cohort-specific median. The time variable was defined as days after diagnosis (GE and CRC) or treatment start (ovarian). Living patients were censored at the date of last assessment. Because age at diagnosis was significantly associated with poor prognosis in CRC and male sex was significantly associated with a poor prognosis in GE cancer, we applied a Cox proportional hazards regression model(36); ORF1p was found to be independently prognostic (**Supplementary Tables S10-S11**). Survival objects and KM curves were computed using the survival, ggpubr and survminer packages in R. All tests were performed using R version 4.3.1 (The R Project for Statistical Computing, https://www.R-project.org/). The proportional hazard assumption was tested by plotting the Schoenfeld residuals and applying the Grambsch-Therneau test using the ggcoxdiagnostics function in R. The effect of influential observations was assessed by plotting the Deviance residuals using the ggcoxdiagnostics function in R. Original data for survival are provided in a file "Supplementary Original Survival Data".

Data Availability. Data were generated by the authors and included in the article and its supplementary data files. Survival and related assay data are provided in a supplementary file.

Acknowledgements

We mourn the loss of our dear colleague, Lawrence R. Zukerberg, who died unexpectedly during preparation of this manuscript and made key contributions to our studies of ORF1p in tissue. We thank Bert Vogelstein for plasma samples from colorectal cancer patients. We are grateful to Phil Cole for resources for protein expression and purification and helpful discussions and to Andrew Kruse and Edward Harvey for helpful discussions regarding nanobodies. We thank Zuzana Tothova for helpful discussion and review of the manuscript. This work was supported by the National Institutes of Health grants R01GM130680 (KHB), K08DK129824 (MST), F32EB029777 (CW), R01CA240924 (DTT), U01CA228963 (DTT), P41 GM109824 (MPR, BTC), T32CA009216 (MST, GE), U01CA233364, U2CCA271871, U01CA152990 (SJS), R01GM126170 (JL), P50CA228991 Ovarian SPORE (EJ, T-LW, I-MS, RD), P50CA240243 ovarian SPORE (UAM), P30CA006516 Incubator (MST, DTT, BR, SJS, SJK, KHB); Break *Through* Cancer (KHB); Earlier.Org (KHB and DRW); Minnesota Ovarian Cancer Alliance (KHB); DOD W81XWH-22-1-0852 (EJ, RD); Canary Foundation (RD); Gray Foundation (EJ, RD); The Concord (MA) Detect Ovarian Cancer Early Fund (SJS), Good Ventures (Open Philanthropy Project); Friends of Dana-Farber Cancer Institute; Dana-Farber Cancer Institute; and the Dana-Farber/Harvard Cancer Center (DF/HCC); ACD-Biotechne (DTT, VD); Robert L. Fine Cancer Research Foundation (DTT); Worldwide Cancer Research grant 19-0223 (JL); Robertson Therapeutic Development Fund (JL); Nile Albright Research Foundation (BRR); Vincent Memorial Research Foundation (BRR); Stand Up to Cancer Gastric Cancer Interception Research Team Grant (SU2C-AACR-DT-30-20, SJK, DTT, administered by the American Association for Cancer Research, the Scientific Partner of SU2C).

Author Contributions

MST, CW, ÖHY, SJK, VD, DTT, JL, DRW, and KHB formulated the research plan and interpreted experimental results with assistance from SJZ, TC, LC, YS, JCW, WCC, JH, BDM, and HJ. CW, SZJ, LC, and YS performed Simoa and MOSAIC experiments. WCC, JH, HJ, and BDM performed biochemical experiments. GE performed mouse experiments and interpreted results. PCF, MST, CW, HJ, KRM, BTC, MPR, and JL developed and engineered nanobody constructs. PCF performed SPR affinity measurements. MST, BDM, JCW, and JL designed and performed mass spectrometry experiments. MG, IB, JWF, XY, MET, XW, DC, BEJ, MM, RU, AME, END, LMS, TLW, IMS, EJ, BV, SC, TS, GC, BLN, LP, ARP, MS, UAM, BRR, RD, SJK, and DTT provided patient samples and data and interpreted clinical results. SJS, TC, and KM carried out bioinformatic analysis. MST, LRZ, OHY, and VD diagnosed biopsies, scored cases, and interpreted results. JEF illustrated the manuscript along with MST. MST, CW, DRW, and KHB wrote the manuscript. All authors edited and approved the manuscript.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin **2022**;72(1):7-33 doi 10.3322/caac.21708.
2. Sawyers CL. The cancer biomarker problem. Nature **2008**;452(7187):548-52 doi 10.1038/nature06913.
3. Crosby D, Bhatia S, Brindle KM, Coussens LM, Dive C, Emberton M*, et al.* Early detection of cancer. Science **2022**;375(6586):eaay9040 doi 10.1126/science.aay9040.
4. Ignjatovic V, Geyer PE, Palaniappan KK, Chaaban JE, Omenn GS, Baker MS*, et al.* Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection

to Achieving Translational Data. J Proteome Res **2019**;18(12):4085-97 doi 10.1021/acs.jproteome.9b00503.
5. Jamshidi A, Liu MC, Klein EA, Venn O, Hubbell E, Beausang JF*, et al.* Evaluation of cell-free DNA approaches for multi-cancer early detection. Cancer Cell **2022**;40(12):1537-49 e12 doi 10.1016/j.ccell.2022.10.022.
6. Bast RC, Jr., Lu Z, Han CY, Lu KH, Anderson KS, Drescher CW, Skates SJ. Biomarkers and Strategies for Early Detection of Ovarian Cancer. Cancer Epidemiol Biomarkers Prev **2020**;29(12):2504-12 doi 10.1158/1055-9965.EPI-20-1057.
7. Menon U, Gentry-Maharaj A, Burnell M, Singh N, Ryan A, Karpinskyj C*, et al.* Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. Lancet **2021**;397(10290):2182-93 doi 10.1016/S0140-6736(21)00731-5.
8. Rodic N, Sharma R, Sharma R, Zampella J, Dai L, Taylor MS*, et al.* Long interspersed element-1 protein expression is a hallmark of many human cancers. Am J Pathol **2014**;184(5):1280-6 doi 10.1016/j.ajpath.2014.01.007.
9. Ardeljan D, Wang X, Oghbaie M, Taylor MS, Husband D, Deshpande V*, et al.* LINE-1 ORF2p expression is nearly imperceptible in human cancers. Mob DNA **2020**;11:1 doi 10.1186/s13100-019-0191-2.
10. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ, 3rd*, et al.* Landscape of somatic retrotransposition in human cancers. Science **2012**;337(6097):967-71 doi 10.1126/science.1222077.
11. Helman E, Lawrence MS, Stewart C, Sougnez C, Getz G, Meyerson M. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. Genome Res **2014**;24(7):1053-63 doi 10.1101/gr.163659.113.
12. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J*, et al.* Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet **2020**;52(3):306-19 doi 10.1038/s41588-019-0562-0.
13. Burns KH. Transposable elements in cancer. Nature Reviews Cancer **2017**;17(7):415-24 doi 10.1038/nrc.2017.35.
14. McKerrow W, Kagermazova L, Doudican N, Frazzette N, Kaparos EI, Evans SA*, et al.* LINE-1 retrotransposon expression in cancerous, epithelial and neuronal cells revealed by 5' single-cell RNA-Seq. Nucleic Acids Res **2023**;51(5):2033-45 doi 10.1093/nar/gkad049.
15. Pisanic TR, 2nd, Asaka S, Lin SF, Yen TT, Sun H, Bahadirli-Talbott A*, et al.* Long Interspersed Nuclear Element 1 Retrotransposons Become Deregulated during the Development of Ovarian Cancer Precursor Lesions. Am J Pathol **2019**;189(3):513-20 doi 10.1016/j.ajpath.2018.11.005.
16. Cohen L, Cui N, Cai Y, Garden PM, Li X, Weitz DA, Walt DR. Single Molecule Protein Detection with Attomolar Sensitivity Using Droplet Digital Enzyme-Linked Immunosorbent Assay. ACS Nano **2020**;14(8):9491-501 doi 10.1021/acsnano.0c02378.
17. Rajurkar M, Parikh AR, Solovyov A, You E, Kulkarni AS, Chu C*, et al.* Reverse Transcriptase Inhibition Disrupts Repeat Element Life Cycle in Colorectal Cancer. Cancer Discov **2022**;12(6):1462-81 doi 10.1158/2159-8290.CD-21-1117.
18. Solyom S, Ewing AD, Rahrmann EP, Doucet T, Nelson HH, Burns MB*, et al.* Extensive somatic L1 retrotransposition in colorectal tumors. Genome Res **2012**;22(12):2328-38 doi 10.1101/gr.145235.112.
19. Ewing AD, Gacita A, Wood LD, Ma F, Xing D, Kim MS*, et al.* Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. Genome Res **2015**;25(10):1536-45 doi 10.1101/gr.196238.115.

