

# **A novel proteomics-based plasma test for early detection of multiple cancers in the general population**

Bogdan Budnik\*, Hossein Amirkhani, Mohammad H. Forouzanfar, Ashkan Afshin\*

\*Corresponding authors: [bbudnik@novelna.com](mailto:bbudnik@novelna.com), [ashkan@novelna.com](mailto:ashkan@novelna.com)

## **Background**

Early detection of cancer is crucial for reducing the global burden of cancer and saving lives, but effective screening tests for many cancers do not exist. Genomics-based liquid biopsy tests for screening multiple cancers at once have been developed, but they have low sensitivity for early-stage cancers and are expensive. Recent advancements in measuring protein abundances in plasma offer new opportunities for developing multi-cancer screening tests.

## **Methods**

We collected plasma samples from 440 individuals, healthy and diagnosed with 18 various types of early-stage solid tumours. Using Proximity Extension Assay, we measured more than 3000 high and low-abundance proteins in each sample. Then, using a multi-step statistical approach, we identified a limited set of proteins that could detect early-stage cancers and their tissue of origin with high diagnostic accuracy.

## **Findings**

Our sex-specific cancer detection consisting of 10 proteins showed high accuracy for both males (AUC: 0.98) and females ((AUC: 0.983). At stage I and at the specificity of 99%, our detection panels were able to identify 89% of cancers among males and 75% of cancers among females.

Our sex-specific localization panels consisted of 150 proteins and were able to identify the tissue of origin of most cancers in more than 80% of cases. The analysis of the plasma concentrations of proteins selected showed that almost all the proteins were in the low-concentration part of the human plasma proteome.

## **Interpretation**

The proteome-based screening test showed promising performance compared with other technologies and could be a starting point for developing a new generation of screening tests for the early detection of cancer and potentially other chronic diseases. This new approach may provide a more accessible and cost-effective alternative to existing methods for cancer detection and may help reduce cancer mortality rates globally.

## 1 **Introduction**

2 Cancer is a leading cause of mortality globally, accounting for one in every six deaths.<sup>1</sup> In the  
3 absence of established risk factors for many cancers, early detection and early treatment remain  
4 the cornerstone of clinical and public health strategies for reducing the global burden of cancer  
5 and saving lives. However, currently, no effective test exists for the early detection of many  
6 cancers. Nearly 60% of cancer-related deaths are due to cancers for which no screening test exists.<sup>2</sup>  
7 Additionally, existing screening tests (i.e., colonoscopy, CT scan, mammography, pap test) have  
8 major limitations, including invasiveness, high cost, and low accuracy for early stages.  
9 Liquid biopsy, the analysis of biomarkers in non-solid specimens, has emerged as a promising  
10 approach for developing novel biomarkers.<sup>3</sup> In recent years, efforts have been made to develop a  
11 genomics-based liquid biopsy test for screening multiple cancers at once.<sup>4 5</sup> For example, a blood  
12 test has been developed to identify the presence of over 50 cancers based on their methylation  
13 signatures in cell-free DNA.<sup>6</sup> However, these genomics-based multi-cancer tests have shown low  
14 sensitivity for early-stage cancers (<50%).<sup>7</sup> Additionally, they are too expensive (>500 USD) to  
15 be covered by most insurance companies and incorporated as routine screening tests in the health  
16 care system.  
17 Protein biomarkers in the blood have the potential to be used for early detection and ongoing  
18 monitoring of diseases, but the current options have significant limitations due to a lack of  
19 diagnostic specificity.<sup>8</sup> In this paper, we explore the potential use of plasma proteins as biomarkers  
20 for solid tumours (excluding melanoma) in specific organs and the need to search for biomarkers  
21 in the depths of the proteome that are currently undetectable. We discuss the lower sensitivity of  
22 current protein assays compared to nucleic acid detection methods and the need to be able to detect  
23 very small amounts of proteins to identify early stages of cancer growth through liquid biopsy.

