

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

Martin S. Taylor^{*,1,‡}, Connie Wu^{*,2,3,a,‡}, Peter C. Fridy⁴, Stephanie J. Zhang^{2,3}, Yasmeen Senussi^{2,3}, Justina C. Wolters⁵, Wen-Chih Cheng⁶, John Heaps⁶, Bryant D. Miller⁶, Kei Mori^{2,7}, Limor Cohen^{2,3,8}, Hua Jiang⁴, Kelly R. Molloy⁹, Brian T. Chait⁹, Michael Goggins¹⁰, Irun Bhan¹¹, Joseph W. Franses¹¹, Xiaoyu Yang¹², Mary-Ellen Taplin¹², Xinan Wang¹³, David C. Christiani^{11,13}, Bruce E. Johnson¹², Matthew Meyerson¹², Ravindra Uppaluri¹⁴, Ann Marie Egloff¹⁴, Elyssa N. Denault¹¹, Laura M. Spring¹¹, Tian-Li Wang¹⁰, Ie-Ming Shih¹⁰, Euihye Jung¹⁵, Kshitij S. Arora¹, Lawrence R. Zukerberg², Osman H. Yilmaz¹⁶, Gary Chi¹¹, Bryanna L. Norden¹¹, Yuhui Song¹¹, Linda Nieman¹¹, Aparna R. Parikh¹¹, Matthew Strickland¹¹, Ryan B. Corcoran¹¹, Tomas Mustelin¹⁷, George Eng^{1,18}, Ömer H. Yilmaz^{1,18}, Ursula A. Matulonis¹², Steven J. Skates¹⁹, Bo R. Rueda²⁰, Ronny Drapkin¹⁵, Samuel J. Klempner¹¹, Vikram Deshpande¹, David T. Ting¹¹, Michael P. Rout⁴, John LaCava^{4,21}, David R. Walt^{2,3,‡}, and Kathleen H. Burns^{2,6,‡}

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

20. Department of Obstetrics and Gynecology, Massachusetts General Hospital, and Harvard Medical School, Boston, MA, USA.

21. European Research Institute for the Biology of Ageing, University Medical Center Groningen, The Netherlands.

* Equal Contribution

‡ Correspondence : mstaylor@mgh.harvard.edu, conniewu@umich.edu, dwalt@bwh.harvard.edu, kathleenh_burns@dfci.harvard.edu

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

1 **Abstract**

2 Improved biomarkers are needed for early cancer detection, risk stratification, treatment selection,
3 and monitoring treatment response. While proteins can be useful blood-based biomarkers, many
4 have limited sensitivity or specificity for these applications. Long INterspersed Element-1 (LINE-
5 1, L1) open reading frame 1 protein (ORF1p) is a transposable element protein overexpressed in
6 carcinomas and high-risk precursors during carcinogenesis with negligible detectable expression
7 in corresponding normal tissues, suggesting ORF1p could be a highly specific cancer biomarker.
8 To explore the potential of ORF1p as a blood-based biomarker, we engineered ultrasensitive
9 digital immunoassays that detect mid-attomolar (10^{-17} M) ORF1p concentrations in patient plasma
10 samples across multiple cancers with high specificity. Plasma ORF1p shows promise for early
11 detection of ovarian cancer, improves diagnostic performance in a multi-analyte panel, and
12 provides early therapeutic response monitoring in gastric and esophageal cancers. Together,
13 these observations nominate ORF1p as a multi-cancer biomarker with potential utility for disease
14 detection and monitoring.

15

16 **Statement of Significance (50 word)**

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

19 blood. Ultrasensitive ORF1p assays from as little as 25 μ L plasma are novel, rapid, cost-effective
20 tools in cancer detection and monitoring.

21

22 **Introduction**

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

39

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

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

53

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

63

64 **Results**

65 Because our preliminary survey of plasma ORF1p levels by Simoa in patients with advanced
66 stage colorectal cancer (CRC) indicated detectable ORF1p levels in 90% of cases(18), higher
67 than the proportion of CRCs we previously reported to express ORF1p by immunohistochemistry
68 (50%, n=18)(8), we first sought to benchmark ORF1p in tissues. Using a re-optimized protocol,
69 we stained 211 CRCs [178 sequential cases included on a tissue microarray (TMA) as well as an
70 additional 33 with matched plasma] and found 91% of CRC cases were immunoreactive for

