bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

Martin S. Taylor*[,1,‡], Connie Wu*[,2,3,a ‡], Peter C. Fridy[4], Yasmeen Senussi[2,3], Wen-Chih Cheng[5], John Heaps[5], Kei Mori[2,6], Limor Cohen[2,3,7], Kelly R. Molloy[8], Brian T. Chait[8], Michael 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], Euihye Jung[14], Kshitij S. Arora[1], Lawrence R. Zukerberg[1], Osman H. Yilmaz[15], Gary Chi[10], Bryanna L. Norden[10], Yuhui Song[10], Linda Nieman[10], Aparna R. Parikh[10,] Matthew Strickland[10], Tomas Mustelin[16], George Eng[1,17], Ömer H. Yilmaz[1,17], Ursula A. Matulonis[11], Steven J. Skates[18], Bo R. Rueda[19], Ronny Drapkin[14], Samuel J. Klempner[10], Vikram Deshpande[1], David T. Ting[10], Michael P. Rout[4], John LaCava[4,20], David R. Walt[2,3,‡], and Kathleen H. Burns[2,5,‡]

1. Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.
2. Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
3. Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, MA, USA.
4. Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY, USA.
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. Division of Rheumatology, Department of Medicine, University of Washington, Seattle, WA, USA.
17. The David H. Koch Institute for Integrative Cancer Research at MIT, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
18. MGH Biostatistics, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.
19. Department of Obstetrics and Gynecology, Massachusetts General Hospital, and Harvard Medical School, Boston, MA, USA.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

20. 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
Present addresses: [a] University of Michigan Life Sciences Institute, Department of Biomedical Engineering, Ann Arbor, MI, USA.

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, L1) open reading frame 1 protein (ORF1p) is a transposable element protein overexpressed in carcinomas and high-risk precursors during carcinogenesis with negligible detectable expression in corresponding normal tissues, suggesting ORF1p could be a highly specific cancer biomarker. To explore the potential of ORF1p as a blood-based biomarker, we engineered ultrasensitive digital immunoassays that detect mid-attomolar (10^{-17} M) ORF1p concentrations in patient plasma samples across multiple cancers with high specificity. Plasma ORF1p shows promise for early detection of ovarian cancer, improves diagnostic performance in a multi-analyte panel, and provides early therapeutic response monitoring in gastric and esophageal cancers. Together, these observations nominate ORF1p as a multi-cancer biomarker with potential utility for disease detection and monitoring.

Main

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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], although most readily detectable proteins are not sufficiently sensitive at the required high specificity[6] for cancer screening, including CA125 and HE4[7], FDA-cleared markers for the differential diagnosis of pelvic masses, and/or are expressed in normal tissues and therefore lack the requisite specificity.

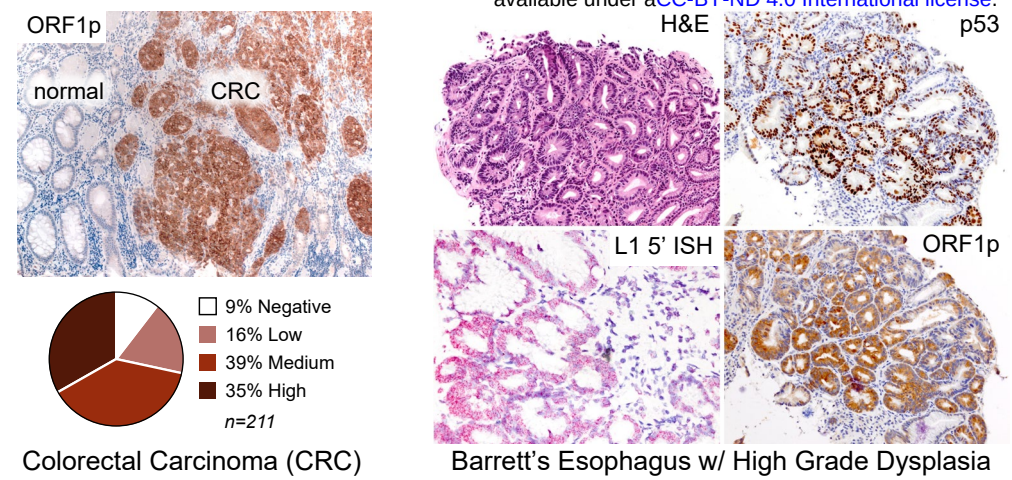
We have previously shown that expression of long interspersed element-1 (L1, LINE-1)-encoded open reading frame 1 protein (ORF1p) is a hallmark of many cancers[8], particularly p53-deficient epithelial cancers. These encompass many of the most commonly occurring and lethal human cancers, including esophageal, 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 begin 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.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 plasma ORF1p may be an early indicator of therapeutic response.

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[18], 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, we stained 211 CRCs [178 sequential cases included on a tissue microarray (TMA) as well as an additional 33 with matched plasma] and found 91% of CRC cases were immunoreactive for ORF1p (**Fig. 1a**). This result is consistent with genetic studies demonstrating somatic L1 retrotransposition in most CRCs[19], including activity in precancerous lesions antedating *APC* tumor suppressor loss[20-22]. Similarly, genetic evidence shows esophageal adenocarcinoma (EAC) has high L1 activity[12], and L1 insertions occur in the highly prevalent Barrett's esophagus (BE) precursor early in carcinogenesis[23,24]. We therefore assembled a cross-sectional cohort of 72 BE cases with consensus diagnosis reached by three expert gastrointestinal pathologists. L1 RNA and ORF1p expression were pervasive in dysplastic BE and present in 100% of 51 esophageal carcinomas (**Fig. 1b,c**); all five BE cases indefinite for dysplasia and positive for ORF1p and/or L1 RNA developed high grade dysplasia on subsequent biopsies (not shown). Overall, this picture is similar to high grade serous ovarian cancers (HGSOC), where ORF1p is expressed in 90% of

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



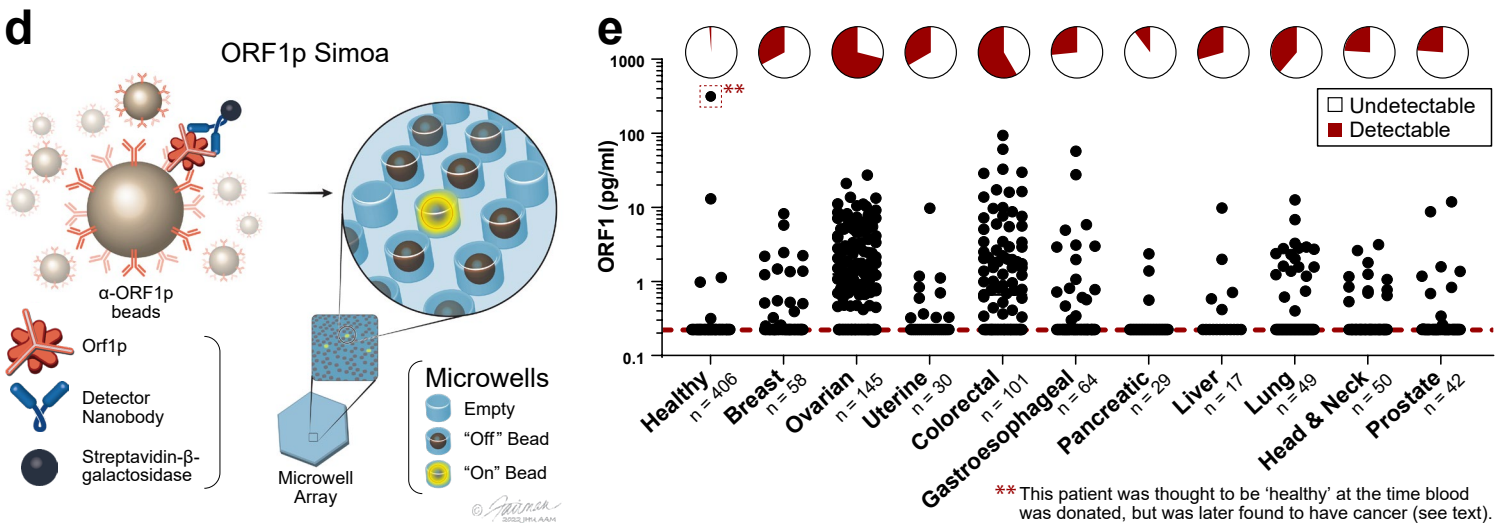


Figure 1. ORF1p expression is early and pervasive in carcinomas. **a**, ORF1p immunostaining in a cohort of 211 colorectal cancers. **b**, Representative BE case: lesional cells overexpress p53, the L1 RNA, and ORF1p. **c**, L1 RNA and ORF1p overexpression across a cohort of 72 consensus BE cases and 51 carcinomas. **d**, Schematic of single-molecule protein detection by Simoa; a second generation assay is shown. Antibody/nanobody-coated magnetic beads, present in excess relative to target, capture single target ORF1p molecules. Enzyme-labeled detection reagent (here, a homodimeric nanobody) 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. First generation Simoa instead uses Nb5-coated beads and Ab6 detector. **e**, 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 control patient was thought to be 'healthy' at the time blood was donated to the biobank but was later found to have prostate cancer and lymphoma.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

cases and 90% of fallopian tube precursor lesions (serous tubal intraepithelial carcinomas, STICs)[8,15,25]. Taken together, ORF1p tissue expression is highly prevalent in gastrointestinal and gynecologic carcinomas and high-risk precursor lesions.

We next sought to extend our tissue findings and explore plasma ORF1p. We optimized our previously reported ORF1p Simoa assay and assessed the landscape of ORF1p levels in pretreatment plasma from patients with advanced cancers. This “first-generation” assay uses a recombinant, single-domain camelid nanobody (Nb5) as the capture reagent and a monoclonal antibody (Ab6) as the detector reagent and has a limit of detection of 0.056 pg/mL (~470 aM trimeric ORF1p), corresponding to 1.9 fM in plasma after correcting for sample dilution (**Fig. 1d, Table S1**). With this assay, we surveyed multiple cancer types and >400 ‘healthy’ control individuals, who were without known cancer at the time blood was donated to the biobank. Plasma ORF1p appears to be a highly specific cancer biomarker, with undetectable levels in ~99% of controls (ages 20-90, **Fig. 1e, S1**). Of the five control patients with detectable ORF1p, the one with the highest ORF1p was later found to have advanced prostate cancer and 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, **Fig. S2-S4**). Notably, four of eight stage I ovarian cancers in the cohort were positive (**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 (**Fig. S5**)[26]. Detectable ORF1p was seen in 1 of 30 patients with

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

chronic liver disease; the one positive patient was subsequently diagnosed with hepatocellular carcinoma (**Fig. S5**). Size exclusion chromatography analysis of patient plasma further showed that the majority of ORF1p resides outside extracellular vesicles (**Fig. S6**). 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) (**Fig. S7-S11, Table S2-S4**). 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. 2a-c, S11-S15**). Adding detergent further improved performance by limiting bead aggregation and improving bead loading into microwells. These second-generation assays comprised capture::detection pairs of 34H7::Nb5-5LL, 62H12::Nb5-5LL, and 62H12::Ab6, achieving 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.2a-c, Table S1-S6**). Somewhat unexpectedly, analytical sensitivity did not perfectly correspond to clinical sensitivity. While the second-generation assays demonstrated less than an order-of-magnitude improvement in