## 24 **Methods**

25 We collected plasma samples from a total of 440 individuals, healthy and those diagnosed with 18  
26 various types of early-stage solid tumours. The 18 solid tumours (14 tumours in men and 16 in  
27 women) were chosen because of the invaluable opportunity of the cure at early stages before local  
28 invasion or spreading metastases happens and a potential diagnostic test at the early stage could  
29 reduce the mortality and cost of the cancer significantly. Deep proteomics profiling of plasma  
30 samples was conducted using Proximity Extension Assay technology<sup>9-11</sup> and measured nearly  
31 3,000 proteins in each sample. Then, we attempted to identify cancer protein signatures for the  
32 purpose of cancer detection. Our approach involved two main steps. In the first step, we searched  
33 for a limited number of proteins that could identify any cancer in its early stages. In the second  
34 step, we classified each type of cancer against the others to find a cancer-specific signature for  
35 localization (i.e., tissue of origin). Both steps were done for male and female samples separately  
36 and probability-based score was calculated at each step for a person's protein array sample and the  
37 patient was classified by the calculated score. As with other biomarker studies, our experiment  
38 size was relatively small.<sup>12</sup> Therefore, we used a multi-step approach to define a minimal set of  
39 proteins (also known as the "best set") that could classify samples from a stable model and be  
40 generalizable. To select the features forming the minimum set, 100 bootstrap samples of the  
41 original dataset were analysed by logistic regressions with L1 penalty to identify the proteins with  
42 stronger association and best statistical significance for all cancers, pan cancer analysis, and  
43 differentiation of each cancer from other solid tumours (localization). The use of the L1 penalty  
44 ensures the sparsity of the selected biomarkers, preventing the simultaneous selection of correlated  
45 biomarkers. A leave-one-out validation was performed to evaluate the performance of the model  
46 with selected features. For cancer localization, the proteins were evaluated to identify the site of  
47 origin and/or subtype of the cancer for each cancer versus other cancers in the dataset. The top

48 cancer specific proteins with the strongest associations with the target cancers were selected  
49 iteratively, which resulted in different size minimum sets for the cancers with the overall highest  
50 average performance. Then, for a given sample, the predicted probabilities of it belonging to  
51 different cancers were calculated using the selected proteins, and the cancer with the highest  
52 probability score was determined as the predicted cancer.

53

## 54 **Results**

### 55 **The effect of protein correlation and sex**

56 Out of the 3072 proteins that were analysed, 280 did not pass the quality measurements and were  
57 excluded from further analysis. The abundance of proteins showed varying correlation levels, with  
58 positive correlations potentially being a result of shared biological pathways. Many protein pairs  
59 displayed high positive correlation, as indicated by the correlation matrix (Fig 1). However, highly  
60 correlated proteins could make the analysis unstable, despite being related to cancer. Additionally,  
61 different biological pathways contribute to the initiation and progression of cancer, which  
62 highlights the need to capture diverse cancer types through different pathways and potentially less  
63 correlated proteins.<sup>13</sup> Therefore, in our analysis, we only picked the most useful proteins out of  
64 internally correlated sets to be able to classify more patients.

65 We found that the protein-cancer association varied significantly between male and females (Fig  
66 1). A simple comparison between cancer and normal samples for men and women showed a poor  
67 correlation across protein profiles. Many proteins with a p-value of less than 0.001 in one sex  
68 showed no significant difference in the other one. An analysis of volcano plots showed minimal  
69 overlap between top 100 proteins selected based on their differential expression and their p-value.  
70 An analysis of the top 10 proteins using AUC (area under the curve) of the ROC curve (receiver

71 operating characteristic curve) showed the order of importance of the proteins varied between  
72 males and females. The 10 proteins with the best AUC to differentiate cancer from normal in male  
73 samples performed poorly in females and vice versa. Thus, despite the sample size limitation, we  
74 decided to perform the remainder of the analyses on male and female samples separately to capture  
75 the sex specific patterns.

76

### 77 **Optimal number of proteins for cancer detection and localization**

78 Finding the best sex-specific sets to identify a cancer was performed in two steps. At the first step,  
79 we detected the protein signature of any cancer (pan cancer/any cancer classifier) to classify any  
80 cancer from normal, followed by the second step, identifying the tissue of origin of cancer and  
81 cancer subtypes (i.e., small cell and non-small cell cancers of lung, and cervical and endometrial  
82 cancers of uterine).