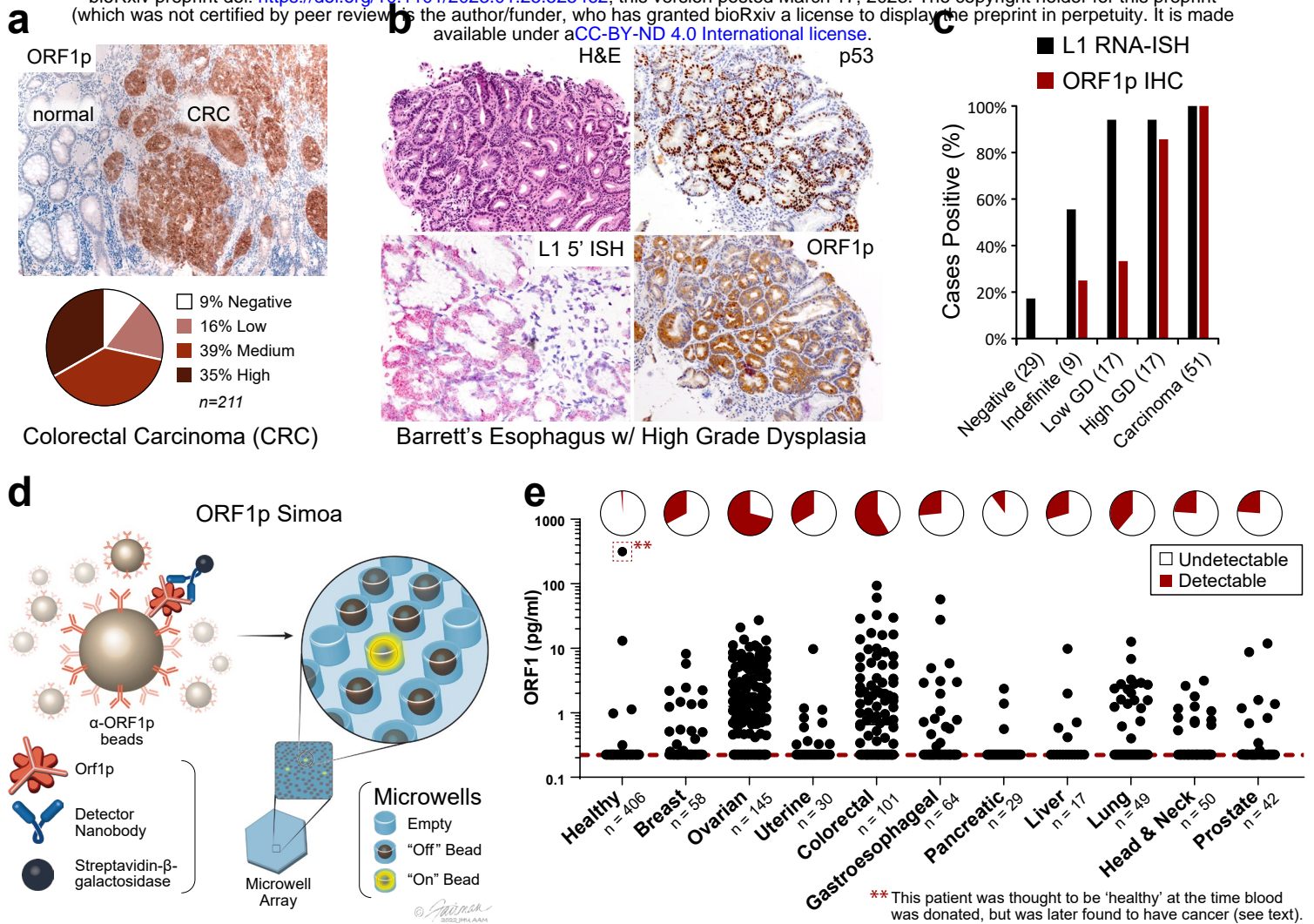


Figure 1. ORF1p expression is early and pervasive in carcinomas. **a**, ORF1p immunostaining in a cohort of 211 colorectal cancers. **b**, Representative BE case: lesional cells overexpress p53, the L1 RNA, and ORF1p. **c**, L1 RNA and ORF1p overexpression across a cohort of 72 consensus BE cases and 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.

71 ORF1p (**Fig. 1a**). This result is consistent with genetic studies demonstrating somatic L1
72 retrotransposition in most CRCs(19), including activity in precancerous lesions antedating APC
73 tumor suppressor loss(20-22). Similarly, genetic evidence shows esophageal adenocarcinoma
74 (EAC) has high L1 activity(12), and L1 insertions occur in the highly prevalent Barrett's esophagus
75 (BE) precursor early in carcinogenesis(23,24). We therefore assembled a cross-sectional cohort
76 of 72 BE cases with consensus diagnosis reached by three expert gastrointestinal pathologists.
77 L1 RNA and ORF1p expression were pervasive in dysplastic BE and present in 100% of 51
78 esophageal carcinomas (**Fig. 1b,c**); all five BE cases indefinite for dysplasia and positive for
79 ORF1p and/or L1 RNA developed high grade dysplasia on subsequent biopsies (not shown).
80 Overall, this picture is similar to high grade serous ovarian cancers (HGSOC), where ORF1p is
81 expressed in 90% of cases and 90% of fallopian tube precursor lesions (serous tubal
82 intraepithelial carcinomas, STICs)(8,15,25). Taken together, ORF1p tissue expression is highly
83 prevalent in gastrointestinal and gynecologic carcinomas and high-risk precursor lesions.

84

85 We next sought to extend our tissue findings and explore plasma ORF1p. We optimized our
86 previously reported ORF1p Simoa assay and assessed the landscape of ORF1p levels in
87 pretreatment plasma from patients with advanced cancers. This “first-generation” assay uses a
88 recombinant, single-domain camelid nanobody (Nb5) as the capture reagent and a monoclonal
89 antibody (Ab6) as the detector reagent and has a limit of detection of 0.056 pg/mL (~470 aM
90 trimeric ORF1p), corresponding to 1.9 fM in plasma after correcting for sample dilution (**Fig. 1d**,
91 **Table S1**). With this assay, we surveyed multiple cancer types and >400 ‘healthy’ control
92 individuals, who were without known cancer at the time blood was donated to the biobank. Plasma
93 ORF1p appears to be a highly specific cancer biomarker, with undetectable levels in ~99% of
94 controls (ages 20-90, **Fig. 1e, S1**). Of the five control patients with detectable ORF1p, the one
95 with the highest ORF1p was later found to have advanced prostate cancer and a cutaneous T
96 cell lymphoma; limited clinical information is available for the other four positive ‘healthy’

97 individuals. With a cutoff set at 98% specificity in healthy controls, the highest proportions of
98 ORF1p(+) cases were observed in colorectal (58%, n=101) and ovarian cancers (71%, n=145).
99 While most of these patients had advanced-stage disease, plasma ORF1p remained detectable
100 in several early-stage patients in the cohort, including in those with ovarian and lung cancers and
101 in 5/18 with intraductal papillary mucinous neoplasms in the pancreas (IPMN, **Fig. S2-S4**).
102 Notably, four of eight stage I ovarian cancers in the cohort were positive (**Fig. S2**), suggesting
103 that plasma ORF1p may be an indicator of early-stage disease. As L1 expression is also
104 dysregulated in autoimmune disease and autoantibodies against ORF1p are prevalent in patients
105 with systemic lupus erythematosus (SLE), we measured plasma ORF1p in 30 SLE patients and
106 observed no detectable levels (**Fig. S5**)(26). Detectable ORF1p was seen in 1 of 30 patients with
107 chronic liver disease; the one positive patient was subsequently diagnosed with hepatocellular
108 carcinoma (**Fig. S5**). Size exclusion chromatography analysis of patient plasma further showed
109 that the majority of ORF1p resides outside extracellular vesicles (**Fig. S6**). Together, these
110 findings support the hypothesis that tumor-derived ORF1p can be found in the peripheral blood
111 of cancer patients and may act as a cancer-specific biomarker.