20. Scott EC, Gardner EJ, Masood A, Chuang NT, Vertino PM, Devine SE. A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Res **2016**;26(6):745-55 doi 10.1101/gr.201814.115.
21. Cajuso T, Sulo P, Tanskanen T, Katainen R, Taira A, Hanninen UA*, et al.* Retrotransposon insertions can initiate colorectal cancer and are associated with poor survival. Nat Commun **2019**;10(1):4022 doi 10.1038/s41467-019-11770-0.
22. Doucet-O'Hare TT, Rodic N, Sharma R, Darbari I, Abril G, Choi JA*, et al.* LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. Proc Natl Acad Sci U S A **2015**;112(35):E4894-900 doi 10.1073/pnas.1502474112.
23. Katz-Summercorn AC, Jammula S, Frangou A, Peneva I, O'Donovan M, Tripathi M*, et al.* Multi-omic cross-sectional cohort study of pre-malignant Barrett's esophagus reveals early structural variation and retrotransposon activity. Nat Commun **2022**;13(1):1407 doi 10.1038/s41467-022-28237-4.
24. Xia Z, Cochrane DR, Tessier-Cloutier B, Leung S, Karnezis AN, Cheng AS*, et al.* Expression of L1 retrotransposon open reading frame protein 1 in gynecologic cancers. Hum Pathol **2019**;92:39-47 doi 10.1016/j.humpath.2019.06.001.
25. Carter V, LaCava J, Taylor MS, Liang SY, Mustelin C, Ukadike KC*, et al.* High Prevalence and Disease Correlation of Autoantibodies Against p40 Encoded by Long Interspersed Nuclear Elements in Systemic Lupus Erythematosus. Arthritis Rheumatol **2020**;72(1):89-99 doi 10.1002/art.41054.
26. Wu C, Dougan TJ, Walt DR. High-Throughput, High-Multiplex Digital Protein Detection with Attomolar Sensitivity. ACS Nano **2022**;16(1):1025-35 doi 10.1021/acsnano.1c08675.
27. Jacobs IJ, Menon U, Ryan A, Gentry-Maharaj A, Burnell M, Kalsi JK*, et al.* Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. Lancet **2016**;387(10022):945-56 doi 10.1016/S0140-6736(15)01224-6.
28. Wylie A, Jones AE, D'Brot A, Lu WJ, Kurtz P, Moran JV*, et al.* p53 genes function to restrain mobile elements. Genes Dev **2016**;30(1):64-77 doi 10.1101/gad.266098.115.
29. De Cecco M, Ito T, Petrashen AP, Elias AE, Skvir NJ, Criscione SW*, et al.* L1 drives IFN in senescent cells and promotes age-associated inflammation. Nature **2019**;566(7742):73-8 doi 10.1038/s41586-018-0784-9.
30. Simon M, Van Meter M, Ablaeva J, Ke Z, Gonzalez RS, Taguchi T*, et al.* LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient Mice Drives Inflammation. Cell Metab **2019**;29(4):871-85 e5 doi 10.1016/j.cmet.2019.02.014.
31. Fridy PC, Li Y, Keegan S, Thompson MK, Nudelman I, Scheid JF*, et al.* A robust pipeline for rapid production of versatile nanobody repertoires. Nat Methods **2014**;11(12):1253-60 doi 10.1038/nmeth.3170.
32. Mast FD, Fridy PC, Ketaren NE, Wang J, Jacobs EY, Olivier JP*, et al.* Highly synergistic combinations of nanobodies that target SARS-CoV-2 and are resistant to escape. Elife **2021**;10 doi 10.7554/eLife.73027.
33. Wolters JC, Ciapaite J, van Eunen K, Niezen-Koning KE, Matton A, Porte RJ*, et al.* Translational Targeted Proteomics Profiling of Mitochondrial Energy Metabolic Pathways in Mouse and Human Samples. J Proteome Res **2016**;15(9):3204-13 doi 10.1021/acs.jproteome.6b00419.
34. Rodić N, Sharma R, Sharma R, Zampella J, Dai L, Taylor MS*, et al.* Long Interspersed Element-1 Protein Expression Is a Hallmark of Many Human Cancers. The American Journal of Pathology **2014**;184(5):1280-6 doi 10.1016/j.ajpath.2014.01.007.
35. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American statistical association **1958**;53(282):457-81.

36. Cox DR. Regression models and life‐tables. Journal of the Royal Statistical Society: Series B (Methodological) **1972**;34(2):187-202.