83 For each step, we tested several numbers of proteins and expected to see an increasing performance  
84 of the model by AUC with more proteins to capture different cancer populations. The model  
85 performance increased quickly with adding a few more proteins but after 10 proteins in any cancer  
86 model, no more improvement in AUC was observed (Fig 2). A similar phenomenon was observed  
87 in the cancer localization step (Fig 4). Since the performance of different predictive models for  
88 specific cancers plateaued at different numbers, we employed an algorithm to pick the proteins-  
89 cancer pairs for additional proteins if improved overall AUC more. For the localization panels, the  
90 highest performance was observed at the level of 150 proteins. As expected, different numbers of  
91 proteins were picked as the best set by the algorithm. Moreover, some proteins were helpful in  
92 localizing more than one cancer.

93

## 94 **Sex-specific cancer detection panel**

95 Our final detection panels for male and female each consisted of 10 proteins that were differentially  
96 expressed among normal and cancer plasma samples (Fig 2). Each protein of the panel alone had  
97 a low to medium detection accuracy but when assessed in combination with other proteins as a  
98 panel they achieved a very high accuracy in detection of early-stage cancers (Fig 3). The proteins  
99 in the panel showed low to medium correlation indicating each protein contributing new  
100 information and presenting a different pathway to the panel.

101 Overall, at the detection step, our protein panels showed high sensitivity and specificity among  
102 males and females (Fig 3). At the specificity of 99%, the overall sensitivity of our test was 88%  
103 among males and 74% among females. We also observed high accuracy across all stages of cancer  
104 among males and females. At stage I and at the specificity of 99%, our panel was able to identify  
105 89% of cancers among males and 75% of cancers among females. The performance of the panels  
106 varied across cancers. Overall, some cancers were easier to detect (e.g., pancreas and kidney  
107 cancers). On the other hand, the detection of cancers like astrocytoma and thyroid was more  
108 challenging but achieved with relatively high accuracy after optimization.

109 To further evaluate the performance of the cancer set to detect any cancer, we excluded one cancer  
110 at a time and evaluated the fitted model on other cancers by the algorithm above on the excluded  
111 (unseen) cancer. Our analysis showed an acceptable performance with an AUC of, on average,  
112 above 0.9 for an unseen cancer (Fig S3). The cancer-out model has better performance on unseen  
113 cancers of pancreas and liver cancer but a lower performance to detect unseen thyroid cancer and  
114 astrocytoma. This shows that the protein set correlates very well with the cancer and serves as a  
115 cancer detection test. A plausible baseline cancer detection signature will enable efficient  
116 expansion to other cancers with a limited number of new samples added to the training set.



## 117 **Sex-specific cancer localization panel**

118 Our localization panels consisted of 150 proteins. Each sample was fed into separate cancer versus  
119 other cancers and the prediction probability for that cancer was calculated. The top two highest  
120 probability was used to identify the tissue of origin. The number of proteins allocated to each  
121 cancer was selected in a way that optimized the overall performance of the test. In males, the  
122 highest number of proteins was allocated to bladder cancer and the lowest number of proteins were  
123 allocated to liver cancer. In females, the highest and the lowest number of proteins were allocated  
124 to ovarian cancer and bladder cancer, respectively.

125 We evaluated the overall performance of the test by its ability to correctly classify the samples at  
126 the detection and localization level. Overall, at the specificity of 99% our sex-specific test showed  
127 the detection sensitivity of 88% for male and 74% for female. The localization accuracy of the  
128 tests was 89% for males and 77% for females.

129

## 130 **The role of downregulated and low-concentration proteins**

131 Through our analysis of 18 cancers and normal plasma samples, we found that only a few cancers  
132 can be uniquely identified by up-regulated proteins, which are typically preferred as biomarkers.  
133 We discovered that many cancers showed much higher specificity using downregulated proteins  
134 rather than just upregulated proteins. As the number of cancers included in a single pan-cancer test  
135 increases, it will be crucial to have both types of regulation biomarkers to achieve high cancer  
136 specificity among many different cancers.

137 Our evaluation of the plasma concentrations of proteins that were selected showed that almost all  
138 the proteins were in the low abundance group. The critical proteins such as CCL11 required no  
139 dilution, compared with high abundance proteins such as FGA which require 100,000-fold dilution

140 to be detected reliably by proximity technology. This highlights the importance of low-  
141 concentration proteins to see precancerous states and early stages where the tumour has little  
142 systemic impact and generated footprints.