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

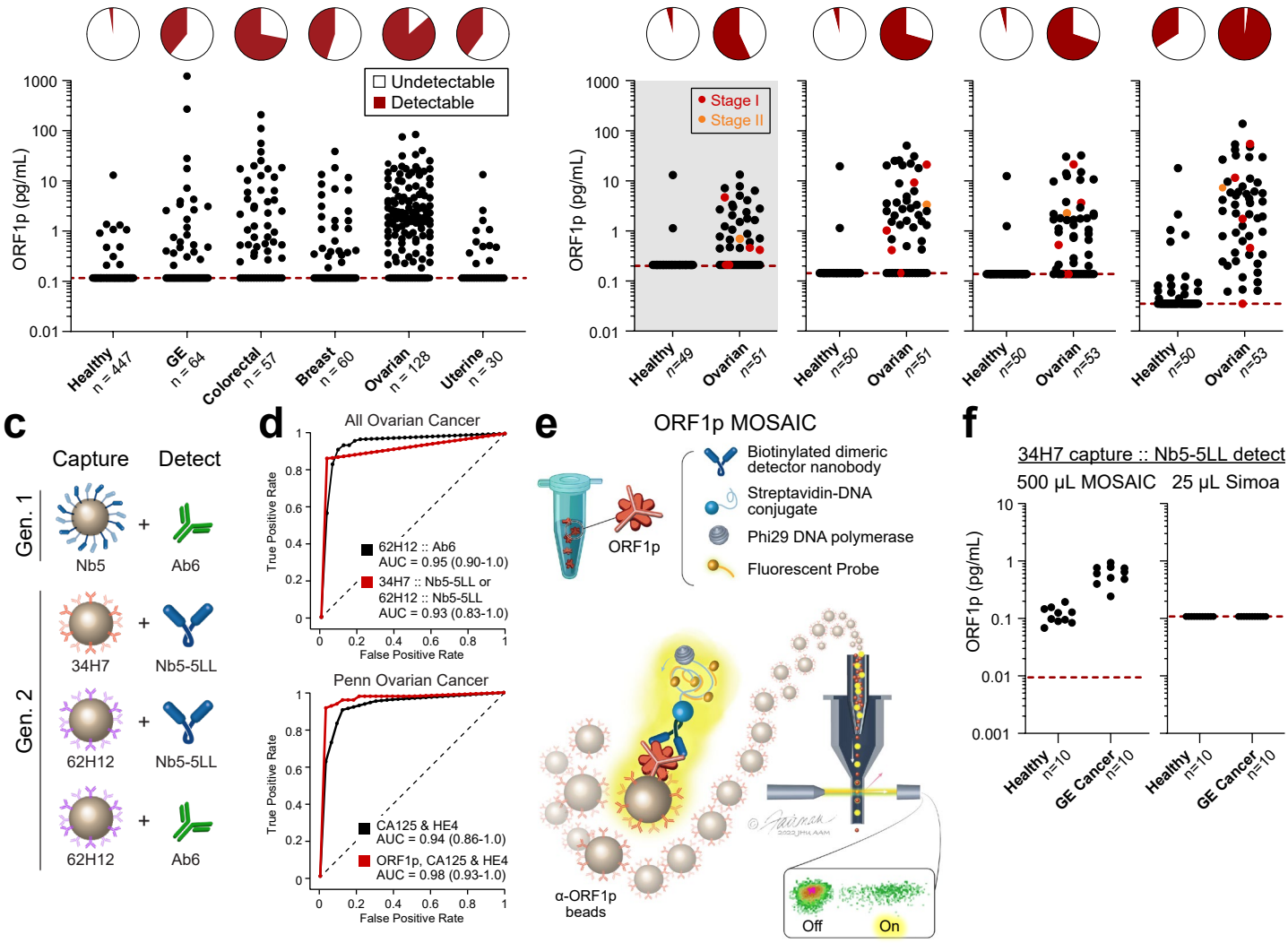


Figure 2. Improved detection of ORF1p with second-generation assays and potential clinical utility. **a**, 34H7::Nb5-5LL second-generation assay measurements across a multi-cancer cohort. **b**, Ovarian cancer patients with age- and gender-matched controls in first- and second-generation assays; patients are a subset of those in 2a; red dots: stage I disease, orange dots: stage II disease. **c**, Schematic of affinity reagents used. 34H7 and 62H2 are custom mAbs; Nb5-5LL is an engineered homodimeric nanobody. **d**, ROC curves with single marker ORF1p across all healthy and ovarian cancer patients (top, n=128-132 cancer, 447-455 healthy), and multivariate models for ovarian (bottom, n=51-53 cancer, 50 healthy). **e**, Schematic of MOSAIC assays. Captured single molecule "immunosandwiches" are formed analogously to Simoa assays. DNA-conjugated streptavidin enables rolling circle amplification to be carried out, generating a strong local fluorescent signal on the bead surface, and then "on" and "off" beads are quantified by flow cytometry. **f**, 37H7::Nb5-5LL MOSAIC and Simoa assays in 10 previously-undetectable gastroesophageal (GE) cancer and healthy control patients. **g**, ORF1p is an early predictor of response in 19 gastro-esophageal patients undergoing chemo/chemoradiotherapy; plasma was measured in all three second-generation Simoa assays before and during/post treatment; p<0.0001, Fisher's Exact test. **h**, Representative CT and PET-CT from patients in the cohort.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 2a, S16**). 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. 2d top panel, Table S7**). This large cohort included pre-treatment plasma samples from ovarian cancer patients (mostly high-grade serous ovarian carcinoma) with age-matched controls (n=51-53 women, **Fig 2b**); 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 2d bottom panel, S17; 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 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.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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. 2e**). 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[27]. 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, **Fig. S18**. 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. 2f**). 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.

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. 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 (Supplement, **Table S9)**. Clinical response ('Responders' and 'Non-Responders') was determined by review of re-staging CT and PET-CT imaging. 12 patients died, six were alive at last follow-up (all 'Responders'), and one was lost to follow-up, over an average of 465 days (range 98-1098). 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

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

imaging (**Fig. 2g**, 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 post-treatment. Representative PET and PET-CT images are shown (**Fig. 2h**). 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.

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, 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. 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; and (3), plasma levels pre- and on/post treatment correlated with therapeutic response. Nonetheless, the low levels of circulating ORF1p makes orthogonal confirmation 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

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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[3,5,28]. We demonstrate that ORF1p is an early indicator of chemotherapeutic response in gastric and esophageal cancers at timepoints where other parameters are often ambiguous, opening possibilities for monitoring minimal residual disease or relapse. 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 and a small and heterogeneous gastroesophageal therapeutic cohort. Larger cohorts will be needed for further validation. Further optimizations to both assay design and reagents will likely be possible. Finally, it is unclear how ORF1p, which is normally cytosolic, enters the blood and what clinicopathologic factors might affect these levels. Future work will also be needed to understand whether there is a normal baseline level of circulating ORF1p and what factors affect this level.

Acknowledgements

We thank Jeni Fairman for illustrations and 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

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

manuscript. This work was supported by the National Institutes of Health grants R01GM130680 (KHB), K08 DK129824-01 (MST), F32EB029777 (CW), R01CA240924 (DTT), U01CA228963 (DTT), P41 GM109824 (MR, BTT), T32CA009216 (MST, GE);), P50CA228991 Ovarian SPORE (EJ, T-LW, I-MS, RD); U01CA233364, U2CCA271871, U01CA152990 (SJS); R01GM126170 (JL); 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); Nile Albright Research Foundation (BRR); Vincent Memorial Research Foundation (BRR); SU2C 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 LC, YS, WCC, and JH. CW, LC, and YS performed Simoa and MOSAIC experiments. WCC and JH performed biochemical experiments. GE performed mouse experiments. PCF, MST, CW, KRM, BTC, MPR, and JL developed and engineered nanobody constructs. PCF performed SPR affinity measurements. MG, IB, JWF, XY, MET, XW, DC, BEJ, MM, RU, AME, END, LMS, TLW, IMS, EJ, BV, GC, BLN, ARP, MS, UAM, BRR, RD, SJK, and DTT provided patient samples and data and interpreted clinical results. SJS and KM carried out bioinformatic analysis. MST, LRZ, OHY, and VD diagnosed biopsies, scored cases, and interpreted results. MST, CW, DRW, and KHB wrote the manuscript. All authors edited and approved the manuscript.

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

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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.

References

1. Siegel, R.L., Miller, K.D., Fuchs, H.E. & Jemal, A. Cancer statistics, 2022. *CA Cancer J Clin* **72**, 7-33 (2022).
2. Sawyers, C.L. The cancer biomarker problem. *Nature* **452**, 548-552 (2008).
3. Crosby, D.*, et al.* Early detection of cancer. *Science* **375**, eaay9040 (2022).
4. Ignjatovic, V.*, et al.* Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. *J Proteome Res* **18**, 4085-4097 (2019).
5. Jamshidi, A.*, et al.* Evaluation of cell-free DNA approaches for multi-cancer early detection. *Cancer Cell* **40**, 1537-1549 e1512 (2022).
6. Menon, U.*, 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* **397**, 2182-2193 (2021).
7. Bast, R.C., Jr.*, et al.* Biomarkers and Strategies for Early Detection of Ovarian Cancer. *Cancer Epidemiol Biomarkers Prev* **29**, 2504-2512 (2020).
8. Rodic, N.*, et al.* Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am J Pathol* **184**, 1280-1286 (2014).
9. Ardeljan, D.*, et al.* LINE-1 ORF2p expression is nearly imperceptible in human cancers. *Mob DNA* **11**, 1 (2020).
10. Lee, E.*, et al.* Landscape of somatic retrotransposition in human cancers. *Science* **337**, 967-971 (2012).
11. Helman, E.*, et al.* Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res* **24**, 1053-1063 (2014).
12. Rodriguez-Martin, B.*, et al.* Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat Genet* **52**, 306-319 (2020).
13. Burns, K.H. Transposable elements in cancer. *Nature Reviews Cancer* **17**, 415-424 (2017).
14. McKerrow, W.*, et al.* LINE-1 Retrotransposon expression in cancerous, epithelial and neuronal cells revealed by 5' single-cell RNA-Seq. *bioRxiv*, 2021.2001.2019.427347 (2022).
15. Pisanic, T.R., 2nd*, et al.* Long Interspersed Nuclear Element 1 Retrotransposons Become Deregulated during the Development of Ovarian Cancer Precursor Lesions. *Am J Pathol* **189**, 513-520 (2019).
16. Cohen, L.*, et al.* Single Molecule Protein Detection with Attomolar Sensitivity Using Droplet Digital Enzyme-Linked Immunosorbent Assay. *ACS Nano* **14**, 9491-9501 (2020).
17. Rajurkar, M.*, et al.* Reverse Transcriptase Inhibition Disrupts Repeat Element Life Cycle in Colorectal Cancer. *Cancer Discov* (2022).
18. Rajurkar, M.*, et al.* Reverse Transcriptase Inhibition Disrupts Repeat Element Life Cycle in Colorectal Cancer. *Cancer Discov* **12**, 1462-1481 (2022).
19. Solyom, S.*, et al.* Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res* **22**, 2328-2338 (2012).
20. Ewing, A.D.*, et al.* Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res* **25**, 1536-1545 (2015).
21. Scott, E.C.*, et al.* A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res* **26**, 745-755 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

22. Cajuso, T.*, et al.* Retrotransposon insertions can initiate colorectal cancer and are associated with poor survival. *Nat Commun* **10**, 4022 (2019).
23. Doucet-O'Hare, T.T.*, et al.* LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc Natl Acad Sci U S A* **112**, E4894-4900 (2015).
24. Katz-Summercorn, A.C.*, et al.* Multi-omic cross-sectional cohort study of pre-malignant Barrett's esophagus reveals early structural variation and retrotransposon activity. *Nat Commun* **13**, 1407 (2022).
25. Zhouchunyang, X.*, et al.* Expression of L1 retrotransposon open reading frame protein 1 (L1ORF1p) in gynecologic cancers. *Hum Pathol* (2019).
26. Carter, V.*, et al.* High Prevalence and Disease Correlation of Autoantibodies Against p40 Encoded by Long Interspersed Nuclear Elements in Systemic Lupus Erythematosus. *Arthritis Rheumatol* **72**, 89-99 (2020).
27. Wu, C., Dougan, T.J. & Walt, D.R. High-Throughput, High-Multiplex Digital Protein Detection with Attomolar Sensitivity. *ACS Nano* (2022).
28. Jacobs, I.J.*, et al.* Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet* **387**, 945-956 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Supplementary Information

Materials and Methods

Materials. All affinity reagents used in this work are listed in the Supplementary Information (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 (**Figure S22**).

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[1]; 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[2,3] 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

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 (**Figure S7**).

Nanobody and multimeric nanobody purification. C-terminally His6-tagged nanobody constructs were expressed and purified essentially as described[2]. 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 **Figures S2-S3, and S19-21, and Table S10**.

ORF1p large-volume MOSAIC assays. MOSAIC assays were performed as previously described, using Eppendorf tubes for the initial capture step. [insert ACS Nano MOSAIC ref] For each sample, 500 μL plasma

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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.

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).

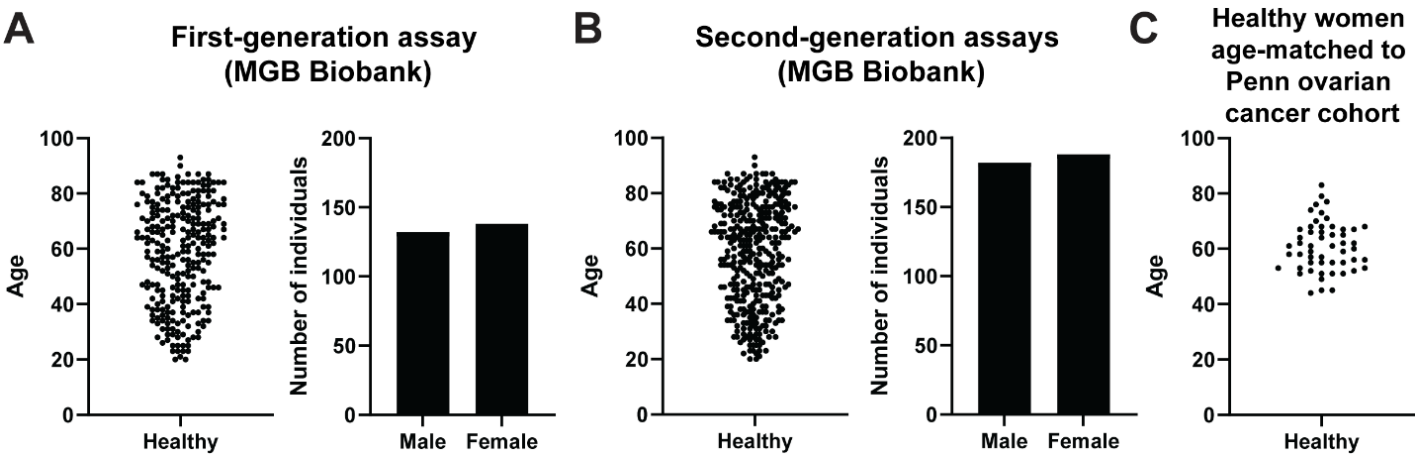
Colon cancer tissue microarray. 178 sequential CRCs resected by a single surgeon from 2011-2013 were assembled on a 3 mm core tissue microarray. All cases were independently scored by two pathologists. The mean age of the cohort was 65 years with 49.8% males. Mean follow-up was 25 months. At resection, 23% were stage I, 33% were stage II, 44% were stage III, and 1% were stage IV.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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.

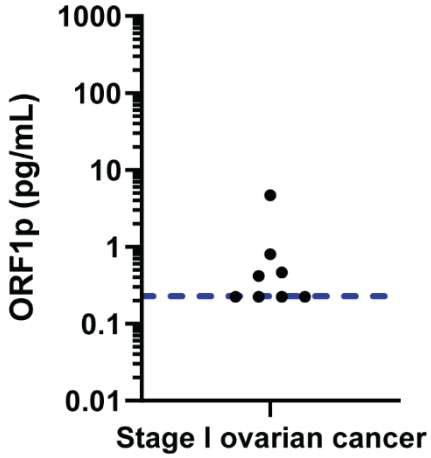
Gastroesophageal cancer treatment cohort. Nineteen patients received systemic therapy, 3 of which also underwent surgical resection. Patients were treated with concurrent chemotherapy (carboplatin/taxol) and radiation (N=3), fluorouracil/ leucovorin/ oxaliplatin/ docetaxel (FLOT, N=2), fluorouracil/ 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.