Figure Legends

Figure 1. ORF1p expression is early and pervasive in carcinomas. **a**, ORF1p immunostaining in a cohort of 211 colorectal cancers. **b**, Representative BE case: lesional cells overexpress p53, the L1 RNA, and ORF1p. **c**, L1 RNA and ORF1p overexpression across a cohort of 72 consensus BE cases and 51 carcinomas. **d**, Summary overview of current data on ORF1p tissue expression in carcinomas (by IHC); early data are from Rodić et al. (8), large cohort colon and gastroesophageal and small cohorts (n<30) are from this study, large cohort ovarian and uterine are from Pisanic et al. (15) and Zhouchunyang et al. (24).

Figure 2. Highly Specific Detection of Carcinomas with the First-Generation ORF1p Simoa Assay. **a**, Schematic of single-molecule protein detection by Simoa; a first-generation assay is shown. Antibody/nanobody-coated magnetic beads, present in excess relative to target, capture single target ORF1p molecules; in the first-generation assay, beads are conjugated with α-ORF1p capture nanobody 5 (Nb5). Enzyme-labeled α-ORF1p detection reagent (here, an antibody, Ab6) is added, forming an “immunosandwich”, beads are loaded into microwells that each can hold at most one bead, and ORF1p molecules are then digitally detected using a fluorogenic substrate by counting “on” wells. **b**, First-generation ORF1p Simoa detects plasma ORF1p with high specificity across major carcinomas. Pie charts indicate percentage of samples with detectable levels; dashed red line, LOD. **, this patient was thought to be ‘healthy’ at the time of blood donation but was six months later found to have prostate cancer and 19 months later found to have lymphoma.

Figure 3. Improved detection of ORF1p with second-generation assays. **a**, Schematic of affinity reagents used. 34H7 and 62H2 are custom mAbs; Nb5-5LL is an engineered homodimeric

730 nanobody. **b**, 34H7::Nb5-5LL second-generation assay measurements across a multi-cancer
731 cohort. **c**, Ovarian cancer patients with age- and gender-matched controls in first- and second-
732 generation assays; patients are a subset of those in 3b; red dots: stage I disease, orange dots:
733 stage II disease. **d**, ROC curves with single marker ORF1p across all healthy and ovarian
734 cancer patients (top, n=128-132 cancer, 447-455 healthy), and multivariate models for ovarian
735 (bottom, n=51-53 cancer, 50 healthy).
736
737 **Figure 4**. Targeted proteomics measurements of plasma ORF1p from large sample volumes. a,
738 ORF1p measured from two gastric cancer patients using two quantotypic peptides
739 (LSFISEGEIK and NLEECIR, red traces) with internal isotopically labeled standards (blue
740 traces) ; a high-ORF1p cancer patient (1231 pg/ml by Simoa, 3.5 ml plasma used for
741 immunoprecipitation (IP)) and high-ORF1p healthy patient (3.0 pg/ml by Simoa, 5 ml plasma
742 used for IP) are shown with 900 amol standard injected. **b**, Correlation between measured
743 ORF1p by Simoa and targeted proteomics assays; r=0.97 (Simoa vs LSFISEGEIK) and r=0.99
744 (Simoa vs NLEECIR, t test), p<0.0001 for both.
745
746 **Figure 5.** Improved detection of ORF1p with third-generation Simoa assays and with MOSAIC
747 assays**.** **a**, Comparison of 2^{nd} and 3^{rd} generation Simoa assays (25 µL) in 25 mostly
748 undetectable gastroesophageal (GE) cancer and healthy control patients. **b**, Schematic of
749 MOSAIC assays. Captured single molecule “immunosandwiches” are formed analogously to
750 Simoa assays. DNA-conjugated streptavidin enables rolling circle amplification to be carried out,
751 generating a strong local fluorescent signal on the bead surface, and then “on” and “off” beads
752 are quantified by flow cytometry, allowing efficient sampling of larger numbers of capture beads.
753 This results in improved sensitivity and multiplexing capabilities. **c**, 37H7::Nb5-5LL MOSAIC and
754 Simoa assays in 10 previously-undetectable GE cancer and healthy control patients. Red
755 dashed lines indicate analytical limit of detection (LoD) for recombinant ORF1p in buffer. Blue

dashed line in panel c indicates plasma-specific background in large volume MOSAIC assays, which is used to determine positivity in the pie-charts.