112
113 Given the gap between proportions of ORF1p(+) cancers by tumor immunohistochemistry (~90%
114 for CRC and HGSOC) versus by blood testing (~60-70%), we evaluated the possibility of
115 increasing plasma assay sensitivity by decreasing the assay's lower limit of detection. To this end,
116 we developed a panel of ORF1p affinity reagents, including new recombinant rabbit monoclonal
117 antibodies (RabMAbs) and engineered camelid nanobodies raised against recombinant human
118 ORF1p. Because ORF1p is homotrimeric, we engineered multimeric nanobody reagents with the
119 goal of enhancing binding affinity via increased avidity. These parallel development efforts
120 ultimately yielded both improved nanobody and rabbit monoclonal antibody reagents with at least
121 low-picomolar equilibrium dissociation constants (K_D) (**Fig. S7-S12, Table S2-S4**). Iterative
122 screening of these reagents with Simoa using recombinant antigen and select patient plasma

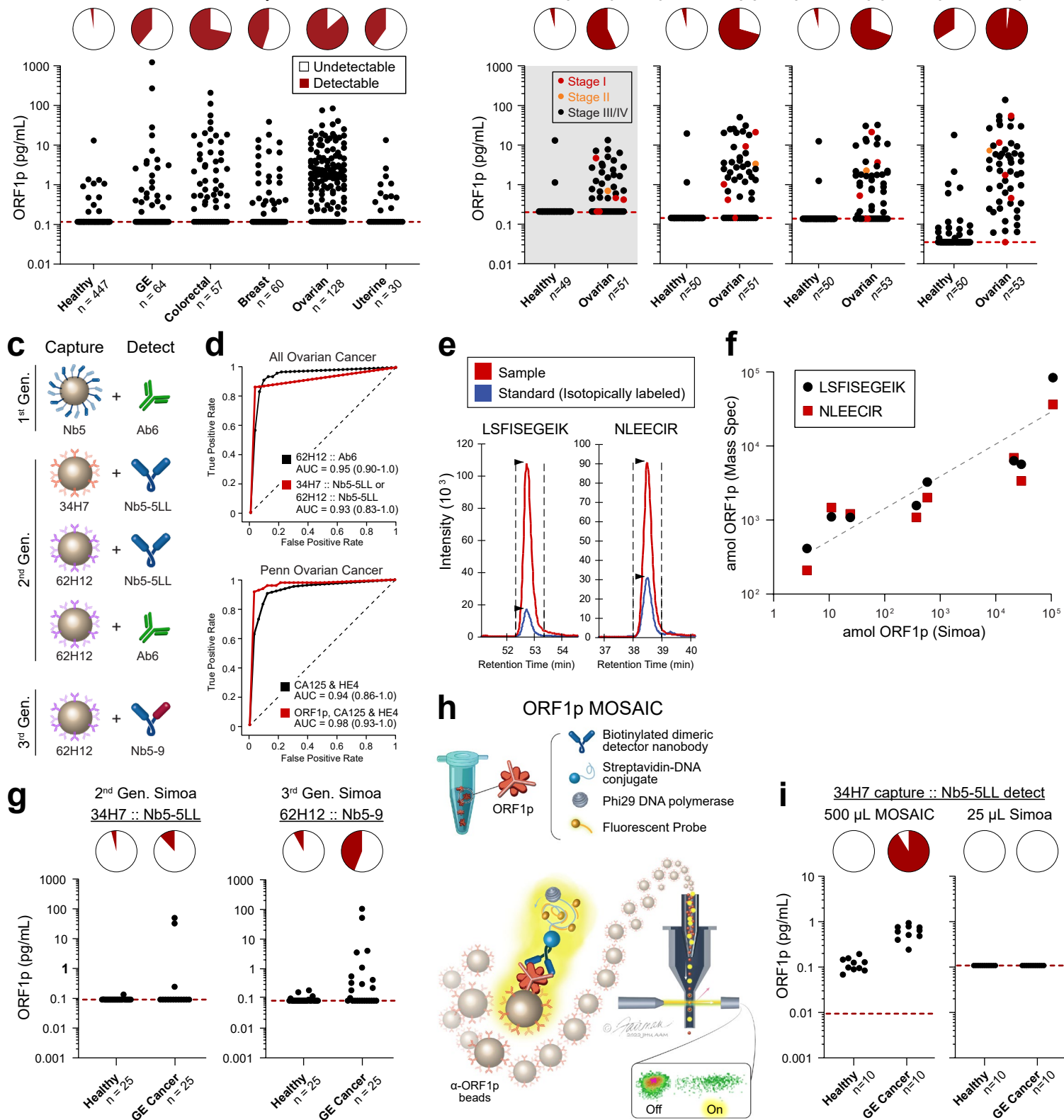


Figure 2. Improved detection of ORF1p with second- and third-generation assays. **a**, 34H7::Nb5-5LL second-generation assay measurements across a multi-cancer cohort. **b**, Ovarian cancer patients with age- and gender-matched controls in first- and second-generation assays; patients are a subset of 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 and Nb5-9 are an engineered homodimeric and heterodimeric nanobodies, respectively. **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**, Targeted proteomics measurements of plasma ORF1p from a gastric cancer patient using two quantotypic peptides (LSFISEGEIK and NLEECIR) with internal standards. **f**, correlation between measured ORF1p by Simoa and targeted proteomics assays; $r=0.97$ (Simoa vs LSFISEGEII) and $r=0.99$ (Simoa vs NLEECIR, t test), $p<0.0001$ for both. **g**, Comparison of 2nd and 3rd generation Simoa assays (25 μ L) in 25 mostly undetectable gastroesophageal (GE) cancer and healthy control patients. **h**, 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. **i**, 34H7::Nb5-5LL MOSAIC and Simoa assays in 10 previously-undetectable GE cancer and healthy control patients.