Histochemistry: ORF1p immunohistochemistry was performed as described using anti-ORF1 4H1 (Millipore)[4] diluted 1:3000 on a Leica Bond system. Cases were scored by three experienced gastrointestinal pathologists (MST, VD, OHY). **LINE-1 in situ hybridization** was performed as described using RNAscope catalog 565098 (Advanced Cell Diagnostics) on a Leica Bond system[5]. 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).

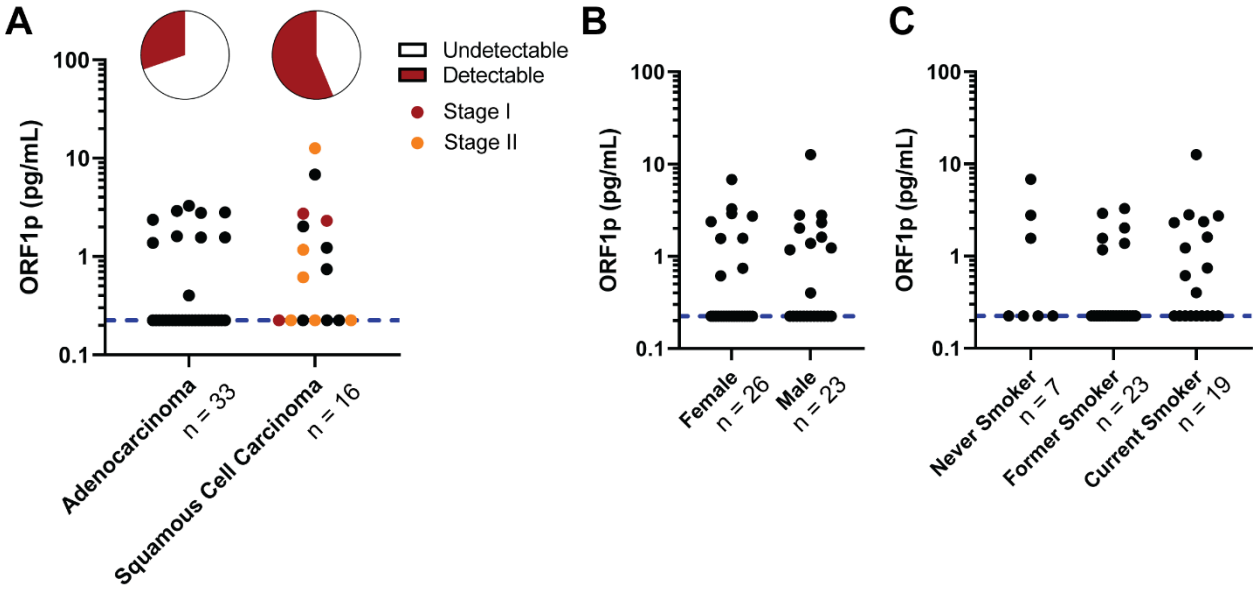


Supplementary Figure 1. Age and gender distributions of healthy control patients used in the first- and second-generation ORF1p Simoa assays from the (A-B) Mass General Brigham Biobank in the first-generation (A) and second-generation (B) ORF1p Simoa assays; and (C) Penn Medicine Biobank. Healthy female control patients from the Penn Medicine Biobank were age-matched with the ovarian cancer patients from the Penn Ovarian Cancer Research Center Tumor BioTrust Collection.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

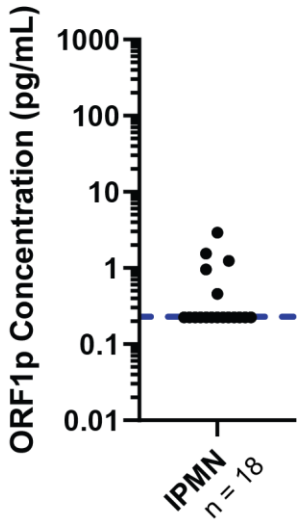


Supplementary Figure 2. Circulating ORF1p levels in eight stage I ovarian cancer patients, as measured by the first-generation Simoa assay. Blue dashed line indicates the assay limit of detection, accounting for the four-fold dilution factor.

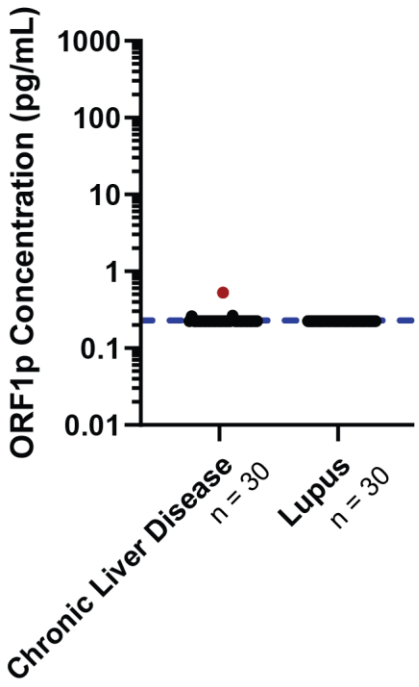


Supplementary Figure 3. Circulating ORF1p levels in lung cancer patient cohort, as classified by (A) disease subtype, (B) gender, and (C) smoking status. Blue dashed line indicates the assay limit of detection, accounting for the four-fold dilution factor.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

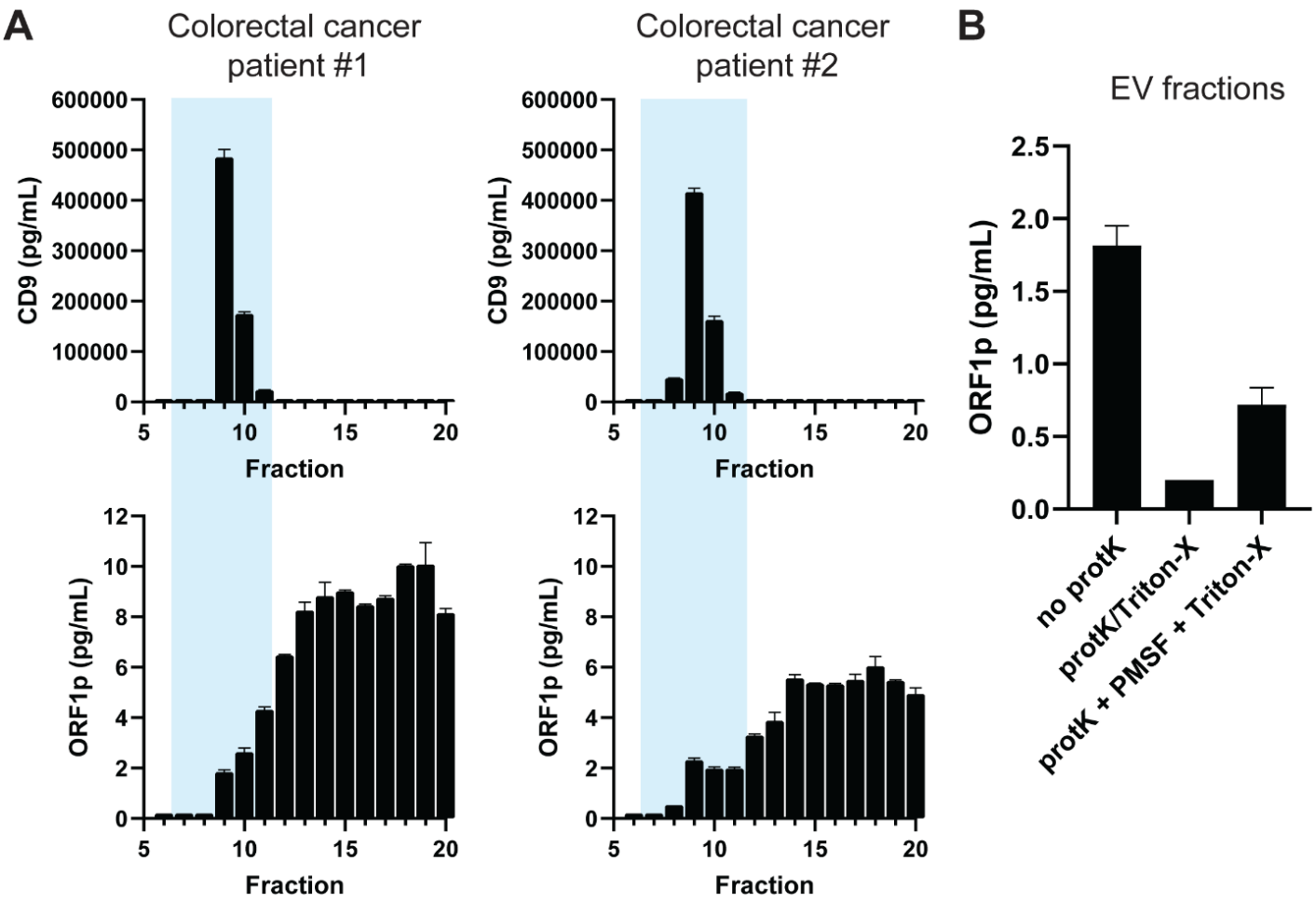


Supplementary Figure 4. Circulating ORF1p levels in patients with IPMN, measured using the first-generation Simoa assay. IPMN, intraductal papillary mucinous neoplasm. Blue dashed line denotes the assay limit of detection, accounting for four-fold dilution.



Supplementary Figure 5. Circulating ORF1p levels in patients with chronic liver disease or systemic lupus erythematous, measured using the first-generation Simoa assay. The red dot represents a chronic liver disease patient diagnosed with hepatocellular carcinoma after the time of sampling. Blue dashed line denotes the assay limit of detection, accounting for four-fold dilution.

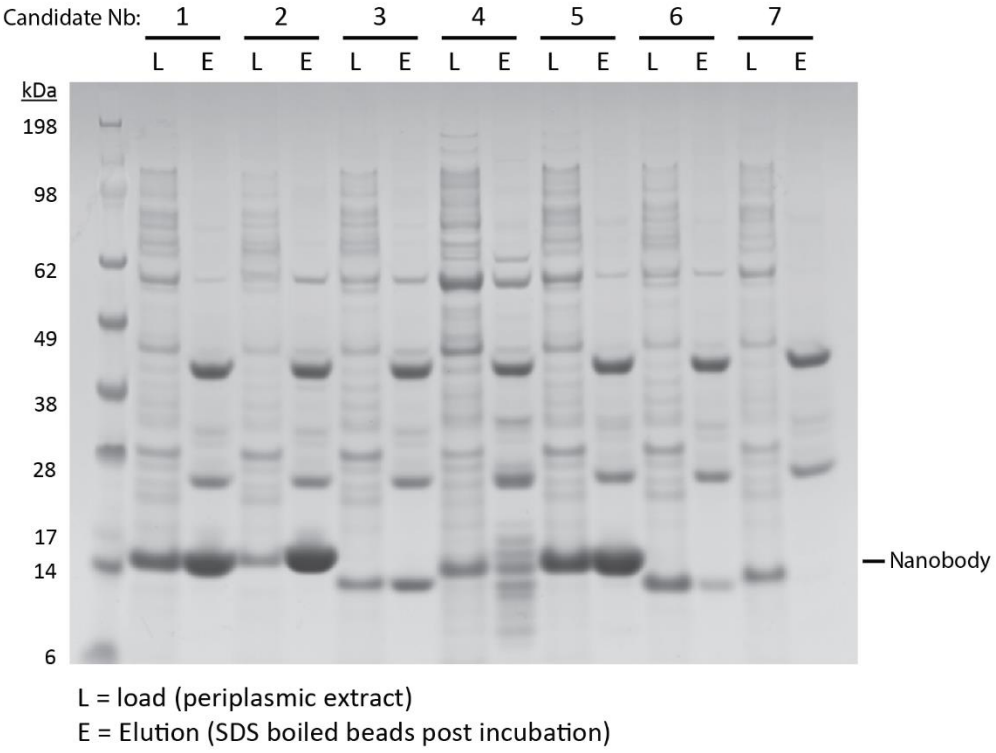
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



Supplementary Figure 6. Analysis of circulating ORF1p levels in size exclusion chromatography (SEC) fractions of colorectal cancer patient plasma. 500 µL plasma was filtered through a 0.45 µm centrifugal filter and fractionated with a Sepharose CL-6B resin packed column (A) CD9 and ORF1p levels in each SEC fraction as measured by Simoa. Blue highlighted boxes denote extracellular vesicle (EV)-containing fractions. The majority of circulating ORF1p is measured in free protein fractions. (B) Determination of ORF1p levels inside EVs via proteinase K (protK) protection assays. EV-containing fractions were pooled and concentrated and treated with proteinase K, followed by the serine protease inhibitor phenylmethylsulfonyl fluoride (PMSF) and 0.5% Triton-X. Controls without proteinase K digestion (no protK) and with simultaneous proteinase K and Triton-X treatment (protK/Triton-X) were performed. Decreased ORF1p concentrations were measured after proteinase K digestion and subsequent EV lysis by Triton-X, further suggesting that only a very small amount of circulating ORF1p is in EVs.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A