Figure 6.ORF1p is an early predictor of response in 19 gastroesophageal (GE) patients undergoing chemo/chemoradiotherapy and is prognostic in GE and colorectal cancers (CRC). Responders and Non-Responders were characterized retrospectively by medical oncologists blinded to the assays results by post-therapy, pre-surgery imaging. **a,** Plasma ORF1p as measured by all three second-generation Simoa assays before and during/post treatment; left panel: Non-Responders have higher pre-treatment ORF1p than Responders (p=0.02, t-test); right panel: ORF1p pre- and on/post therapy classifies Responders and Non-Responders; p<0.0001, Fisher's exact test. **b**, Representative CT and PET-CT from patients in the cohort. The representative Non-Responder has the second-highest plasma ORF1p pre-treatment (25.8 pg/ml), which increased to 43.0 pg/ml at day 28 of FOLFOX therapy (47 days after diagnosis), concomitant with increased sizes and number of hepatic metastases seen on CT at day 61. The representative Responder has the fourth-highest plasma ORF1p value in the cohort of Responders (0.83 pg/ml), which decreased to undetectable at day 26 of CROSS therapy (48 days after diagnosis); the displayed PET-CT is 59 days after initiation of therapy, 31 days after the second ORF1p measurement. **c**, Kaplan-Meier survival analysis of patients categorized as plasma ORF1p-high and ORF1p-low based on the median plasma ORF1p assay value shows significantly longer survival for ORF1p-low patients with GE (stages III-IV, p=0.0017, log rank test) and CRC (all stage IV, p=0.011, log rank test). Shaded regions represent 95% confidence intervals.

Figure 1

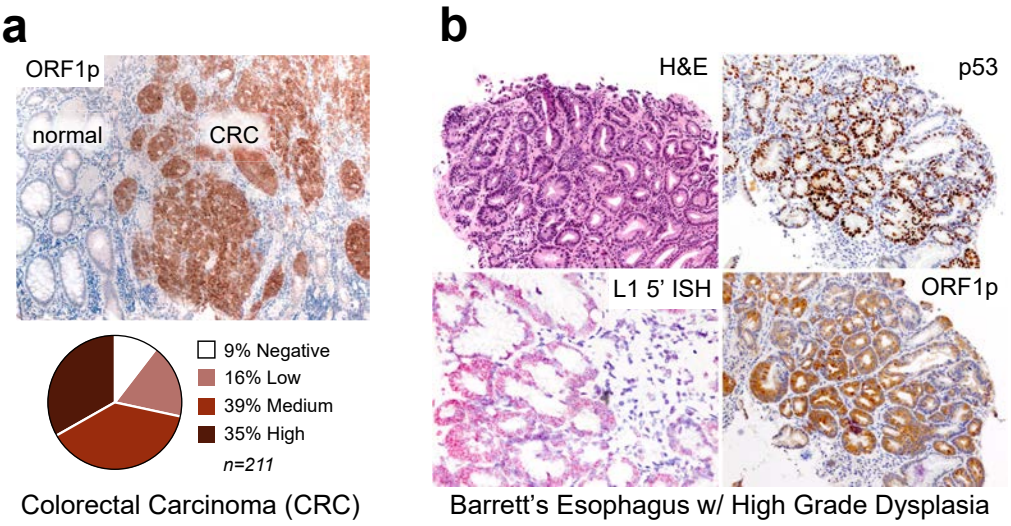
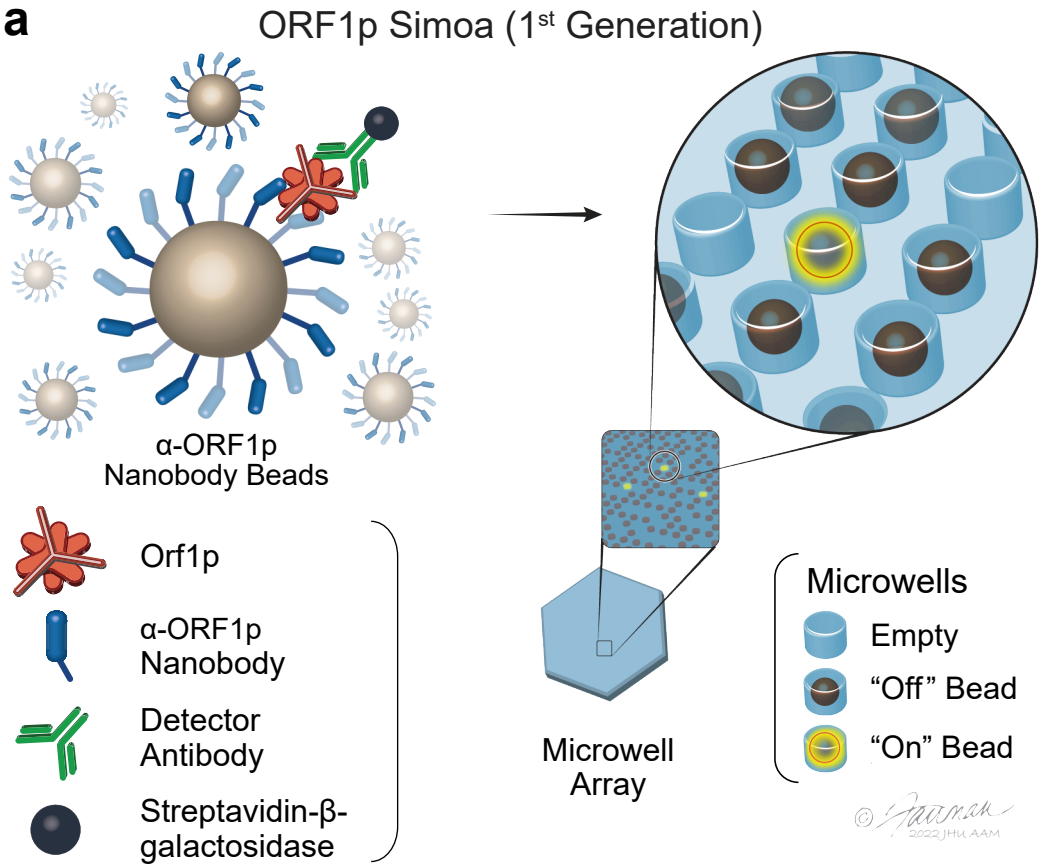