143

## 144 **Discussions**

145 In this study, we showed that a measurement of limited set of plasma proteins could classify cancer  
146 samples from normal and differentiate different cancers. This finding is the foundation for a multi-  
147 cancer screening test for the early detection of 18 solid tumours that cover all major human organs  
148 of origin for such cancers at the earliest stage of their development with high accuracy. It is  
149 important to diagnose cancer at very early stages where curative treatments are achievable with  
150 surgery and available treatments. Additionally, for the first time to our knowledge, we found  
151 compelling evidence that the cancer protein signatures are most likely sex specific for all cancers.  
152 Our study also showed that biological signals for early-stage cancers are much more evident in the  
153 low-concentration part of the human plasma proteome. It was also promising to observe that a set  
154 of proteins could differentiate all cancers from normal and sensitive to detect unseen cancers

155 In our study, we analysed a range of proteins found in classical cancer pathways. However, we  
156 discovered that only a very small number of these proteins could be used as biomarkers for early-  
157 stage cancer. In contrast, many proteins that were effective biomarkers for early-stage cancer were  
158 found at low concentrations across the entire plasma proteome. This finding may be due to the fact  
159 that most of our knowledge about the role of proteins in cancer pathways comes from studies of  
160 transcriptome at the tissue level in advanced stages of cancer, and the expression of proteins at the  
161 mRNA and protein levels do not always correspond. In addition, the concentration of proteins in  
162 tissue and plasma may not be strongly correlated. Finally, our samples were mainly from early-

163 stage cancers, where classic cancer pathways may not be highly active. This finding has major  
164 implication for developing the next generation of diagnostics highlighting the role low-abundance  
165 protein in early detection of disease.

166 The proteome-based diagnostic test showed promising performance compared with other  
167 technologies such as circulating tumour DNA tests (ctDNA)<sup>14</sup> by significantly outperforming  
168 existing multi-cancer screening tests in detecting cancer across all stages (I, II, III) and among all  
169 types of cancers. At the specificity of 99% and in stage I of cancer, our test had a sensitivity that  
170 was much greater than Galleri<sup>15</sup> and CancerSEEK<sup>16</sup> tests. Additionally, our study demonstrated  
171 ability of our “best-test” to achieve much higher accuracy in identifying the tissue of origin of  
172 cancers in each sample in comparison to other tests. At the cancer-specific level, all our best-tests  
173 were more accurate than other available screening tests. Among the four screening tests that have  
174 received the highest recommendation (Level A) from the U.S. Preventive Services Task Force  
175 (colonoscopy for colon cancer, pap test for cervical cancer, mammography for breast cancer, and  
176 low-dose CT scan for lung cancer), only coloscopy and low-dose CT scan had an accuracy of  
177 above 90% for cancer detection. However, the sensitivity of our test for detecting early-stage  
178 cancer was still higher than the sensitivity of these tests.

179 Over the past decade, mRNA large-scale sequencing has provided a comprehensive view of gene  
180 expression in specific tissues, revealing the proteins that are present in different organs of the  
181 human body. The Human Protein Atlas is a useful resource for understanding mRNA and protein  
182 expression in multiple healthy tissues.<sup>17</sup> However, it is important to note that tissues are typically  
183 composed of complex assemblies of distinct cells that may have different functions and  
184 developmental histories. Increasing amounts of information about RNA and protein expression in  
185 specific cell types is now becoming available for the individual cells that make up tissues and

186 organs. A challenge in using protein detection for liquid biopsy is that cancer-specific protein  
187 biomarkers may be present at ultra-low levels in the blood.<sup>18</sup> This is because proteins that are  
188 present at high concentrations in the blood of healthy individuals are unlikely to be significantly  
189 increased in patients with early stages of the disease or at early recurrence. The long history of  
190 plasma proteome analysis by mass spectrometry show that even proteome coverage was increased  
191 from several hundred proteins thirty years ago to more than five thousand proteins based on latest  
192 development in chromatography separation technique and DIA (Data Independent Acquisition)  
193 type of acquisition.<sup>19</sup> Still the major problem of cheap and reproducible sample preparation  
194 protocols and reliably measuring proteins after first thousands of most abundant proteins prevent  
195 development of early stage multi cancer test by mass spectrometry at acceptable price per sample  
196 and general population scale. Thus, assays with greater sensitivity for biomarker proteins that are  
197 normally present at very low or undetectable levels in the blood may enable the detection of cancer  
198 at an earlier stage of the disease or even at premalignant stages. Our test is based on sensitive  
199 proximity assays that require the simultaneous binding of three separate antibodies. This ability to  
200 analyse plasma proteome profiles deeply and consistently allowed us to focus our attention on very  
201 low-abundant proteins, which we found to be the most precise and accurate biomarkers of early  
202 stages for all the cancers studied in our study. Advancing the PEA technology to measure ultra-  
203 low protein concentrations will provide better opportunities to detect and classify cancers at a very  
204 early stage and even at the precancerous stage.<sup>20</sup>