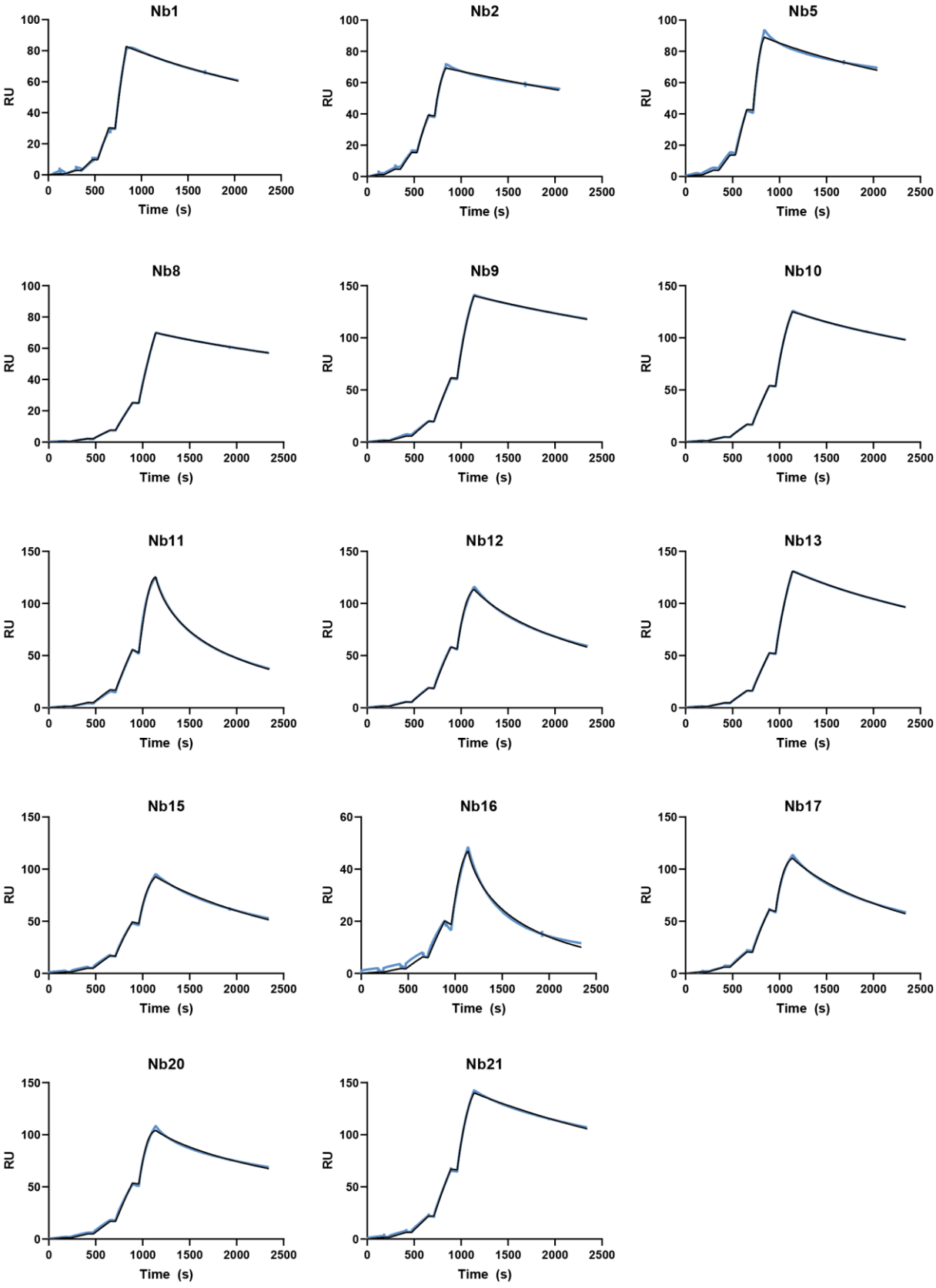


B

	ELISA EC50 (nM)
Nb ORF1-1	**19.4**
Nb ORF1-2	**4.8**
Nb ORF1-3	3.0
Nb ORF1-4	6.0
Nb ORF1-5	**1.8**
Nb ORF1-6	11.4
Nb ORF1-7	NA
Nb ORF1-8	**1.4**
Nb ORF1-9	**1.2**
Nb ORF1-10	**1.2**
Nb ORF1-11	**1.5**
Nb ORF1-12	**1.4**
Nb ORF1-13	**1.1**
Nb ORF1-14	NA
Nb ORF1-15	**0.6**
Nb ORF1-16	**1.0**
Nb ORF1-17	**1.0**
Nb ORF1-18	5.6
Nb ORF1-19	7.9
Nb ORF1-20	**1.8**
Nb ORF1-21	**0.6**

Supplementary Figure 7. Nanobody generation. **(**A) Representative screening results for nanobodies 1-7. Periplasmic extracts from E. Coli expressing the indicated candidate nanobodies (Nb) were bound to ORF1p-conjugated beads and eluted, then analyzed by SDS-PAGE and stained by Coomassie blue. Expressed and purified nanobodies (~15 kDa) are indicated; larger/darker bands indicate higher levels of expression and/or purification. (B) EC50s from ELISAs performed against recombinant ORF1p. Highlighted nanobodies were selected for follow-up characterization, while remainder were rejected for poor affinity and/or protein expression.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



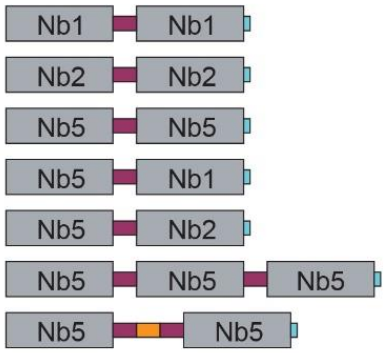
Supplementary Figure 8. Representative SPR sensorgrams of ORF1 nanobodies. Single-cycle kinetics performed over immobilized ORF1, with nanobodies injected sequentially at 0.1, 0.3, 1, 3, and 10nM concentrations. Raw data (blue) and a 1:1 binding model fit (black) are plotted.

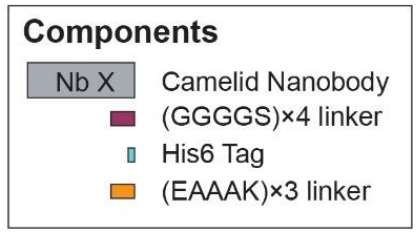
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A

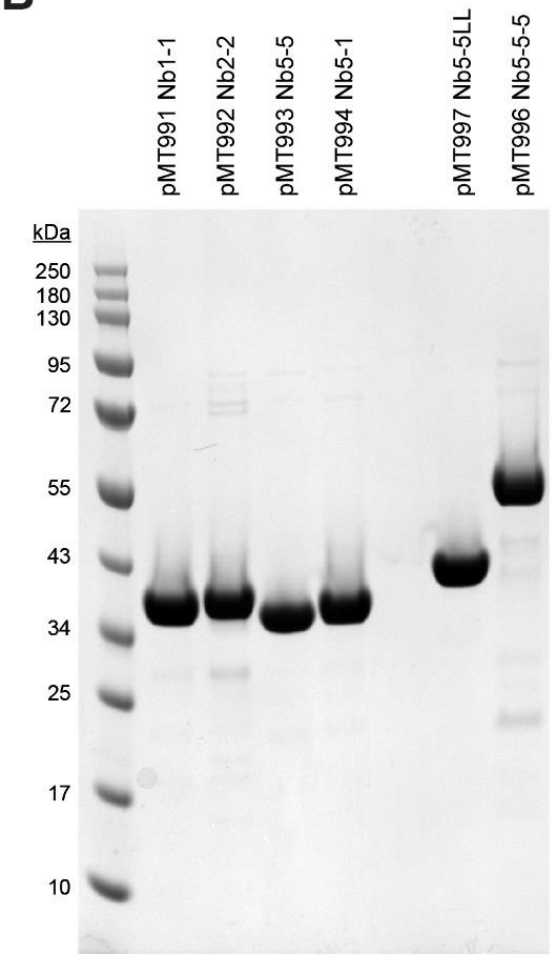
Nanobody Constructs

Construct	Description
MT991	Nb 1-1
MT992	Nb 2-2
MT993	Nb 5-5
MT994	Nb 5-1
MT995	Nb 5-2
MT996	Nb 5-5-5
MT997	Nb 5-5 long "LL"



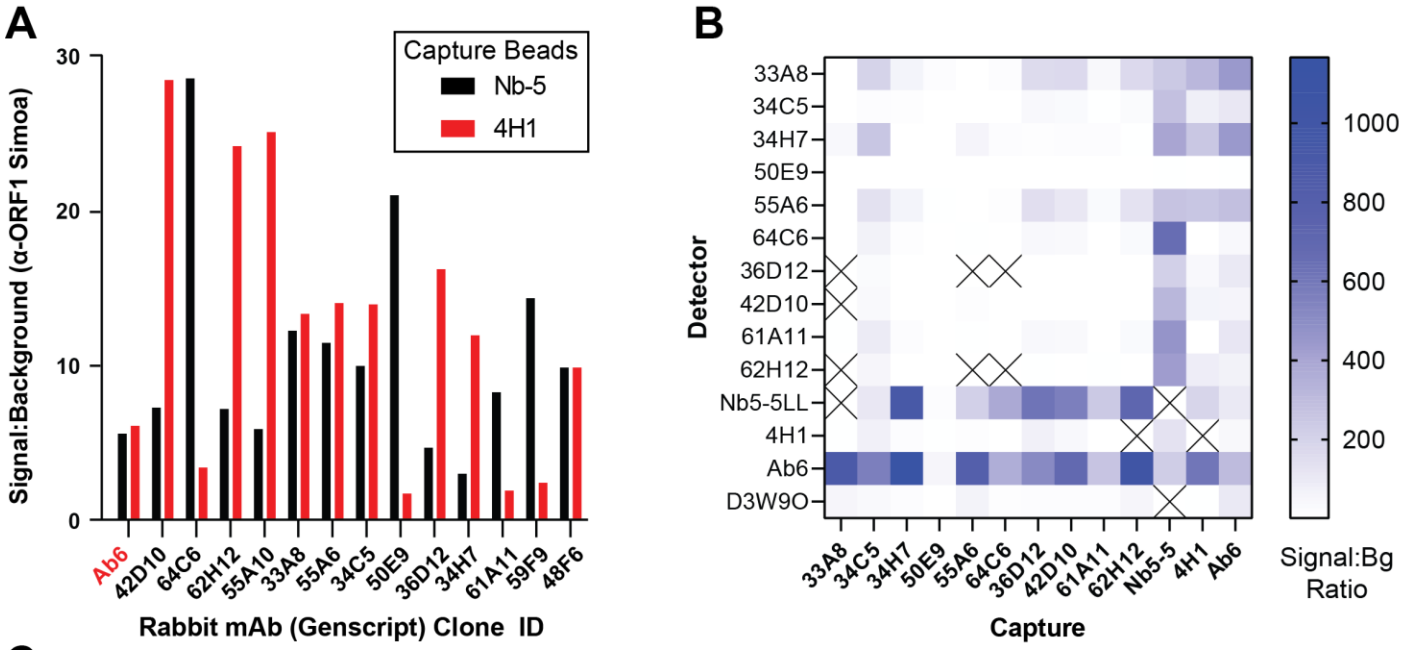


B



Supplementary Figure 9. Engineered Nanobody constructs. (A) Schematic of design of engineered dimeric and trimeric nanobody constructs, with flexible (GGGGS × 4) and rigid helical (EAAAK × 3) linkers. (B), Representative preparation of engineered nanobody constructs, Coomassie stain.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



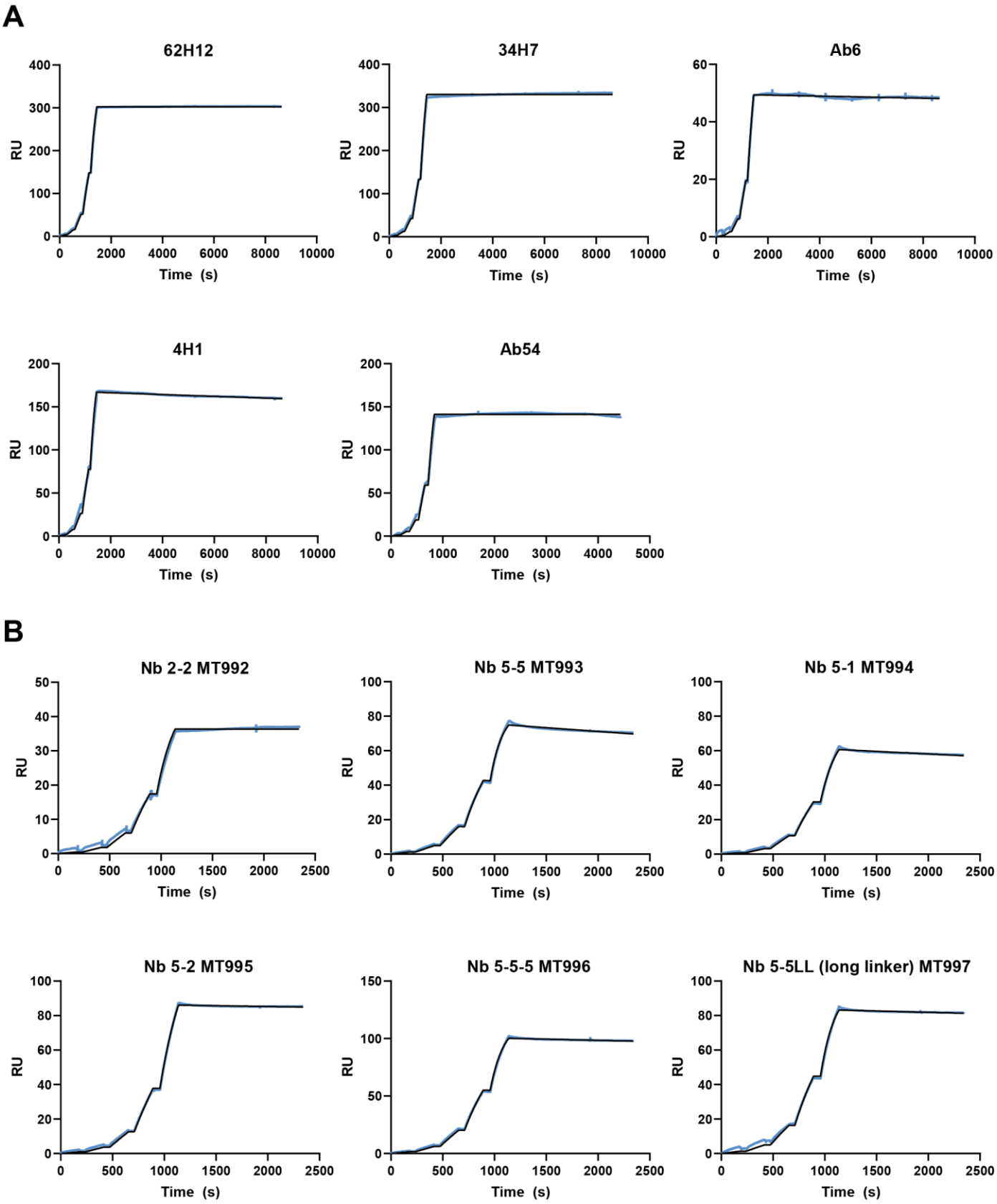
C

		DETECTOR								
		Ab6	Nb5-5LL	33A8	34C5	34H7	64C6	61A11	62H12	55A6
CAPTURE	33A8	0.016								
	34C5	0.026				0.032				
	34H7	0.006	0.043							
	55A6	0.026	0.055							
	64C6	0.034	0.041							
	36D12	0.046	0.050							
	42D10	0.021	0.020							
	61A11	0.042	0.066							
	62H12	0.008	0.062							
	Nb5-5	0.050		0.133	0.022	0.047	0.038	0.042	0.120	
	4H1	0.030	0.056	0.057		0.044				0.036
	Ab6	0.102	0.044	0.063					0.113	