123 samples yielded three best-performing capture::detection pairs, termed “second-generation,”
124 which use rabbit monoclonal antibodies 34H7 and 62H12 as capture reagents and either Ab6 or
125 homodimeric form of Nb5 (Nb5-5LL) as detector (**Fig. 2a-c, S13-S16**). Adding detergent further
126 improved performance by limiting bead aggregation and improving bead loading into microwells.
127 These second-generation assays comprised capture::detection pairs of 34H7::Nb5-5LL,
128 62H12::Nb5-5LL, and 62H12::Ab6, achieving detection limits of 0.016-0.029 pg/mL (130-240 aM
129 trimeric ORF1p), and the four different reagents have predominantly non-overlapping epitopes in
130 binning experiments (34H7 and 62H12 partially overlap, **Fig. 2a-c, Table S1, S5-S6**). Somewhat
131 unexpectedly, analytical sensitivity did not perfectly correspond to clinical sensitivity. While the
132 second-generation assays demonstrated less than an order-of-magnitude improvement in
133 analytical sensitivity over the first-generation assay, they showed considerable improvement in
134 circulating ORF1p detectability over background in buffer in re-measured samples across a large
135 cohort of healthy and cancer patients (**Fig 2a, S17**). This difference may be due to differing
136 accessibilities of circulating ORF1p epitopes or to different nonspecific binding patterns in plasma.
137
138 Undetectable or extremely low ORF1p levels in healthy individuals could readily be discriminated
139 from measured ORF1p levels in ovarian cancer patients, resulting in a strong discriminatory ability
140 with single-marker models (area under the receiver operating characteristic curve, AUCs of 0.93
141 to 0.948, sensitivity of 41% to 81% at 98% specificity, **Fig. 2d top panel, Table S7**). This large
142 cohort included pre-treatment plasma samples from ovarian cancer patients (mostly high-grade
143 serous ovarian carcinoma) with age-matched controls (n=51-53 women, **Fig 2b**); again, second-
144 generation assays showed higher sensitivities while maintaining high specificities, notably
145 achieving detection of five out of six Stage I/II patients at >98% specificity. Furthermore,
146 multivariate models combining ORF1p (34H7::Nb5-5LL assay) with ovarian cancer biomarkers
147 CA125 and HE4 yielded improved diagnostic performance over these existing markers (CA125
148 and HE4 alone, AUC = 0.94, 59% sensitivity at 98% specificity; ORF1p, CA125, and HE4, AUC

149 = 0.98, 91% sensitivity at 98% specificity; **Fig 2d bottom panel, S18; Table S8**). While it is not
150 clear whether the low ORF1p levels detected in several healthy individuals is due to nonspecific
151 binding, true background levels of ORF1p, or an unappreciated pre-malignant state, several
152 positive healthy controls were positive by only one of the three second-generation assays (n=4
153 positive by only 62H12::Nb5-5LL and n=75 positive by only 62H12:Ab6), suggesting nonspecific
154 binding in at least some of these cases and the potential to improve specificity by combining data
155 from multiple assays. Our results indicate that by developing improved affinity reagents, we
156 achieved improved clinical sensitivity in detecting circulating ORF1p in cancer patients, with 83%
157 sensitivity at >98% specificity towards early detection of ovarian cancer.

158

159 To further validate our results, we developed a targeted proteomics approach to measure ORF1p
160 following affinity capture, with two distinct peptides measured vs. internal isotopically labeled
161 control peptides (**Fig. 2e**). With this assay, we applied much larger volumes of plasma (3-6 ml,
162 120-240 fold more than the 25 μ L used in Simoa assays) from a cohort of 10 patients, including
163 2 gastroesophageal (GE) cancer patients and one healthy control with very high ORF1p (230-
164 1230 pg/ml), two healthy controls with high ORF1p, (3-5 pg/ml), and 5 healthy controls with low
165 ORF1p (undetectable – 0.2 pg/ml). The results (**Fig. 2f, S19**) show strong correlation with Simoa,
166 providing further confidence in our results ($r=0.97-0.99$, $p<0.0001$).

167

168 Building on the improvements made through nanobody engineering in our second-generation
169 assays, we developed an expanded set of homodimeric, heterodimeric, and heterotrimeric anti-
170 ORF1p nanobodies and screened them in combination with 34H7 and 62H12 capture antibodies,
171 resulting in “third-generation” assays (**Figs. S9, S12, S20-21**). We noticed that reagents
172 containing Nb2 performed very well in SPR but poorly in Simoa detection, and we hypothesized
173 this was because Nb2 contains a lysine in the CDR, which would be biotinylated in the procedure,
174 reducing affinity. We therefore engineered the new reagents to be C-terminally biotinylated on

175 cysteine residues and varied linker sequence. Five of these assays, which utilize Nb2- and Nb9-
176 containing constructs, outperform our second-generation assays in a cohort of 25 GE cancer
177 patients with ORF1p measurements that were mostly undetectable previously, while maintaining
178 high specificity versus healthy individuals (**Fig. 2g, S21**).

179

180 To leverage more sensitive assays for ORF1p detection, we next tested ORF1p affinity reagents
181 from one of the second-generation Simoa assays on our recently developed Molecular On-bead
182 Signal Amplification for Individual Counting platform (MOSAIC, **Fig. 2h**). MOSAIC develops
183 localized on-bead signal from single captured molecules, in contrast to the microwell array format
184 in Simoa, and improves analytical sensitivity by an order of magnitude over Simoa via increasing
185 the number of beads counted(27). Furthermore, as the developed Simoa assays used only 25 μ L
186 plasma, we hypothesized that using larger plasma volumes would enhance ORF1p detectability
187 by increasing the number of analyte molecules present. By using a 20-fold higher sample volume
188 (500 μ L plasma) and the MOSAIC platform, we achieved ten-fold higher analytical sensitivity, with
189 a limit of detection of 0.002 pg/ml ORF1p (17 aM trimer, **Fig. S22**). Indeed, in a pilot cohort of
190 gastroesophageal cancer and healthy patients, ORF1p levels in nine of ten previously
191 undetectable cancer patients were readily discriminated from healthy individuals (**Fig. 2i**). Thus,
192 in addition to improved affinity reagents, using larger sample volumes and more analytically
193 sensitive technologies can further enhance both sensitivity and discrimination of circulating
194 ORF1p levels between healthy controls and patients with cancer.