205 Our new generation protein-based plasma test has shown high sensitivity in detecting a variety of  
206 early-stage tumours in asymptomatic patients, making it a strong candidate for use as a population-  
207 wide screening tool that is not currently achievable with existing tests or techniques. Its high  
208 specificity can help alleviate concerns about causing harm to patients, and its low cost allows for

209 widespread implementation. To be suitable for large-scale use, a screening test must have high  
210 sensitivity and the ability to reduce mortality and morbidity, as well as acceptable for healthcare  
211 system cost. In the case of cancer screening, it is also essential for the test to have high specificity  
212 to avoid causing undue harm to patients. Our test exhibits these desirable qualities, making it a  
213 promising option for cancer screening. We expect that the combination of lower cost and higher  
214 accuracy in our test will facilitate its integration into the healthcare system and eventual inclusion  
215 in routine annual check-ups. Early detection of cancer has the potential to greatly reduce the  
216 societal burden of both health and financial costs. In fact, implementing such interventions can not  
217 only be cost-effective but can also result in cost savings for society.

218 Our approach has major strengths, including the total number of proteins measured and accuracy  
219 of such measurements across all measured proteins down to very low abundant proteins, the focus  
220 on early-stage tumours, the number of studied cancers that represents all major organs of unmet  
221 needs included in the study.

222 Limitations should also be considered, including small cohort size and the existence of  
223 comorbidities. Therefore, our test needs to be further validated in a larger population cohort before  
224 being widely accepted into healthcare systems across various populations. We also foresee the  
225 expansion of the test to include all cancers with unmet test needs.

## 226 **Conclusions**

227 In summary, the unique contributions of this study include: the analysis of the most extensive set  
228 of proteomics data available in a multi-cancer sample, the development of a cancer-specific protein  
229 signature for early-stage cancers that focuses on the baseline carcinogenic state rather than the end-  
230 stage tumour behaviours and human response, the identification of a sex-specific protein profiles  
231 and cancers signatures, the highlighting the importance of sex specific proteins set for early cancer

232 detection and importance of downregulated proteins as sensitive biomarkers at the early stage, and  
233 the demonstration of the feasibility and potential performance of this approach for early-stage  
234 diagnosis of all major cancers at the population level.

## References

1. Collaborators GBDCRF. The global burden of cancer attributable to risk factors, 2010-19: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2022;400(10352):563-91. doi: 10.1016/S0140-6736(22)01438-6
2. National Cancer Institute. Crunching Numbers: What Cancer Screening Statistics Really Tell Us 2018 [cited 2023. Available from: <https://www.cancer.gov/about-cancer/screening/research/what-screening-statistics-mean>.
3. Geyer PE, Kulak NA, Pichler G, et al. Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst* 2016;2(3):185-95. doi: 10.1016/j.cels.2016.02.015 [published Online First: 20160323]
4. Marrugo-Ramirez J, Mir M, Samitier J. Blood-Based Cancer Biomarkers in Liquid Biopsy: A Promising Non-Invasive Alternative to Tissue Biopsy. *Int J Mol Sci* 2018;19(10) doi: 10.3390/ijms19102877 [published Online First: 20180921]
5. Ding Z, Wang N, Ji N, et al. Proteomics technologies for cancer liquid biopsies. *Mol Cancer* 2022;21(1):53. doi: 10.1186/s12943-022-01526-8 [published Online First: 20220215]
6. Hasenleithner SO, Speicher MR. How to detect cancer early using cell-free DNA. *Cancer Cell* 2022;40(12):1464-66. doi: 10.1016/j.ccell.2022.11.009
7. Loomans-Kropp HA, Umar A, Minasian LM, et al. Multi-Cancer Early Detection Tests: Current Progress and Future Perspectives. *Cancer Epidemiol Biomarkers Prev* 2022;31(3):512-14. doi: 10.1158/1055-9965.EPI-21-1387
8. Hackshaw A, Cohen SS, Reichert H, et al. Estimating the population health impact of a multi-cancer early detection genomic blood test to complement existing screening in the US and