Limit of Detection (pg/mL)

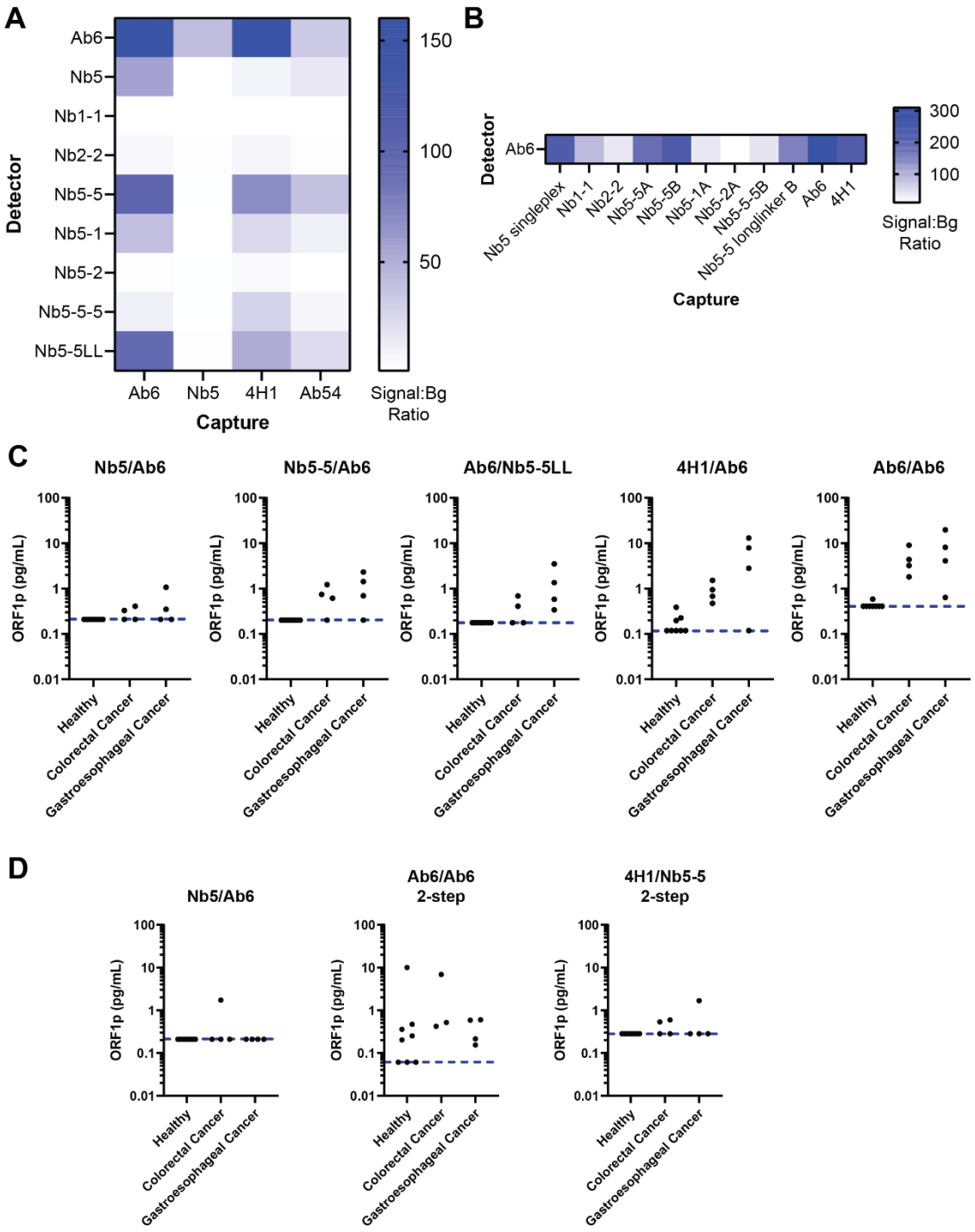
Supplementary Figure 10. Screening of newly developed monoclonal antibody candidates (GenScript). (A) Signal:background ratios of novel rabbit monoclonal α-ORF1p antibodies (B-cell supernatants) in a modified Simoa employing the candidate mAb plus biotinylated secondary anti-Rabbit pAb as detector; these show up to 5-fold improvement vs. our prior best detection antibody, Ab6. Two different capture beads with distinct epitopes were employed. (B-C) Best performers were then synthesized and purified; mAbs were used in further screening with a dimeric nanobody and commercially available monoclonal antibodies for ORF1p detection with Simoa. (B) Signal-to-background comparisons of affinity reagents as capture/detector pairs on Simoa, using recombinant ORF1p protein. All labeled affinity reagents except Nb5-5LL are monoclonal antibodies; Nb5-5LL denotes a homodimeric form of the nanobody Nb5. (C) Limit of detection values for affinity reagent pairs selected from screening in (B). Anti-ORF1 D3W9O is from Cell Signaling Technology.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



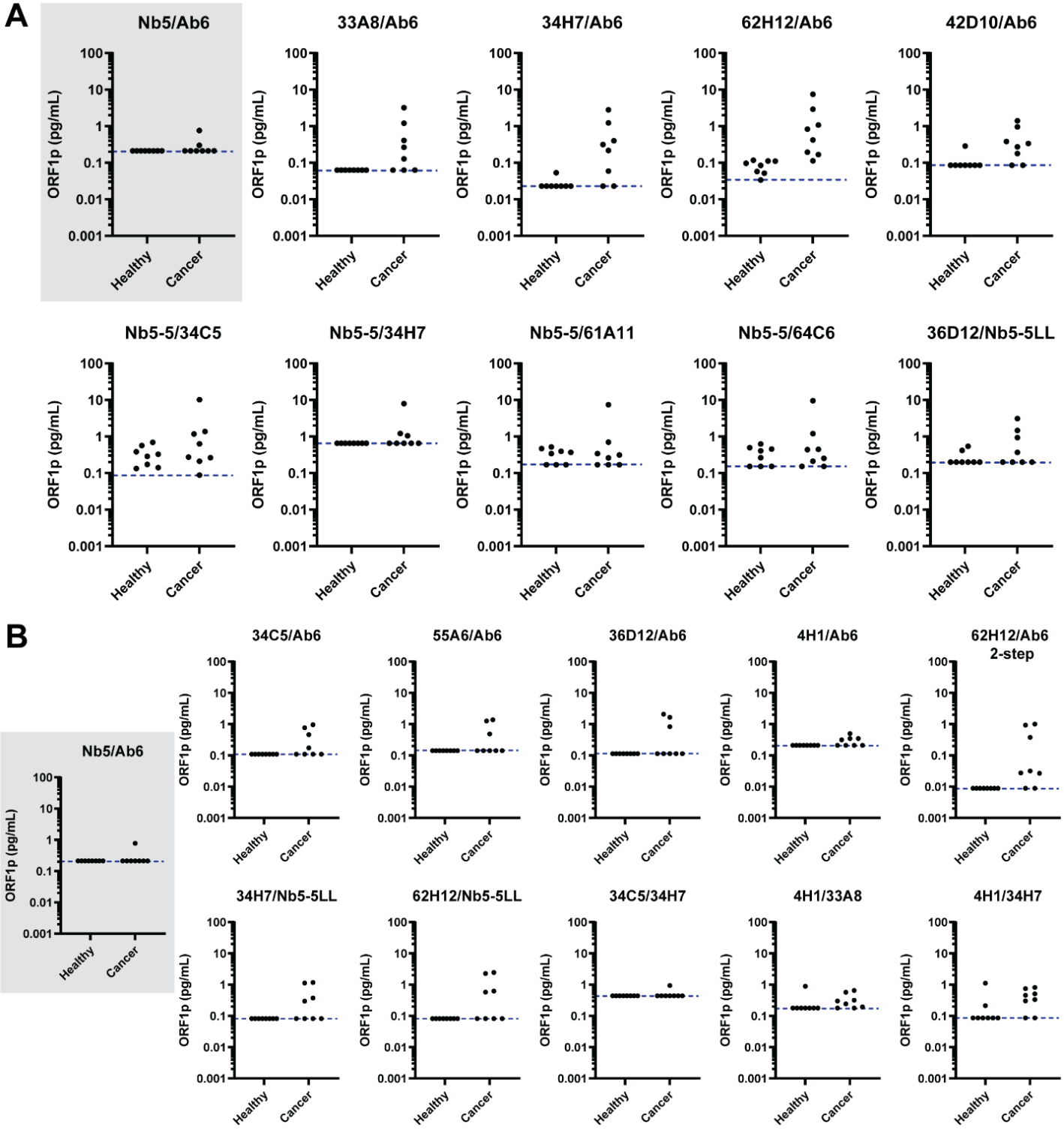
Supplementary Figure 11. Representative SPR sensorgrams of ORF1 monoclonal antibodies and nanobody multimers. Single-cycle kinetics performed over immobilized ORF1, with (A) monoclonal antibodies or (B) nanobody multimers injected sequentially at 0.1, 0.3, 1, 3, and 10nM concentrations. Raw data (blue) and a 1:1 binding model fit (black) are plotted.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



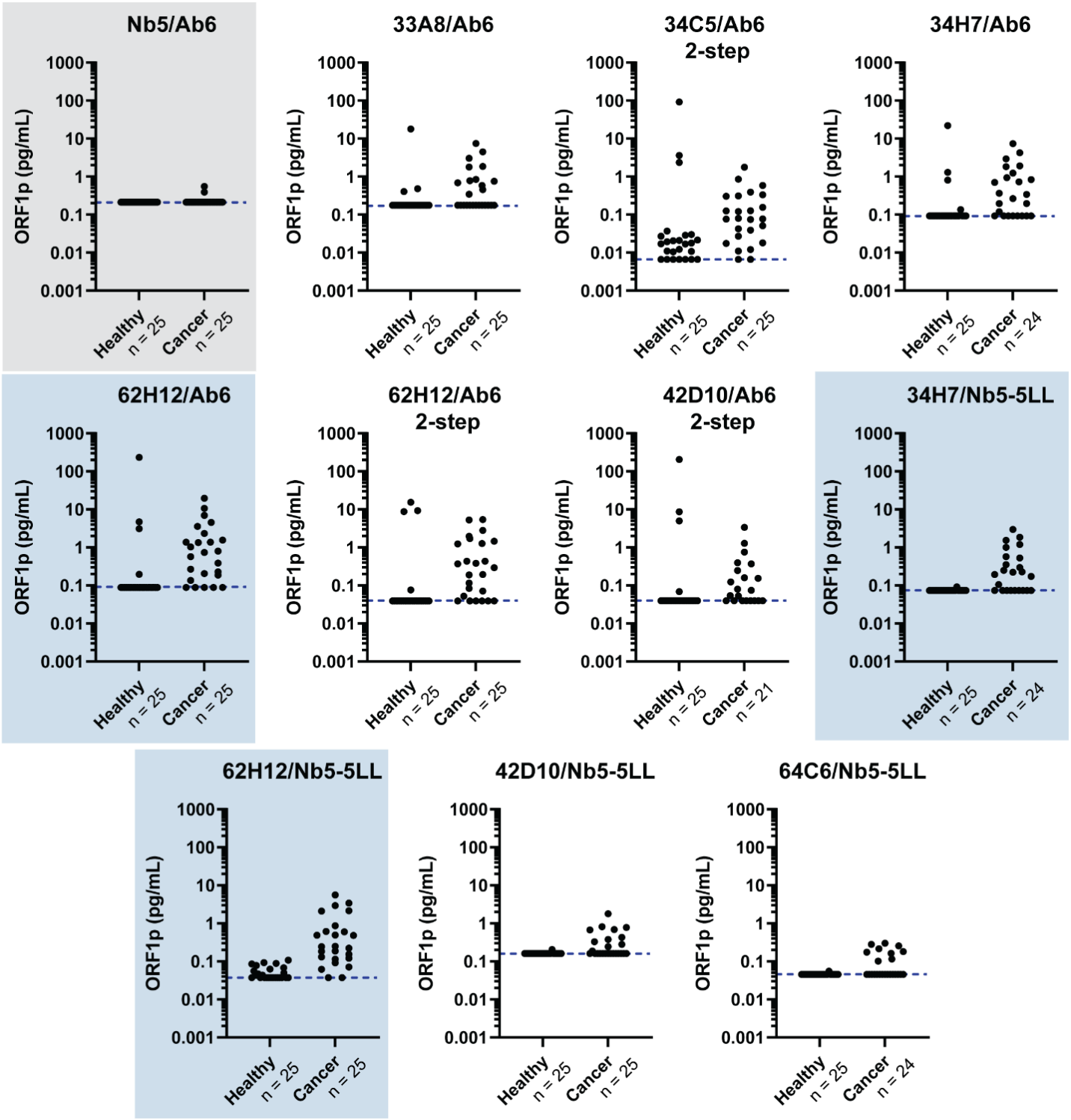
Supplementary Figure 12. Screening of multimeric nanobodies for ORF1p detection with Simoa. (A) Signal-to-background (Signal:Bg) comparisons of multimeric nanobodies as detector reagents on the Simoa platform, using recombinant ORF1p protein. A four-plex assay format, with a unique fluorescent dye-encoded bead type for each of the four capture reagents tested, was used for screening. (B) Signal-to-background comparison of multimeric nanobodies as capture reagents on Simoa, using recombinant ORF1p protein, paired with the monoclonal antibody Ab5 as detector reagent. Sub-labels “A” and “B” refer to different nanobody concentrations used during conjugation to beads. (C-D) Screening of select affinity reagent pairs in small sets of plasma samples from healthy and cancer (colorectal and gastroesophageal) patients. Each assay is denoted by the capture/detector reagent pair. The first-generation assay (Nb5/Ab6) measurements are depicted for comparison. All assays were performed as three-step Simoa assays unless otherwise indicated (two-step). Blue dashed lines indicate the assay limits of detection, accounting for four-fold dilution.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