195

196 To test whether ORF1p might be useful for monitoring therapeutic response, 19 patients with
197 gastroesophageal cancer were identified who had both detectable plasma ORF1p at diagnosis
198 as well as subsequent samples available collected during or after treatment. Primary tumors were
199 all adenocarcinoma and located in the esophagus (n=7), gastroesophageal junction (n=7) and
200 stomach (n=5). All patients received systemic therapy. A smaller fraction of patients also received

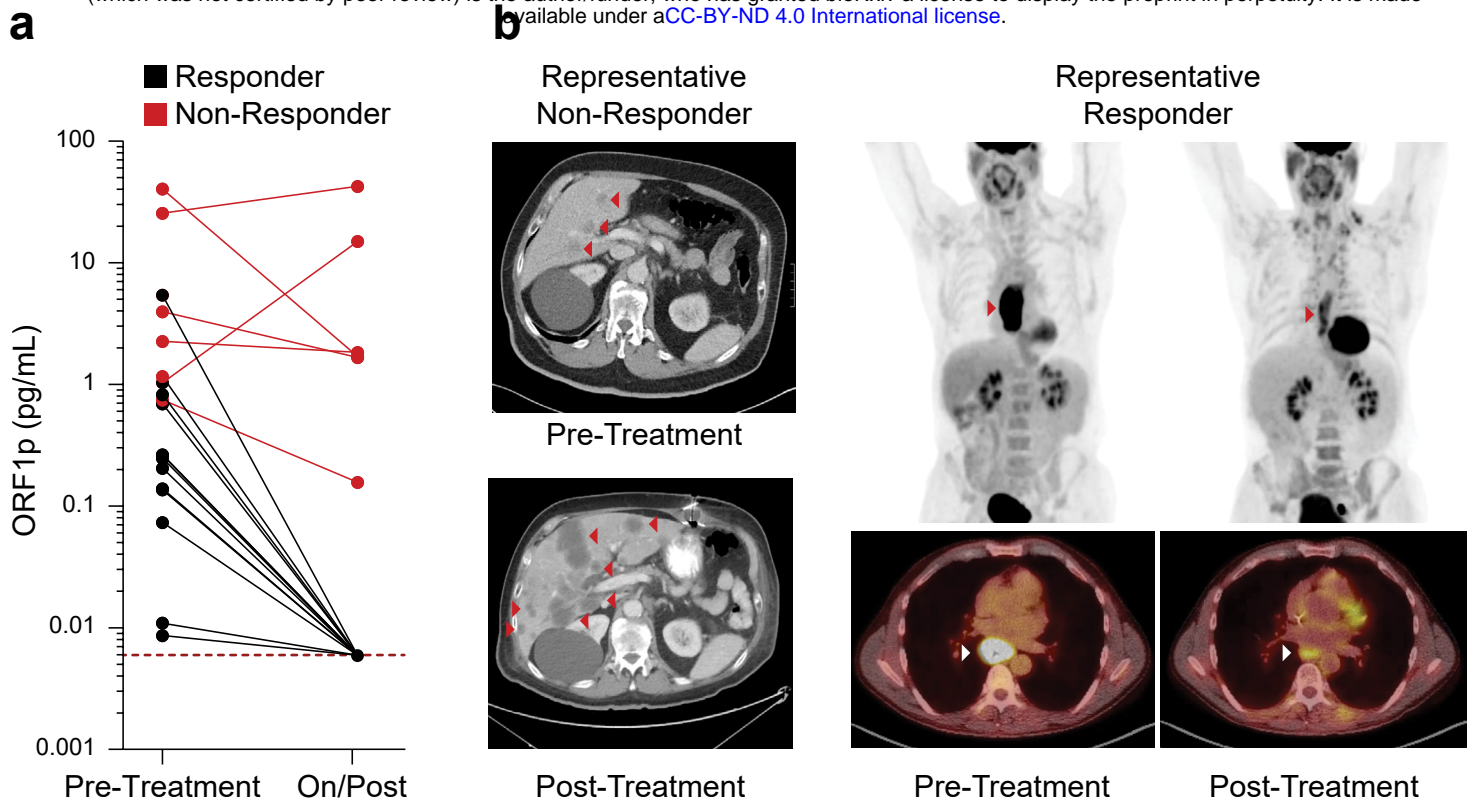


Figure 3. ORF1p is an early predictor of response in 19 gastroesophageal patients undergoing chemo/chemoradiotherapy. **a**, Plasma ORF1p as measured by all three second-generation Simoa assays before and during/post treatment; Responders and Non-Responders were characterized by post-therapy, pre-surgery imaging; $p < 0.0001$, Fisher's exact test. Non-Responders also have higher pre-treatment ORF1p than Responders ($p = 0.02$, t-test). **b**, Representative CT and PET-CT from patients in the cohort.

201 radiation and/or surgery (Supplement, **Table S9**). Clinical response ('Responders' and 'Non-
202 Responders') was determined by review of re-staging CT and PET-CT imaging. Over an average
203 of 465 days (range 98-1098), 12 patients died, six were alive at last follow-up (all 'Responders'),
204 and one was lost to follow-up. All 6 patients with detectable ORF1p at follow-up sampling, as
205 defined by positivity over background in two of three assays, were also Non-Responders by
206 imaging (**Fig. 3a**, $p < 0.0001$, Fisher's Exact test) and had reduced survival ($p = 0.001$ log-rank test
207 for overall survival). In contrast, in all 13 Responders, circulating ORF1p dropped to undetectable
208 levels post-treatment. Representative PET and PET-CT images are shown (**Fig. 3b**). Thus,
209 reduction in circulating ORF1p paralleled treatment response and survival, while persistent
210 circulating ORF1p corresponded to patients with refractory disease, indicating the predictive
211 potential of this marker.

212

213 **Discussion**

214 Taken together, our data reveal for the first time that circulating ORF1p is a multi-cancer protein
215 biomarker with potential utility across clinical paradigms, including early detection, risk
216 stratification, and treatment response. These assays are enabled by ultrasensitive single-
217 molecule detection technologies and high-quality affinity reagents, which are both required due
218 to the attomolar-to-femtomolar circulating levels of ORF1p in cancer patients. Iterative
219 improvements including optimized affinity reagents, buffer, and assay design yield highly sensitive
220 and specific assays. A 20-fold volume scale-up to 500 μ L appears promising for improving
221 sensitivity without obviously compromising specificity, and this volume remains much smaller than
222 a typical 5-10 mL blood draw and could be scaled further without limiting clinical applicability. The
223 data strongly suggest that these assays are measuring *bona fide* tumor-derived circulating ORF1p
224 for the following reasons: (1) four developed assays with predominantly non-overlapping high
225 affinity reagents all measure similar levels across hundreds of samples; (2) levels appear specific
226 to cancer patients, whose tumors overexpress ORF1p; (3) they correlate strongly with