Figure 2



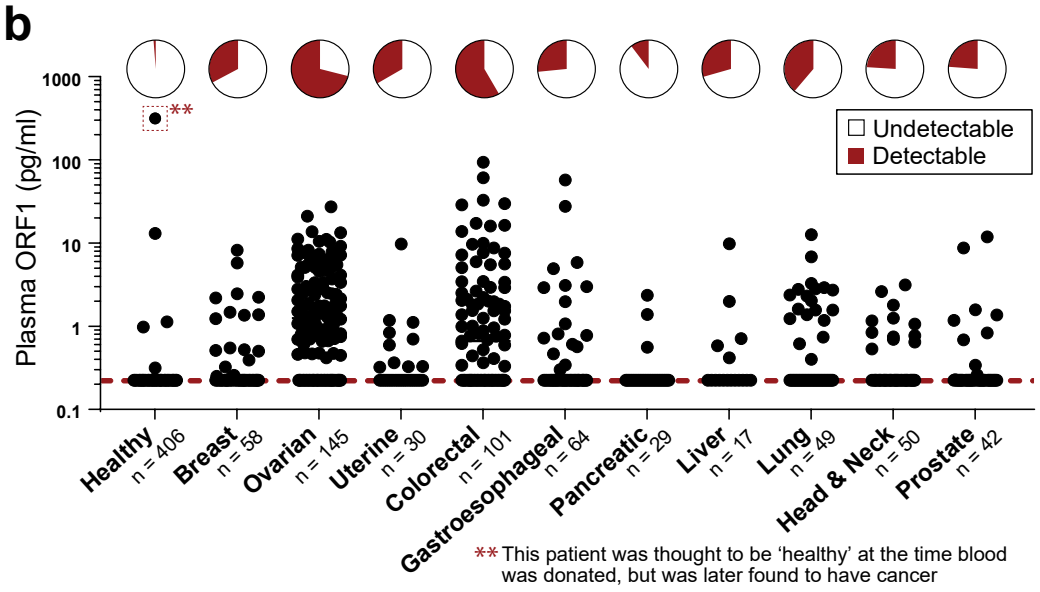


Figure 3