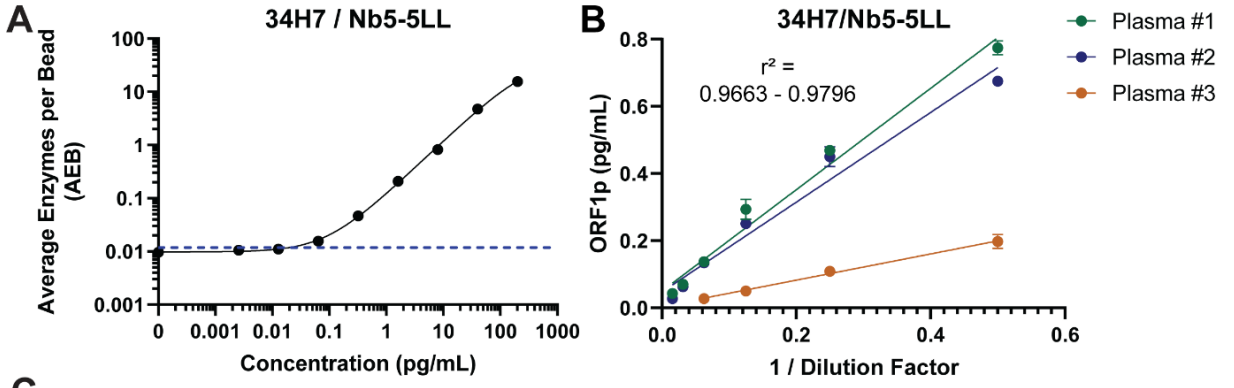
Supplementary Figure 13. First round of screening of newly developed monoclonal antibody and dimeric nanobody reagent pairs in patient plasma. Affinity reagent pairs selected from initial screening with recombinant ORF1 protein (Supplementary Figure 8) were screened in small sets of plasma samples from healthy and cancer (colorectal and gastroesophageal) patients. Eight healthy and eight cancer patients were used in each set of assays in (A) and (B). Blue dashed lines indicate assay limits of detection, accounting for four-fold dilution. Each assay is denoted by the capture/detector reagent pair. The first-generation assay (Nb5/Ab6) measurements are depicted for comparison (gray box).

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



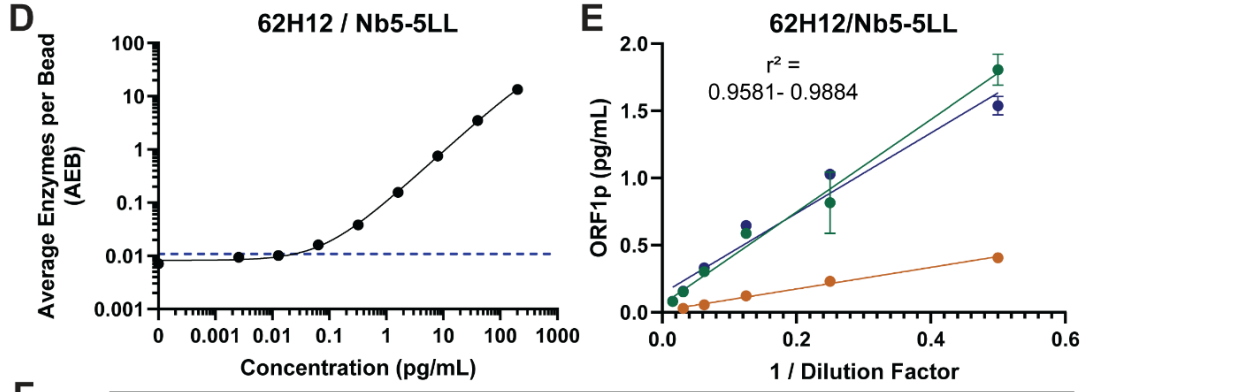
Supplementary Figure 14. Second round of screening of newly developed monoclonal antibody and dimeric nanobody reagent pairs in patient plasma. Affinity reagent pairs selected from a first round of screening in plasma samples (Supplementary Figure 9) were screened in 25 healthy and 25 cancer (colorectal, gastroesophageal, and breast) patient plasma samples. Each assay is denoted by the capture/detector reagent pair. The first-generation assay (Nb5/Ab6) measurements are depicted for comparison (gray box). Blue dashed lines indicate assay limits of detection, accounting for four-fold dilution. Blue boxes indicate the final three selected second-generation assays.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



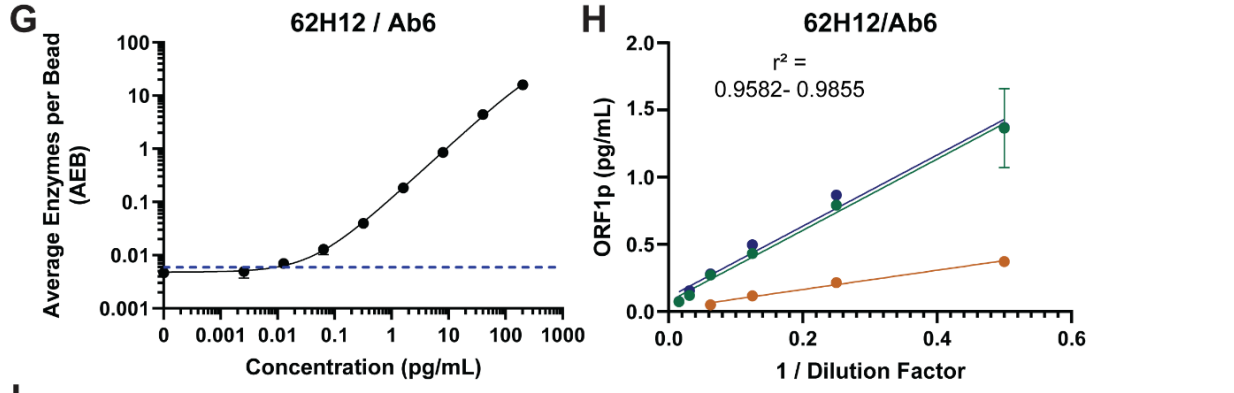
C

RECOVERY (%)				
Spiked Protein (pg/mL)	Plasma #1	Plasma #2	Plasma #3	Plasma #4
0.2	73.8 ± 4.7	78.8 ± 9.6	67.4 ± 7.9	72.1 ± 10.0
2	92.8 ± 0.8	82.3 ± 8.2	70.7 ± 0.3	78.1 ± 5.7
20	83.9 ± 4.3	88.4 ± 6.0	67.8 ± 6.4	77.2 ± 5.8



F

RECOVERY (%)				
Spiked Protein (pg/mL)	Plasma #1	Plasma #2	Plasma #3	Plasma #4
0.2	79.1 ± 2.0	86.9 ± 9.1	64.7 ± 4.9	90.5 ± 4.6
2	92.7 ± 2.6	88.6 ± 3.2	63.9 ± 1.0	79.6 ± 1.3
20	103.8 ± 6.4	85.4 ± 9.9	69.1 ± 5.7	84.1 ± 3.2



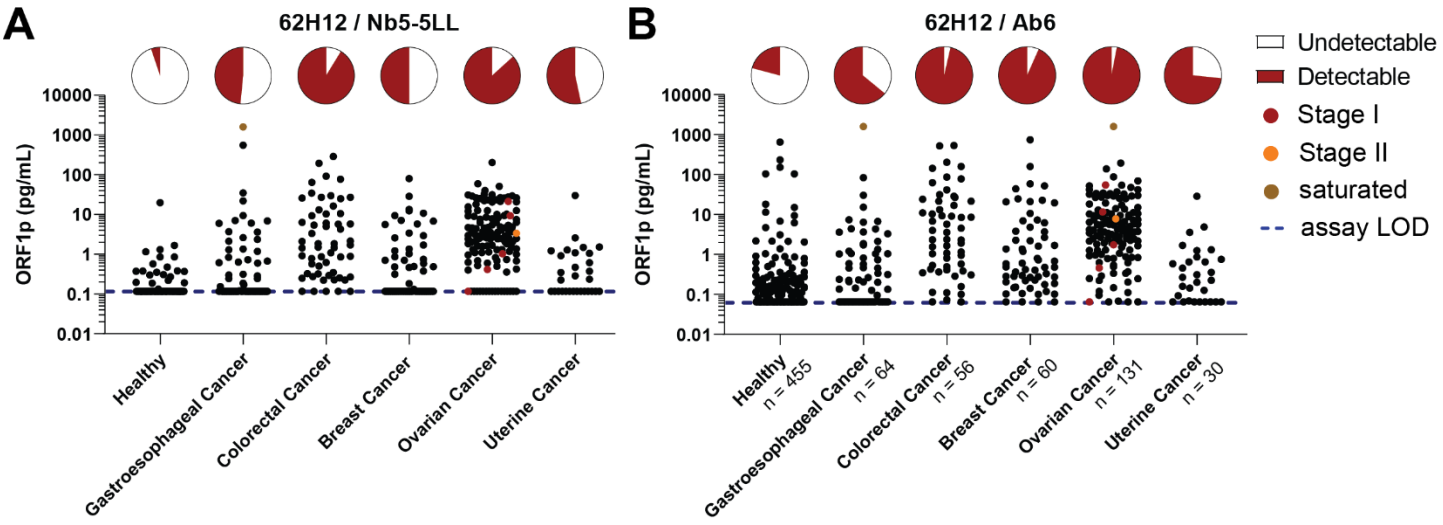
I

RECOVERY (%)				
Spiked Protein (pg/mL)	Plasma #1	Plasma #2	Plasma #3	Plasma #4
0.2	95.8 ± 6.3	100.2 ± 0.2	84.4 ± 16.7	103.4 ± 7.9
2	88.3 ± 6.9	97.0 ± 0.9	64.7 ± 7.1	84.6 ± 0.8
20	85.4 ± 4.2	82.2 ± 2.2	47.2	75.3 ± 1.3

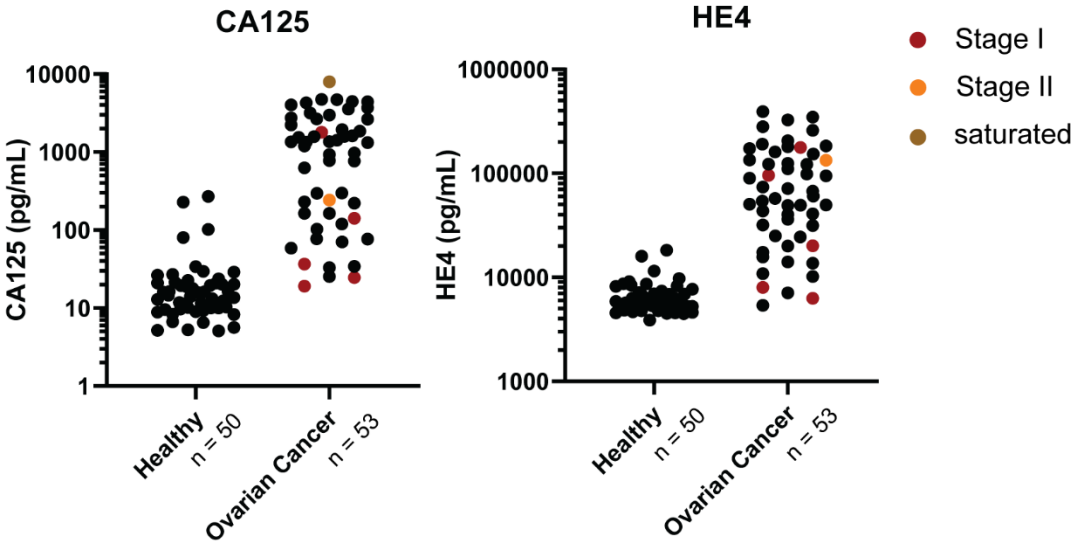
Supplementary Figure 15. Assay validation for second-generation ORF1p Simoa assays. (A-C) Representative calibration curve (A), dilution linearity (B), and spike and recovery results (C) for the 34H7/Nb5-5LL

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(capture/detector) assay. (D-F) Representative calibration curve (D), dilution linearity (E), and spike and recovery results (F) for the 62H12/Nb5-5LL (capture/detector) assay. (G-I) Representative calibration curve (G), dilution linearity (H), and spike and recovery results (I) for the 62H12/Ab6 (capture/detector) assay. Error bars represent the standard deviation of triplicate measurements in each calibration curve and duplicate measurements of all plasma samples.