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

231

232 The results expand our understanding that L1 expression is early and pervasive across
233 carcinomas from multiple organs and high-risk precursor lesions, including dysplastic Barrett's
234 esophagus, which is challenging to diagnose and manage. Circulating ORF1p shows promise in
235 early detection applications such as in ovarian cancer and may be more useful as part of a multi-
236 analyte detection test combined with, for example, cfDNA methylation, longitudinal CA125 in
237 ovarian cancer, or CEA in colorectal cancer(3,5,28). We demonstrate that ORF1p is an early
238 indicator of chemotherapeutic response in gastric and esophageal cancers at timepoints where
239 other parameters are often ambiguous, opening possibilities for monitoring minimal residual
240 disease or relapse. Importantly, ORF1p appears to provide a level of specificity for cancers not
241 achieved by other protein biomarkers, likely due to the unique biology of the retrotransposon, with
242 repression of L1 in normal somatic tissue(9,13,14). ORF1p is therefore attractive as a putative
243 "binary" cancer biomarker, in which a positive signal is highly specific for disease, with diagnostic
244 utility both in tissue and plasma.

245

246 The assays are cost-effective (<\$3 in consumables), rapid (<two hours), simple to perform,
247 scalable, and have clinical-grade coefficients of variation (<15%). Flow cytometers for MOSAIC
248 are common in clinical reference laboratories, and the assay could be modified for DNA-based
249 readout by qPCR or sequencing. Limitations of the current work include the relatively small
250 numbers of early-stage samples and a small and heterogeneous gastroesophageal therapeutic
251 cohort. Larger cohorts will be needed for further validation. Further optimizations to both assay
252 design and reagents will likely be possible, and larger cohorts are needed to further validate and

253 develop third generation Simoa assays and MOSAIC assays. Finally, it is unclear how ORF1p,
254 which is normally cytosolic, enters the blood and what clinicopathologic factors might affect these
255 levels. Future work will also be needed to understand whether there is a normal baseline level of
256 circulating ORF1p, as implied by the trace amounts seen when ORF1p was measured from much
257 larger volumes of plasma using targeted mass spectrometry, and what factors affect this level.

258

259 **Methods**

260 Provided in detail in *Supplementary Information*.

261

262 **Acknowledgements**

263 We thank Jeni Fairman for illustrations and Bert Vogelstein for plasma samples from colorectal
264 cancer patients. We are grateful to Phil Cole for resources for protein expression and purification
265 and helpful discussions and to Andrew Kruse and Edward Harvey for helpful discussions
266 regarding nanobodies. We thank Zuzana Tothova for helpful discussion and review of the
267 manuscript. This work was supported by the National Institutes of Health grants R01GM130680
268 (KHB), K08DK129824 (MST), F32EB029777 (CW), R01CA240924 (DTT), U01CA228963 (DTT),
269 P41 GM109824 (MPR, BTC), T32CA009216 (MST, GE), U01CA233364, U2CCA271871,
270 U01CA152990 (SJS), R01GM126170 (JL), P50CA228991 Ovarian SPORE (EJ, T-LW, I-MS,
271 RD); Break *Through* Cancer (KHB); Earlier.Org (KHB and DRW); Minnesota Ovarian Cancer
272 Alliance (KHB); DOD W81XWH-22-1-0852 (EJ, RD); Canary Foundation (RD); Gray Foundation
273 (EJ, RD); The Concord (MA) Detect Ovarian Cancer Early Fund (SJS), Good Ventures (Open
274 Philanthropy Project); Friends of Dana-Farber Cancer Institute; Dana-Farber Cancer Institute;
275 and the Dana-Farber/Harvard Cancer Center (DF/HCC); ACD-Biotechne (DTT, VD); Robert L.
276 Fine Cancer Research Foundation (DTT); Worldwide Cancer Research grant 19-0223
277 (JL); Robertson Therapeutic Development Fund (JL); Nile Albright Research Foundation
278 (BRR); Vincent Memorial Research Foundation (BRR); SU2C Gastric Cancer Interception
279 Research Team Grant (SU2C-AACR-DT-30-20, SJK, DTT, administered by the American
280 Association for Cancer Research, the Scientific Partner of SU2C).

281

282 **Author Contributions**

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

294

295 **Competing Interests**

296 MST has received consulting fees from ROME Therapeutics and Tessera Therapeutics that are
297 not related to this work. MST and JL have equity in ROME therapeutics. DTT has received
298 consulting fees from ROME Therapeutics, Tekla Capital, Ikena Oncology, Foundation Medicine,
299 Inc., NanoString Technologies, and Pfizer that are not related to this work. DTT is a founder and
300 has equity in ROME Therapeutics, PanTher Therapeutics and TellBio, Inc., which is not related
301 to this work. DTT receives research support from ACD-Biotechne, PureTech Health LLC, Ribon
302 Therapeutics, and Incyte, which was not used in this work. LMS declares the following
303 relationships: Consultant/advisory board: Novartis, Puma, G1 therapeutics, Daiichi Pharma, Astra
304 Zeneca; Institutional research support: Phillips, Merck, Genentech, Gilead, Eli Lilly. SJK declares
305 Consulting/advisory: Eli Lilly, Merck, BMS, Novartis, Astellas, AstraZeneca, Daiichi-Sankyo,
306 Novartis, Sanofi-Aventis, Natera, Exact Sciences, Mersana. Stock/Equity: Turning Point
307 Therapeutics, Nuvalent. BRR serves on SAB for VincenTech and receives research support from
308 Novartis Institutes for Biomedical Research that are not related to this work. DRW has a financial
309 interest in Quanterix Corporation, a company that develops an ultra-sensitive digital immunoassay
310 platform. He is an inventor of the Simoa technology, a founder of the company and also serves
311 on its Board of Directors. KHB declares relationships with Alamar Biosciences, Genscript,
312 Oncolinea/PrimeFour Therapeutics, ROME Therapeutics, Scaffold Therapeutics, Tessera
313 Therapeutics, and Transposon Therapeutics. MST and KHB receive royalties from sales of
314 ORF1p antibodies and MST, CW, PCF, KRM, BTC, MPR, JL, DRW, and KHB are inventors on a
315 patent related to this work. MST, LMS, SJK, BRR, and DTT's interests were reviewed and are
316 managed by Massachusetts General Hospital and Mass General Brigham in accordance with
317 their conflict-of-interest policies. Dr. Walt's interests were reviewed and are managed by Mass
318 General Brigham and Harvard University in accordance with their conflict-of-interest policies.
319 KHB's interests are managed by Dana-Farber Cancer Institute.