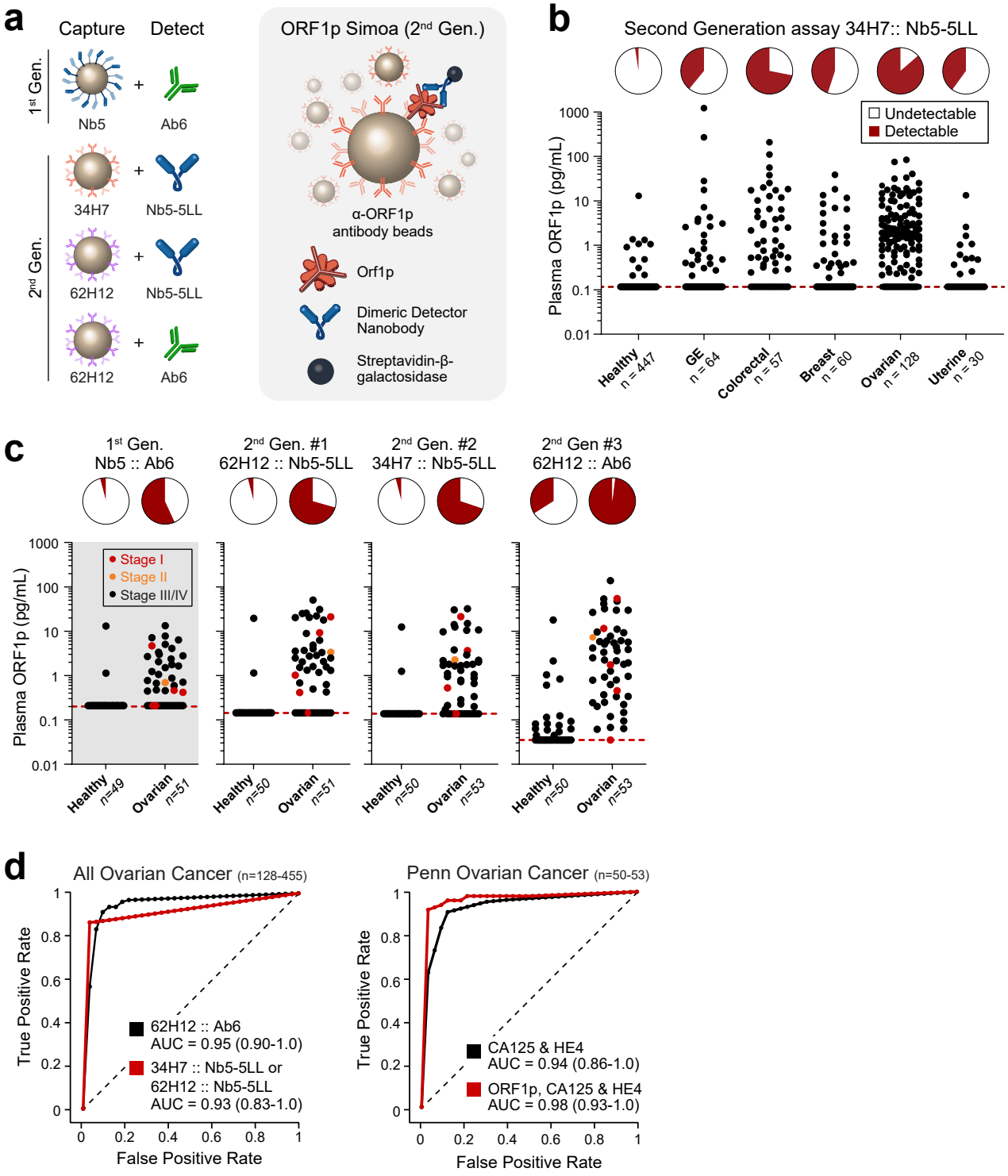
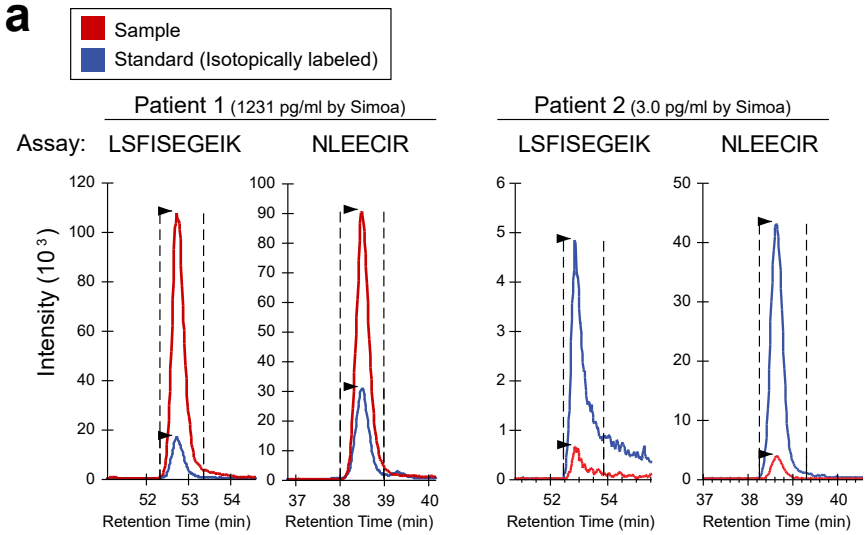