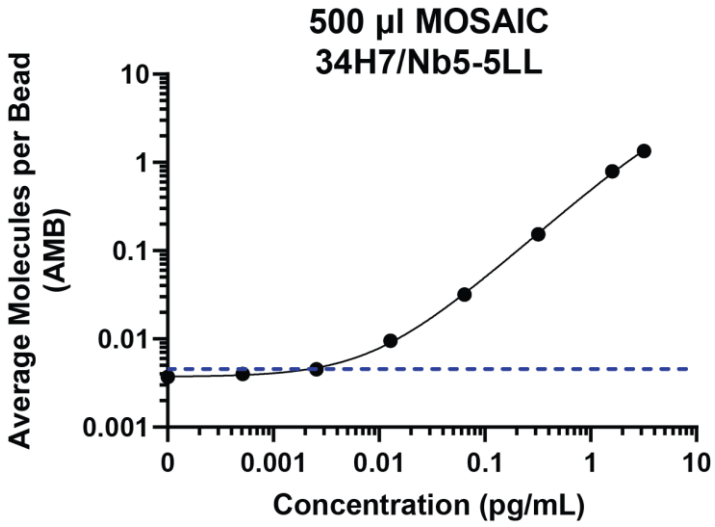


Supplementary Figure 16. Additional second-generation ORF1p Simoa assay measurements in healthy and cancer patients. Each assay is denoted by the capture/detector reagent pair. Blue dashed lines indicate assay limits of detection, accounting for four-fold dilution. The proportion of detectable patients within each cancer type is depicted by corresponding pie charts and early-stage patients (stage I/II) are indicated as red and orange dots, respectively.

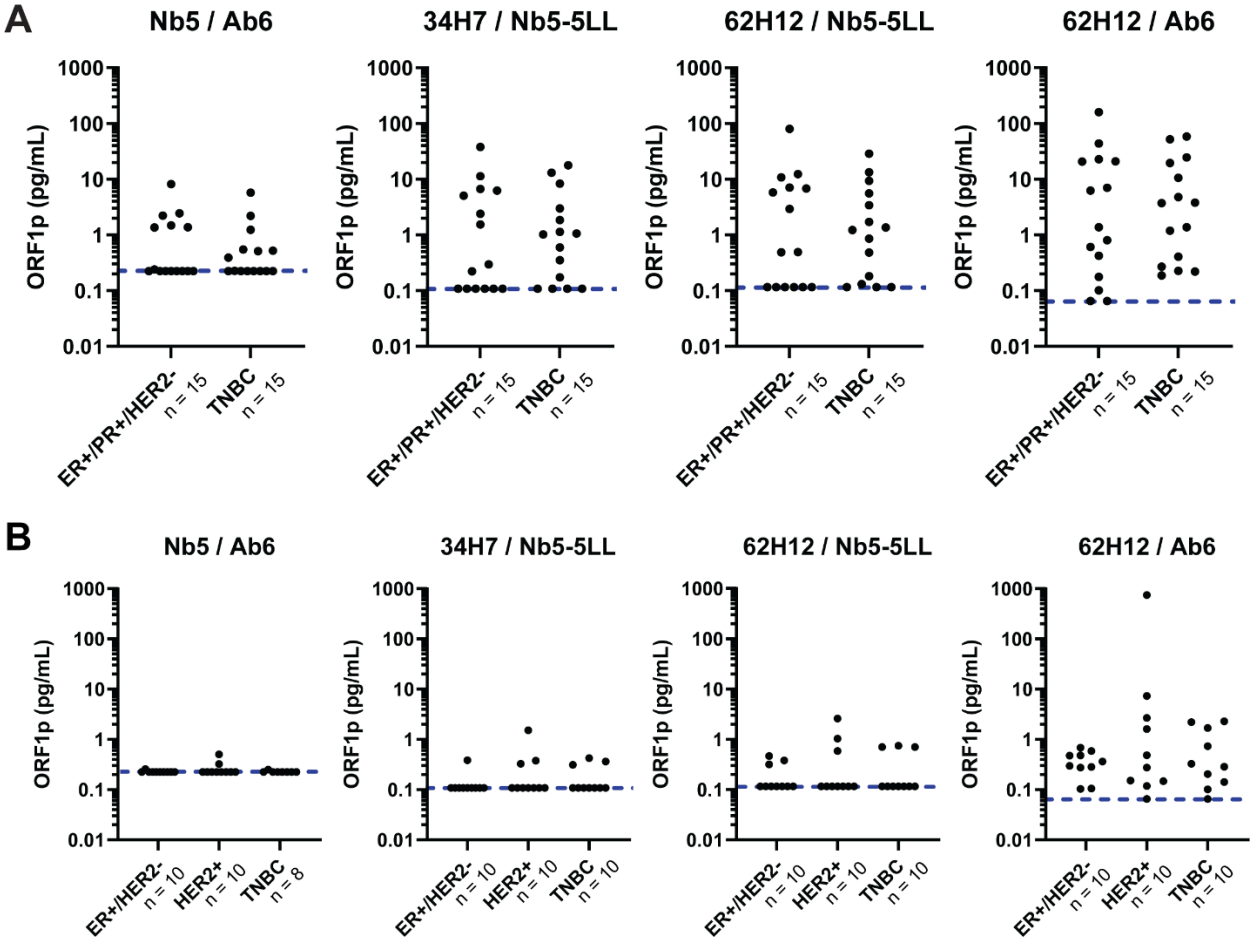


Supplementary Figure 17. Circulating CA125 and HE4 levels, as measured by Simoa, in ovarian cancer patient cohort (obtained from Ronny Drapkin at the University of Pennsylvania). Early-stage patients (stage I/II) are indicated as red and orange dots, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

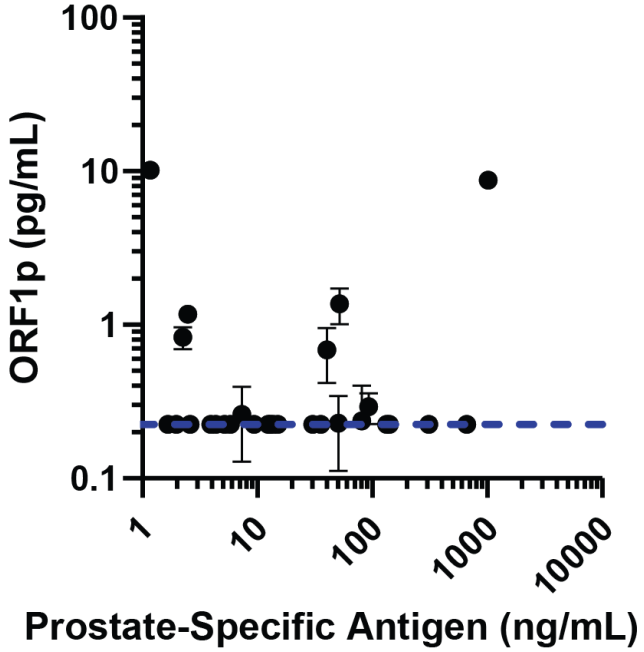


Supplementary Figure 18. Calibration curve for the large-volume MOSAIC assay used in Figure 5. Blue dashed line indicates the assay limit of detection.



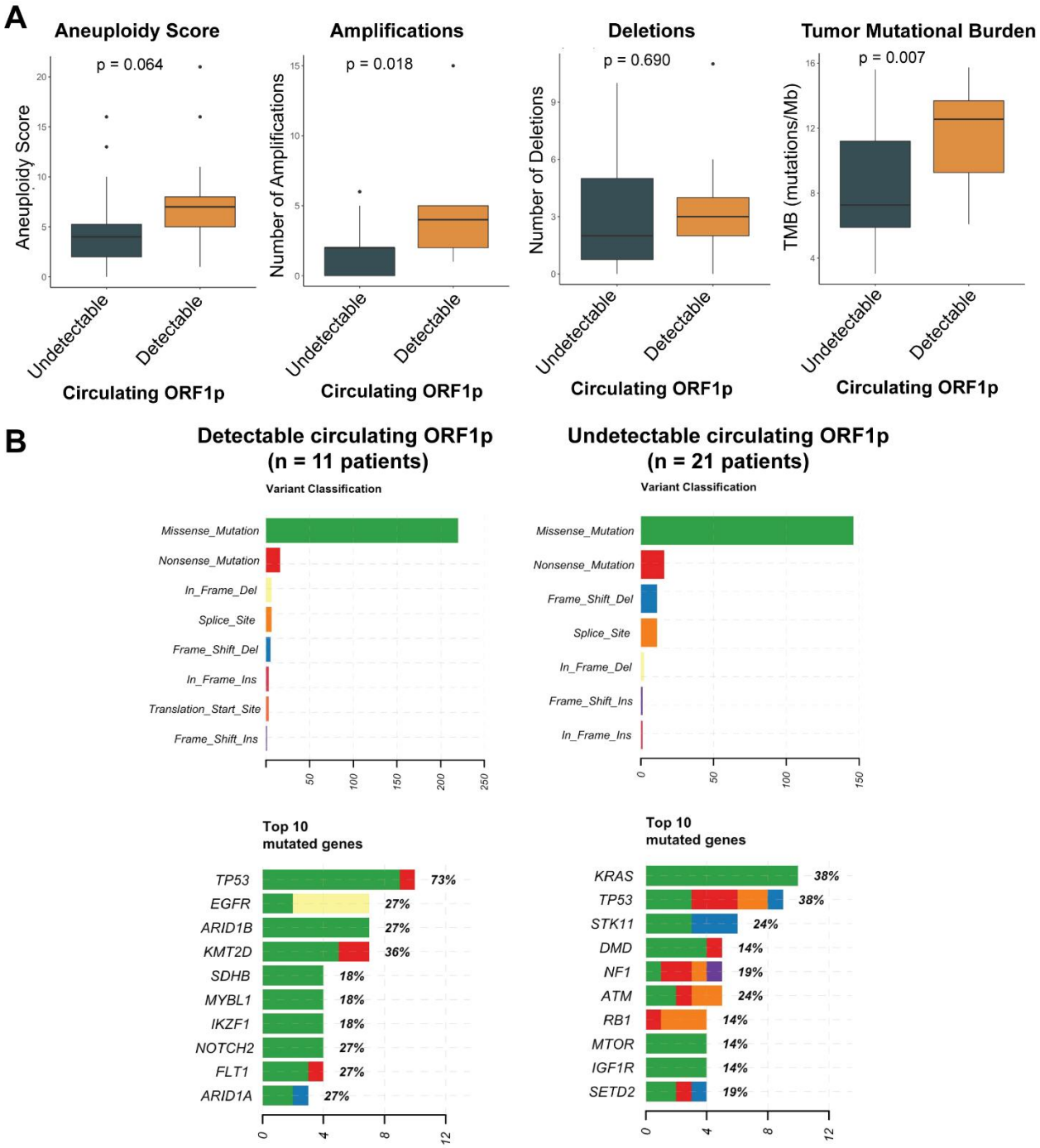
Supplementary Figure 19. Circulating ORF1p levels in breast cancer patients, classified by receptor status, with (A) metastatic and (B) localized disease. First- and second-generation Simoa assays were used to measure circulating ORF1p, with each assay labeled as capture/detector reagent. Blue dashed lines indicate assay limits of detection, accounting for four-fold dilution. ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; TNBC, triple-negative breast cancer.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



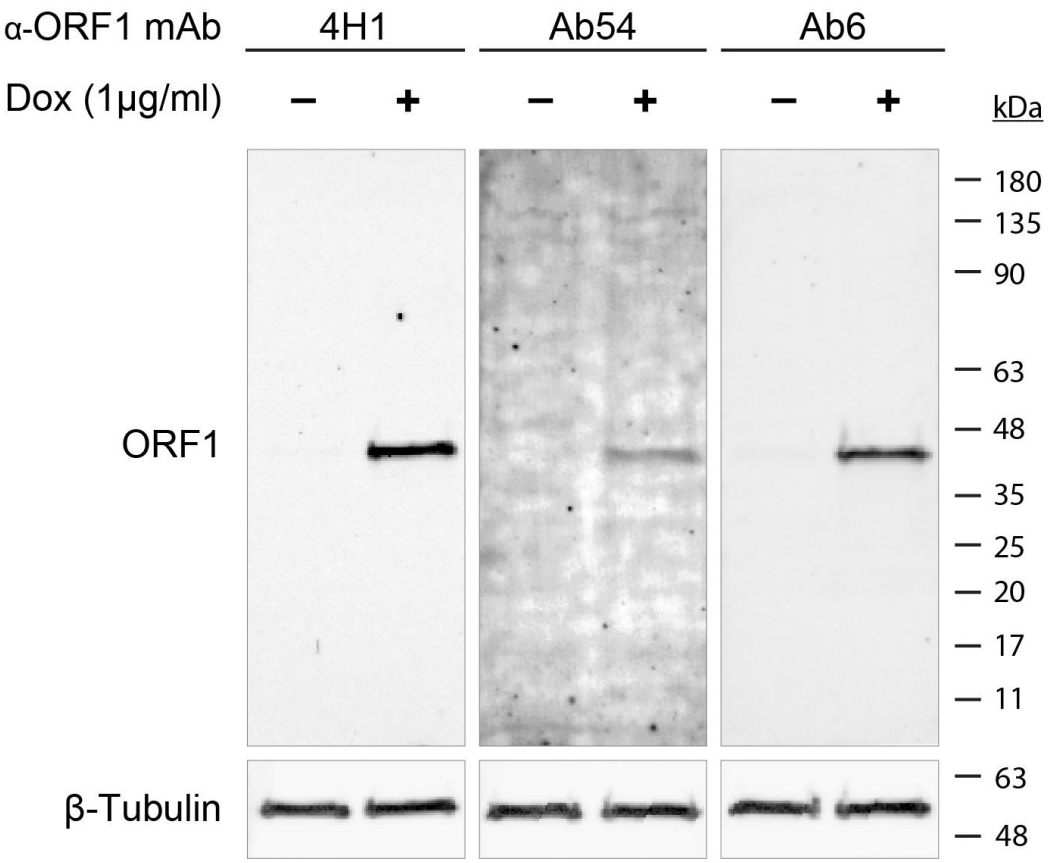
Supplementary Figure 20. Circulating ORF1p (first generation assay) versus prostate-specific antigen levels in prostate cancer patient plasma. Blue dashed line indicates the assay limit of detection, accounting for four-fold dilution.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