- UK. *Br J Cancer* 2021;125(10):1432-42. doi: 10.1038/s41416-021-01498-4 [published Online First: 20210823]
9. Fredriksson S, Gullberg M, Jarvius J, et al. Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol* 2002;20(5):473-7. doi: 10.1038/nbt0502-473
10. Lundberg M, Eriksson A, Tran B, et al. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res* 2011;39(15):e102. doi: 10.1093/nar/gkr424 [published Online First: 20110606]
11. Darmanis S, Nong RY, Hammond M, et al. Sensitive plasma protein analysis by microparticle-based proximity ligation assays. *Mol Cell Proteomics* 2010;9(2):327-35. doi: 10.1074/mcp.M900248-MCP200 [published Online First: 20091127]
12. Skates SJ, Gillette MA, LaBaer J, et al. Statistical design for biospecimen cohort size in proteomics-based biomarker discovery and verification studies. *J Proteome Res* 2013;12(12):5383-94. doi: 10.1021/pr400132j [published Online First: 20131028]
13. Sever R, Brugge JS. Signal transduction in cancer. *Cold Spring Harb Perspect Med* 2015;5(4) doi: 10.1101/cshperspect.a006098 [published Online First: 20150401]
14. Wan JCM, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* 2017;17(4):223-38. doi: 10.1038/nrc.2017.7 [published Online First: 20170224]
15. Klein EA, Richards D, Cohn A, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann Oncol* 2021;32(9):1167-77. doi: 10.1016/j.annonc.2021.05.806 [published Online First: 20210624]



16. Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 2018;359(6378):926-30. doi: 10.1126/science.aar3247 [published Online First: 20180118]
17. Rozenblatt-Rosen O, Stubbington MJT, Regev A, et al. The Human Cell Atlas: from vision to reality. *Nature* 2017;550(7677):451-53. doi: 10.1038/550451a
18. Landegren U, Hammond M. Cancer diagnostics based on plasma protein biomarkers: hard times but great expectations. *Mol Oncol* 2021;15(6):1715-26. doi: 10.1002/1878-0261.12809 [published Online First: 20201117]
19. Robbins JM, Peterson B, Schraner D, et al. Human plasma proteomic profiles indicative of cardiorespiratory fitness. *Nat Metab* 2021;3(6):786-97. doi: 10.1038/s42255-021-00400-z [published Online First: 20210527]
20. Petrer A, von Toerne C, Behler J, et al. Multiplatform Approach for Plasma Proteomics: Complementarity of Olink Proximity Extension Assay Technology to Mass Spectrometry-Based Protein Profiling. *J Proteome Res* 2021;20(1):751-62. doi: 10.1021/acs.jproteome.0c00641 [published Online First: 20201130]

## Figures

**Figure 1. The difference between protein biomarkers in males and females.** The Volcano plots showing differential abundances of proteins among both sexes, males, and females. Top 100 proteins based on their p-value highlighted in red (a). The correlation matrix of the top 100 proteins based on p-value among males, females, and both sexes (b). Ranking of the top 10 proteins based on AUC in males and females as well as the ranking of 10 randomly selected proteins in males and females (c). The scatter plot of p-values for each protein among males and females (d). The Venn diagram of top 100 proteins based on volcano plots in males and females (e).