Figure 4



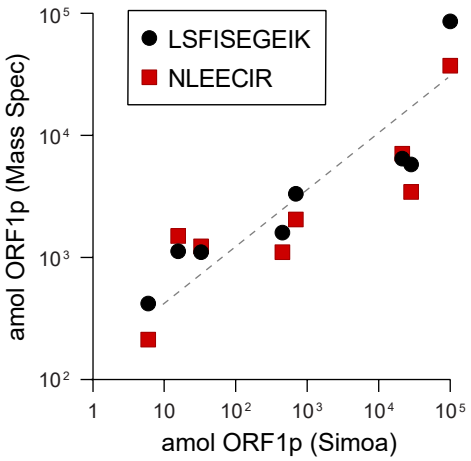


Figure 5

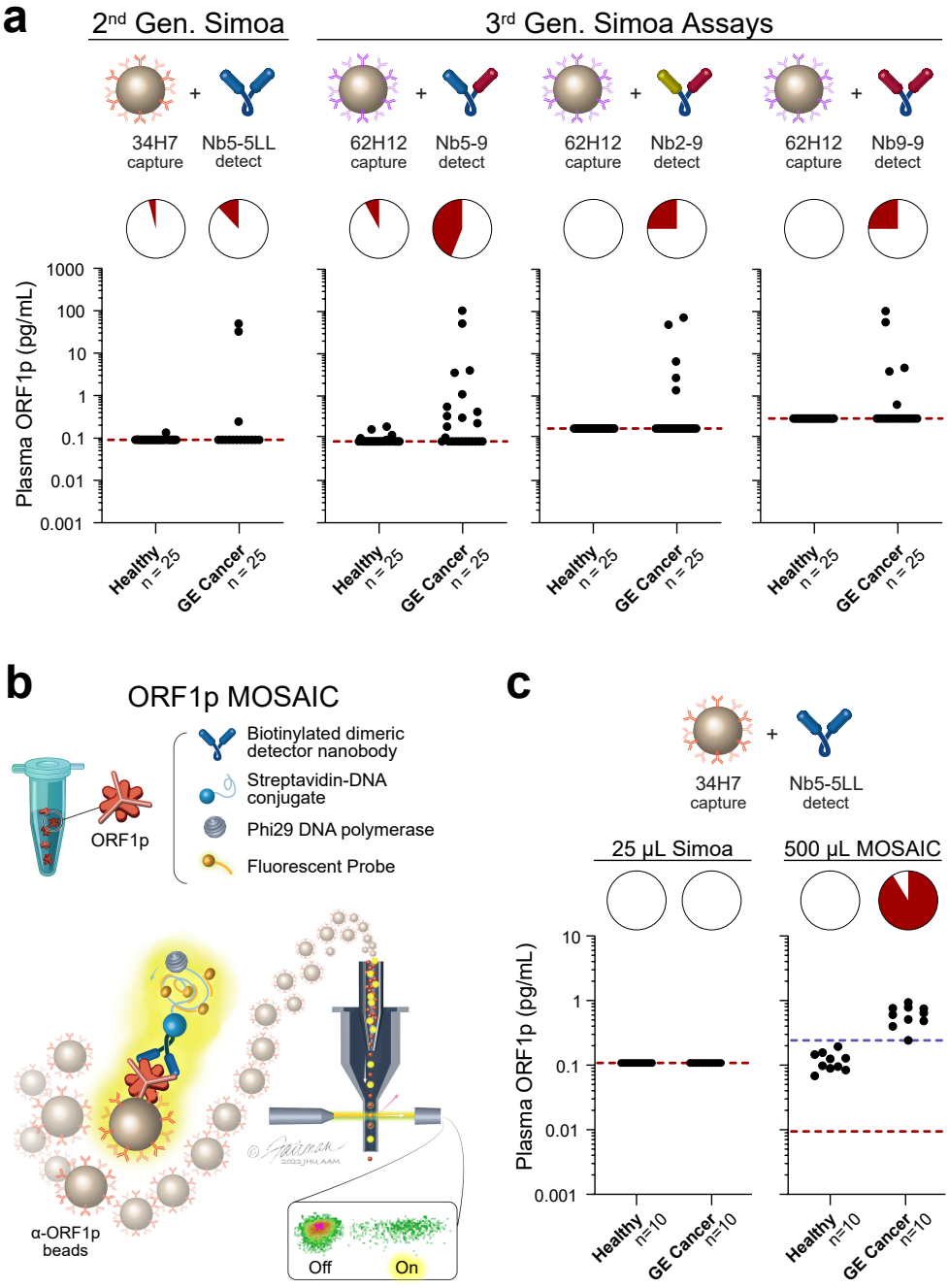


Figure 6