Supplementary Figure 21. Genomics analysis of lung cancer patient cohort. (A) Comparisons of aeuploidy score, amplifications, deletions, and tumor mutational burden between lung cancer patients with detectable versus undetectable circulating ORF1p levels, as measured by the first-generation ORF1p Simoa assay. (B) Mutation types and top mutated genes in lung cancer patients with detectable versus undetectable circulating ORF1p.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



Supplementary Figure 22. Validation of monoclonal antibodies by western blotting. HeLa cells with stably integrated tet-On LINE-1 (L1RP)[6] were treated with DMSO as control or with 1 µg/ml doxycycline (Dox) for 4 days to induce L1 expression, lysed with RIPA buffer, cleared, and 20 µg total protein (BCA) was loaded per lane and blotted with the indicated antibodies.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table S1. Assay conditions and limits of detection for first- and second-generation ORF1p assays.

Platform	Assay (Capture / Detector)	Sample Volume (µL)	Incubation times (capture-detector-streptavidin minutes)	Detector (pg/mL)	Streptavidin β-galactosidase (pM)	Limit of Detection (pg/mL)
Simoa	Nb5 / Ab6	100	15-5-5	0.3	150	0.056 ± 0.032
Simoa	34H7 / Nb5-5LL	100	15-5-5	0.3	300	0.027 ± 0.016
Simoa	62H12 / Nb5-5LL	100	15-5-5	0.3	300	0.029 ± 0.016
Simoa	62H12 / Ab6	100	15-5-5	0.3	300	0.016 ± 0.007
MOSAIC	34H7 / Nb5-5LL	2000	120-10-10	0.3	100 (streptavidin-DNA)	0.002

Table S2. Affinity reagents used across all screening experiments.

Affinity reagent	Type	Source
Ab6	Rabbit monoclonal antibody	Abcam (ab246317)
Ab54	Rabbit monoclonal antibody	Abcam (ab246320)
Nb5	Nanobody	Rockefeller University
4H1	Mouse monoclonal antibody	MilliporeSigma (MABC1152)
JH73	Rabbit monoclonal antibody	
JH74	Rabbit monoclonal antibody	
D3W9O	Rabbit monoclonal antibody	Cell Signaling Technology (88701)
Nb5-5 pMT993	Homodimeric nanobody (short linker)	This study
Nb5-5LL pMT997	Homodimeric nanobody (long linker)	This study
Nb1-1 pMT991	Homodimeric nanobody (short linker)	This study
Nb2-2 pMT992	Homodimeric nanobody (short linker)	This study
Nb5-1 pMT994	Heterodimeric nanobody (short linker)	This study
Nb5-2 pMT995	Heterodimeric nanobody (short linker)	This study
33A8	Rabbit monoclonal antibody	This study
34C5	Rabbit monoclonal antibody	This study
34H7	Rabbit monoclonal antibody	This study
36D12	Rabbit monoclonal antibody	This study
42D10	Rabbit monoclonal antibody	This study
50E9	Rabbit monoclonal antibody	This study
55A6	Rabbit monoclonal antibody	This study
64C6	Rabbit monoclonal antibody	This study
61A11	Rabbit monoclonal antibody	This study
62H12	Rabbit monoclonal antibody	This study

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table S3. Surface plasmon resonance measurements of binding kinetics of anti-ORF1p nanobodies

Construct	ka (1/Ms)	kd (1/s)	KD (M)	KD se	n
NB Orf1-1	6.1E+05	6.4E-04	1.0E-09		1
NB Orf1-2	6.4E+05	4.2E-05	7.1E-11	2.0E-11	3
NB Orf1-5	4.8E+06	2.3E-03	3.7E-10	1.9E-10	3
NB Orf1-8	9.4E+05	2.0E-04	2.0E-10	4.1E-11	3
NB Orf1-9	6.1E+05	1.3E-04	2.2E-10	2.0E-11	3
NB Orf1-10	8.0E+05	1.8E-04	2.7E-10	4.5E-11	3
NB Orf1-11	4.5E+06	3.3E-03	6.4E-10	1.6E-10	3
NB Orf1-12	1.2E+06	4.2E-04	2.9E-10	7.6E-11	3
NB Orf1-13	8.2E+05	2.5E-04	3.0E-10	2.3E-11	3
NB Orf1-15	1.1E+06	5.1E-04	4.9E-10	3.8E-11	3
NB Orf1-16	2.6E+06	3.2E-03	8.4E-10	3.4E-10	3
NB Orf1-17	1.5E+06	4.7E-04	2.9E-10	5.7E-11	3
NB Orf1-20	1.7E+06	3.8E-04	2.4E-10	4.2E-11	3
NB Orf1-21	9.6E+05	1.8E-04	1.7E-10	3.3E-11	3

Table S4. Surface plasmon resonance measurements of binding kinetics of affinity reagents used in single molecule assays. Orange-highlighted cells indicate dissociation rate constants (k_d) below the limit of detection of the instrument, resulting in upper-bound equilibrium dissociation constant (K_D) estimation. SEM, Standard Error of the Mean.

Construct	#	k_a (1/M·s)	k_d (1/s)	K_D (M)	K_D SEM	n
Dimeric and Trimeric Nanobodies						
Nb 1-1	MT991	Weak signal				
Nb 2-2	MT992	6.4E+05	<5.0E-06	<7.9E-12	7.4E-13	3
Nb 5-5	MT993	9.0E+05	3.4E-05	3.8E-11	1.5E-11	3
Nb 5-1	MT994	5.2E+05	3.1E-05	6.0E-11	3.8E-11	2
Nb 5-2	MT995	4.0E+05	1.3E-05	3.4E-11	4.7E-12	3
Nb 5-5-5	MT996	8.1E+05	2.9E-05	3.6E-11	8.5E-12	3
Nb 5-5 long "LL"	MT997	7.5E+05	2.1E-05	2.8E-11	1.2E-11	3
Monoclonal Antibodies						
mAb	62H12	4.8E+05	<5.0E-06	<1.0E-11	4.4E-13	3
mAb	34H7	5.2E+05	<5.0E-06	<9.6E-12	4.1E-12	3
mAb	Ab6	2.2E+05	<5.0E-06	<2.2E-11	2.6E-12	3
mAb	4H1	4.5E+05	<5.0E-06	<1.1E-11	2.5E-12	3
mAb	Ab54	4.2E+05	<5.0E-06	<1.2E-11	5.5E-14	2

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table S5. Epitope binning of ORF1p nanobodies. Fraction of 2nd antibody able to bind after saturation with 1st antibody is indicated, normalized to maximum signal bound. Overlapping epitopes are defined by <0.3 remaining binding activity (red shading). Epitope groups are labeled I-III.

		2nd Ab					
	Group:	I	II	III			
		Nb-2	Nb-5	Nb-8	Nb-9	Nb-10	Nb-21
1st Ab	Nb-2	0.16	0.66	0.94	1.00	1.00	1.00
	Nb-5	0.82	-0.06	1.00	0.99	0.93	0.90
	Nb-8	1.00	0.53	0.07	0.13	0.26	0.24
	Nb-9	0.91	0.66	0.03	0.08	0.19	0.17
	Nb-10	0.70	1.00	-0.08	-0.07	0.08	0.05
	Nb-21	0.83	0.47	-0.06	-0.05	0.08	0.03

Table S6. Epitope binning of ORF1p MAbs and representative nanobodies. Fraction of 2nd antibody able to bind after saturation with 1st antibody is indicated, normalized to maximum signal bound. Overlapping epitopes are defined by <0.3 remaining binding activity (red shading). Epitope groups are labeled I-V.

		2nd Ab					
	Group:	I	II	III	IV	V	
		Nb-2	Nb-5	Nb-9	Ab6	34H7	62H12
1st Ab	Nb-2	0.18	0.91	0.82	0.97	1.00	0.98
	Nb-5	0.67	0.02	0.83	0.95	0.88	0.99
	Nb-9	0.64	0.90	0.04	0.78	0.42	0.86
	Ab6	1.00	0.94	0.87	0.07	0.71	1.00
	34H7	0.87	0.92	0.85	0.81	0.10	0.22
	62H12	0.89	1.00	1.00	1.00	0.59	0.12

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table S7. Performance metrics of classification models built from all healthy and all ovarian cancer patients measured by the second-generation Simoa assays. Classification models were built using five-fold cross validation, with univariate logistic regression for each individual assay and k-nearest neighbors algorithm for the multivariate model combining two assays.

Assay	Accuracy	Sensitivity	Specificity	F1 Score
34H7 / Nb5-5LL	95.0%	85.7%	97.5%	0.876
62H12 / Nb5-5LL	94.6%	86.5%	96.9%	0.870
62H12 / Ab6	92.0%	87.3%	93.3%	0.827
34H7/Nb5-5LL and 62H12/Ab6	94.4%	84.1%	97.3%	0.865

Table S8. Performance metrics of classification models built from ovarian cancer and age-matched healthy female patients (Ronny Drapkin, University of Pennsylvania). Classification models were built using k-nearest neighbors (KNN) and light gradient-boosting machine (LightGBM) algorithms, with five-fold cross validation.

Biomarkers	Accuracy	Sensitivity	Specificity	F1 Score
CA125 and HE4	89.0%	90.0%	88.0%	0.891
CA125, HE4, and ORF1p (34H7/Nb5-5LL)	95.0%	94.0%	96.0%	0.949
CA125, HE4, and ORF1p (62H12/Nb5-5LL)	93.0%	88.2%	98.0%	0.926
CA125, HE4, and ORF1p (62H12/Ab6)	91.1%	86.2%	96.0%	0.906

Table S9. Clinicopathological characteristics for gastroesophageal cancer cohort

Characteristic		**N=19**
Median age (range)-yr		76 (37-81%)
Male sex-no. (%)		19 (100%)
Histology-no. (%)	Adenocarcinoma	19 (100%)
Primary tumor location no. (%)	Esophagus	7 (37%)
	Gastroesophageal Junction	7 (37%)
	Stomach	5 (26%)
Disease stage at initial diagnosis-no. (%)	Locally advanced (II-III)	11 (58%)
	Advanced (IV)	8 (42%)
Treatment paradigm during pre- and post-assay collection*	Neoadjuvant	9 (47%)
	Adjuvant	2 (11%)
	Palliative	8 (42%)
	Surgery	3 (16%)
Treatment response	Responder	13 (68%)
	Non-responder	6 (32%)

*Patients who underwent surgery also received systemic therapy thus total exceeds N=19

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.25.525462; this version posted January 25, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table S10. Demographic and clinicopathological variables of head and neck squamous cell carcinoma patients grouped by detectability of circulating ORF1p.

	Total	ORF1p-negative	ORF1p-positive
Sex			
Male	33	10	1
Female	11	23	10
Cancer site			
Oral cavity	33	24	9
Oropharynx	3	2	1
Larynx	8	7	1
Race			
White	38	28	10
Asian	2	2	0
Black	1	0	1
Other/Unknown	3	3	0
pT Stage			
pT2	6	4	2
pT3	13	11	2
pT4a	20	15	5
Unknown	5	3	2
pN Stage			
pN0	19	15	4
pN1	8	6	2
pN2	6	6	0
pN3	6	3	3
Unknown	5	3	2
PNI			
Positive	18	14	4
Negative	22	16	6
Unknown	4	3	1
ENE			
Positive	9	6	3
Negative	12	9	3
NA	19	15	4
Unknown	4	3	1

Supplementary References

1. Carter, V.*, et al.* High Prevalence and Disease Correlation of Autoantibodies Against p40 Encoded by Long Interspersed Nuclear Elements in Systemic Lupus Erythematosus. *Arthritis Rheumatol* **72**, 89-99 (2020).
2. Fridy, P.C.*, et al.* A robust pipeline for rapid production of versatile nanobody repertoires. *Nat Methods* **11**, 1253-1260 (2014).
3. Mast, F.D.*, et al.* Highly synergistic combinations of nanobodies that target SARS-CoV-2 and are resistant to escape. *Elife* **10**(2021).
4. Rodić, N.*, et al.* Long Interspersed Element-1 Protein Expression Is a Hallmark of Many Human Cancers. *The American Journal of Pathology* **184**, 1280-1286 (2014).
5. Rajurkar, M.*, et al.* Reverse Transcriptase Inhibition Disrupts Repeat Element Life Cycle in Colorectal Cancer. *Cancer Discov* (2022).
6. Ardeljan, D.*, et al.* Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. *Nat Struct Mol Biol* **27**, 168-178 (2020).