320

321 **References**

- 322 1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*
323 **2022**;72(1):7-33 doi 10.3322/caac.21708.
- 324 2. Sawyers CL. The cancer biomarker problem. *Nature* **2008**;452(7187):548-52 doi
325 10.1038/nature06913.
- 326 3. Crosby D, Bhatia S, Brindle KM, Coussens LM, Dive C, Emberton M, *et al.* Early
327 detection of cancer. *Science* **2022**;375(6586):eaay9040 doi 10.1126/science.aay9040.
- 328 4. Ignjatovic V, Geyer PE, Palaniappan KK, Chaaban JE, Omenn GS, Baker MS, *et al.*
329 Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection
330 to Achieving Translational Data. *J Proteome Res* **2019**;18(12):4085-97 doi
331 10.1021/acs.jproteome.9b00503.
- 332 5. Jamshidi A, Liu MC, Klein EA, Venn O, Hubbell E, Beausang JF, *et al.* Evaluation of cell-
333 free DNA approaches for multi-cancer early detection. *Cancer Cell* **2022**;40(12):1537-49
334 e12 doi 10.1016/j.ccell.2022.10.022.
- 335 6. Bast RC, Jr., Lu Z, Han CY, Lu KH, Anderson KS, Drescher CW, *et al.* Biomarkers and
336 Strategies for Early Detection of Ovarian Cancer. *Cancer Epidemiol Biomarkers Prev*
337 **2020**;29(12):2504-12 doi 10.1158/1055-9965.EPI-20-1057.
- 338 7. Menon U, Gentry-Maharaj A, Burnell M, Singh N, Ryan A, Karpinskyj C, *et al.* Ovarian
339 cancer population screening and mortality after long-term follow-up in the UK
340 Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled
341 trial. *Lancet* **2021**;397(10290):2182-93 doi 10.1016/S0140-6736(21)00731-5.

- 342 8. Rodic N, Sharma R, Sharma R, Zampella J, Dai L, Taylor MS, *et al.* Long interspersed
343 element-1 protein expression is a hallmark of many human cancers. *Am J Pathol*
344 **2014**;184(5):1280-6 doi 10.1016/j.ajpath.2014.01.007.
- 345 9. Ardeljan D, Wang X, Oghbaie M, Taylor MS, Husband D, Deshpande V, *et al.* LINE-1
346 ORF2p expression is nearly imperceptible in human cancers. *Mob DNA* **2020**;11:1 doi
347 10.1186/s13100-019-0191-2.
- 348 10. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ, 3rd, *et al.* Landscape of
349 somatic retrotransposition in human cancers. *Science* **2012**;337(6097):967-71 doi
350 10.1126/science.1222077.
- 351 11. Helman E, Lawrence MS, Stewart C, Sougnez C, Getz G, Meyerson M. Somatic
352 retrotransposition in human cancer revealed by whole-genome and exome sequencing.
353 *Genome Res* **2014**;24(7):1053-63 doi 10.1101/gr.163659.113.
- 354 12. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester
355 J, *et al.* Pan-cancer analysis of whole genomes identifies driver rearrangements
356 promoted by LINE-1 retrotransposition. *Nat Genet* **2020**;52(3):306-19 doi
357 10.1038/s41588-019-0562-0.
- 358 13. Burns KH. Transposable elements in cancer. *Nature Reviews Cancer* **2017**;17(7):415-24
359 doi 10.1038/nrc.2017.35.
- 360 14. McKerrow W, Doudican N, Frazzette N, Evans SA, Rocha A, Kagermazova L, *et al.*
361 LINE-1 Retrotransposon expression in cancerous, epithelial and neuronal cells revealed
362 by 5' single-cell RNA-Seq. *bioRxiv* **2022**:2021.01.19.427347 doi
363 10.1101/2021.01.19.427347.
- 364 15. Pisanic TR, 2nd, Asaka S, Lin SF, Yen TT, Sun H, Bahadirli-Talbot A, *et al.* Long
365 Interspersed Nuclear Element 1 Retrotransposons Become Deregulated during the
366 Development of Ovarian Cancer Precursor Lesions. *Am J Pathol* **2019**;189(3):513-20 doi
367 10.1016/j.ajpath.2018.11.005.
- 368 16. Cohen L, Cui N, Cai Y, Garden PM, Li X, Weitz DA, *et al.* Single Molecule Protein
369 Detection with Attomolar Sensitivity Using Droplet Digital Enzyme-Linked
370 Immunosorbent Assay. *ACS Nano* **2020**;14(8):9491-501 doi 10.1021/acsnano.0c02378.
- 371 17. Rajurkar M, Parikh AR, Solovyov A, You E, Kulkarni AS, Chu C, *et al.* Reverse
372 Transcriptase Inhibition Disrupts Repeat Element Life Cycle in Colorectal Cancer.
373 *Cancer Discov* **2022** doi 10.1158/2159-8290.CD-21-1117.
- 374 18. Rajurkar M, Parikh AR, Solovyov A, You E, Kulkarni AS, Chu C, *et al.* Reverse
375 Transcriptase Inhibition Disrupts Repeat Element Life Cycle in Colorectal Cancer.
376 *Cancer Discov* **2022**;12(6):1462-81 doi 10.1158/2159-8290.CD-21-1117.
- 377 19. Solyom S, Ewing AD, Rahrman EP, Doucet T, Nelson HH, Burns MB, *et al.* Extensive
378 somatic L1 retrotransposition in colorectal tumors. *Genome Res* **2012**;22(12):2328-38
379 doi 10.1101/gr.145235.112.
- 380 20. Ewing AD, Gacita A, Wood LD, Ma F, Xing D, Kim MS, *et al.* Widespread somatic L1
381 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res*
382 **2015**;25(10):1536-45 doi 10.1101/gr.196238.115.
- 383 21. Scott EC, Gardner EJ, Masood A, Chuang NT, Vertino PM, Devine SE. A hot L1
384 retrotransposon evades somatic repression and initiates human colorectal cancer.
385 *Genome Res* **2016**;26(6):745-55 doi 10.1101/gr.201814.115.
- 386 22. Cajuso T, Sulo P, Tanskanen T, Katainen R, Taira A, Hanninen UA, *et al.*
387 Retrotransposon insertions can initiate colorectal cancer and are associated with poor
388 survival. *Nat Commun* **2019**;10(1):4022 doi 10.1038/s41467-019-11770-0.
- 389 23. Doucet-O'Hare TT, Rodic N, Sharma R, Darbari I, Abril G, Choi JA, *et al.* LINE-1
390 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma.
391 *Proc Natl Acad Sci U S A* **2015**;112(35):E4894-900 doi 10.1073/pnas.1502474112.

- 392 24. Katz-Summercorn AC, Jammula S, Frangou A, Peneva I, O'Donovan M, Tripathi M, *et*
393 *al.* Multi-omic cross-sectional cohort study of pre-malignant Barrett's esophagus reveals
394 early structural variation and retrotransposon activity. *Nat Commun* **2022**;13(1):1407 doi
395 10.1038/s41467-022-28237-4.
- 396 25. Zhouchunyang X, Cochrane DR, Tessier-Coutier B, Leung S, Karnezis AN, Cheng AS,
397 *et al.* Expression of L1 retrotransposon open reading frame protein 1 (L1ORF1p) in
398 gynecologic cancers. *Hum Pathol* **2019** doi 10.1016/j.humpath.2019.06.001.
- 399 26. Carter V, LaCava J, Taylor MS, Liang SY, Mustelin C, Ukadike KC, *et al.* High
400 Prevalence and Disease Correlation of Autoantibodies Against p40 Encoded by Long
401 Interspersed Nuclear Elements in Systemic Lupus Erythematosus. *Arthritis Rheumatol*
402 **2020**;72(1):89-99 doi 10.1002/art.41054.
- 403 27. Wu C, Dougan TJ, Walt DR. High-Throughput, High-Multiplex Digital Protein Detection
404 with Attomolar Sensitivity. *ACS Nano* **2022** doi 10.1021/acsnano.1c08675.
- 405 28. Jacobs IJ, Menon U, Ryan A, Gentry-Maharaj A, Burnell M, Kalsi JK, *et al.* Ovarian
406 cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer
407 Screening (UKCTOCS): a randomised controlled trial. *Lancet* **2016**;387(10022):945-56
408 doi 10.1016/S0140-6736(15)01224-6.
409

410 **Figure Legends**

411 **Figure 1.** ORF1p expression is early and pervasive in carcinomas. **a**, ORF1p immunostaining in
412 a cohort of 211 colorectal cancers. **b**, Representative BE case: lesional cells overexpress p53,
413 the L1 RNA, and ORF1p. **c**, L1 RNA and ORF1p overexpression across a cohort of 72 consensus
414 BE cases and 51 carcinomas. **d**, Schematic of single-molecule protein detection by Simoa; a
415 second generation assay is shown. Antibody/nanobody-coated magnetic beads, present in
416 excess relative to target, capture single target ORF1p molecules. Enzyme-labeled detection
417 reagent (here, a homodimeric nanobody) is added, forming an “immunosandwich”, beads are
418 loaded into microwells that each can hold at most one bead, and ORF1p molecules are then
419 digitally detected using a fluorogenic substrate by counting “on” wells. First generation Simoa
420 instead uses Nb5-coated beads and Ab6 detector. **e**, First-generation ORF1p Simoa detects
421 plasma ORF1p with high specificity across major carcinomas. Pie charts indicate percentage of
422 samples with detectable levels; dashed red line, LOD. **, this control patient was thought to be
423 ‘healthy’ at the time blood was donated to the biobank but was later found to have prostate cancer
424 and lymphoma.
425

426 **Figure 2.** Improved detection of ORF1p with second- and third-generation assays. **a**, 34H7::Nb5-
427 5LL second-generation assay measurements across a multi-cancer cohort. **b**, Ovarian cancer
428 patients with age- and gender-matched controls in first- and second-generation assays; patients
429 are a subset of those in 2a; red dots: stage I disease, orange dots: stage II disease. **c**, Schematic
430 of affinity reagents used. 34H7 and 62H2 are custom mAbs; Nb5-5LL and Nb5-9 are an
431 engineered homodimeric and heterodimeric nanobodies, respectively. **d**, ROC curves with single
432 marker ORF1p across all healthy and ovarian cancer patients (top, n=128-132 cancer, 447-455
433 healthy), and multivariate models for ovarian (bottom, n=51-53 cancer, 50 healthy). **e**, Targeted
434 proteomics measurements of plasma ORF1p from a gastric cancer patient using two quantotypic
435 peptides (LSFISEGEIK and NLEECIR) with internal standards. **f**, Correlation between measured
436 ORF1p by Simoa and targeted proteomics assays; $r=0.97$ (Simoa vs LSFISEGEII) and $r=0.99$
437 (Simoa vs NLEECIR, t test), $p<0.0001$ for both. **g**, Comparison of 2nd and 3rd generation Simoa
438 assays (25 μ L) in 25 mostly undetectable gastroesophageal (GE) cancer and healthy control
439 patients. **h**, Schematic of MOSAIC assays. Captured single molecule “immunosandwiches” are
440 formed analogously to Simoa assays. DNA-conjugated streptavidin enables rolling circle
441 amplification to be carried out, generating a strong local fluorescent signal on the bead surface,
442 and then “on” and “off” beads are quantified by flow cytometry. **i**, 37H7::Nb5-5LL MOSAIC and
443 Simoa assays in 10 previously-undetectable GE cancer and healthy control patients.

444

445 **Figure 3.** ORF1p is an early predictor of response in 19 gastroesophageal patients undergoing
446 chemo/chemoradiotherapy. **a**, Plasma ORF1p as measured by all three second-generation
447 Simoa assays before and during/post treatment; Responders and Non-Responders were
448 characterized by post-therapy, pre-surgery imaging; $p<0.0001$, Fisher’s exact test. Non-
449 Responders also have higher pre-treatment ORF1p than Responders ($p=0.02$, t-test). **b**,
450 Representative CT and PET-CT from patients in the cohort.