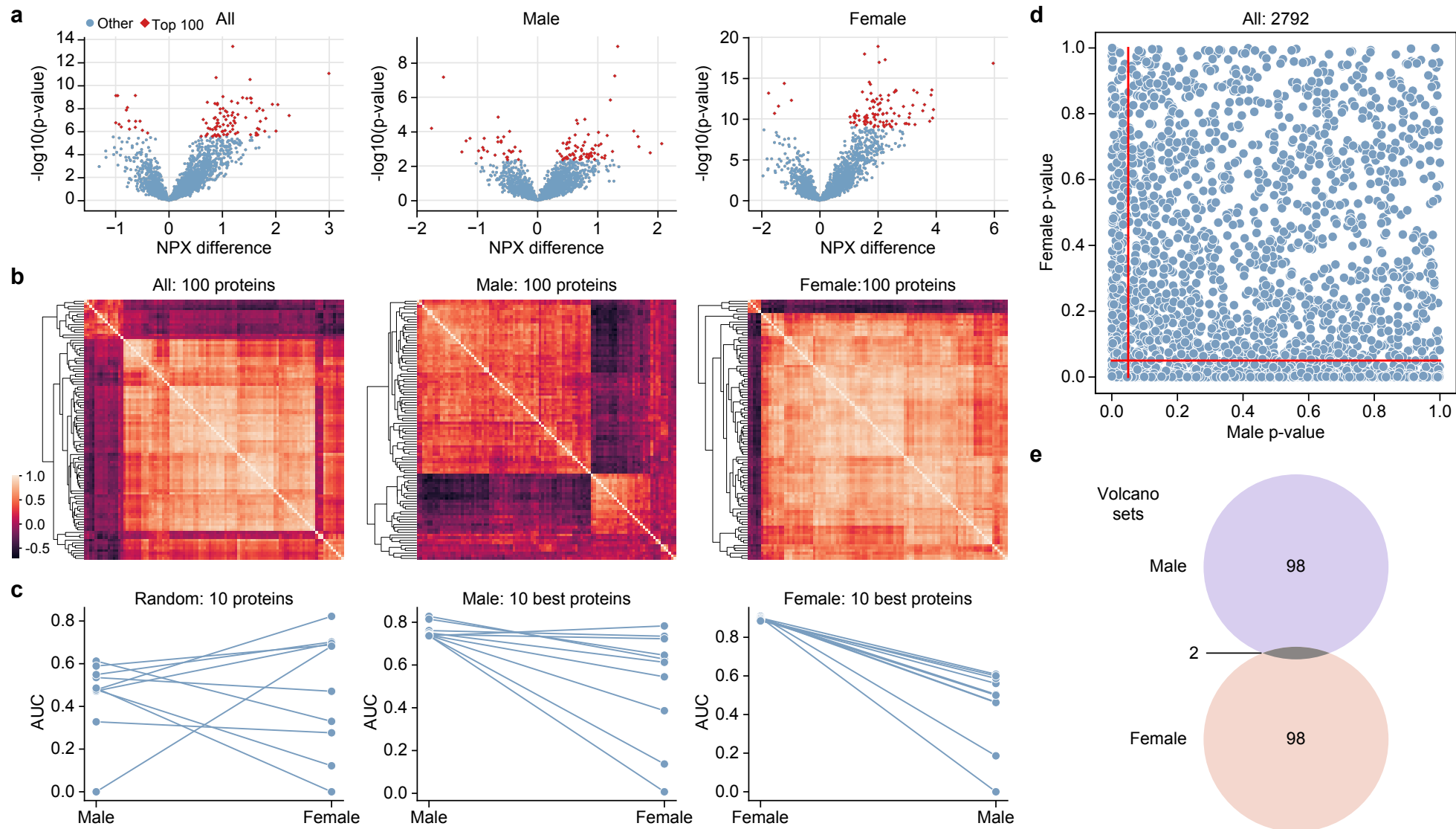
**Figure 2. Selection of cancer detection panel.** The relationship between the number of proteins included in the protein panel and performance of the panel (a). The correlation between proteins included the detection panel by sex (b).

**Figure 3. Performance of the detection panel.** ROC curve (receiver operating characteristic curve) of detection panel for males and females (a). Sensitivity of the detection panel at the specificity of 99% for each stage of the cancer for males and females (b). Sensitivity of detection panel at the specificity of 99% by stage for males and females (c).

**Figure 4. The selection and performance of the localization panels across cancers.** The relationship between the number of proteins included in the localization panel and the overall performance of the panel (a). The confusion matrix showing the performance of the of the test in correctly identifying the source of the cancer in the first prediction (b), and the bar chart of the overall performance of the localization panel (c).

# Figure 1. The difference between protein biomarkers in males and females.

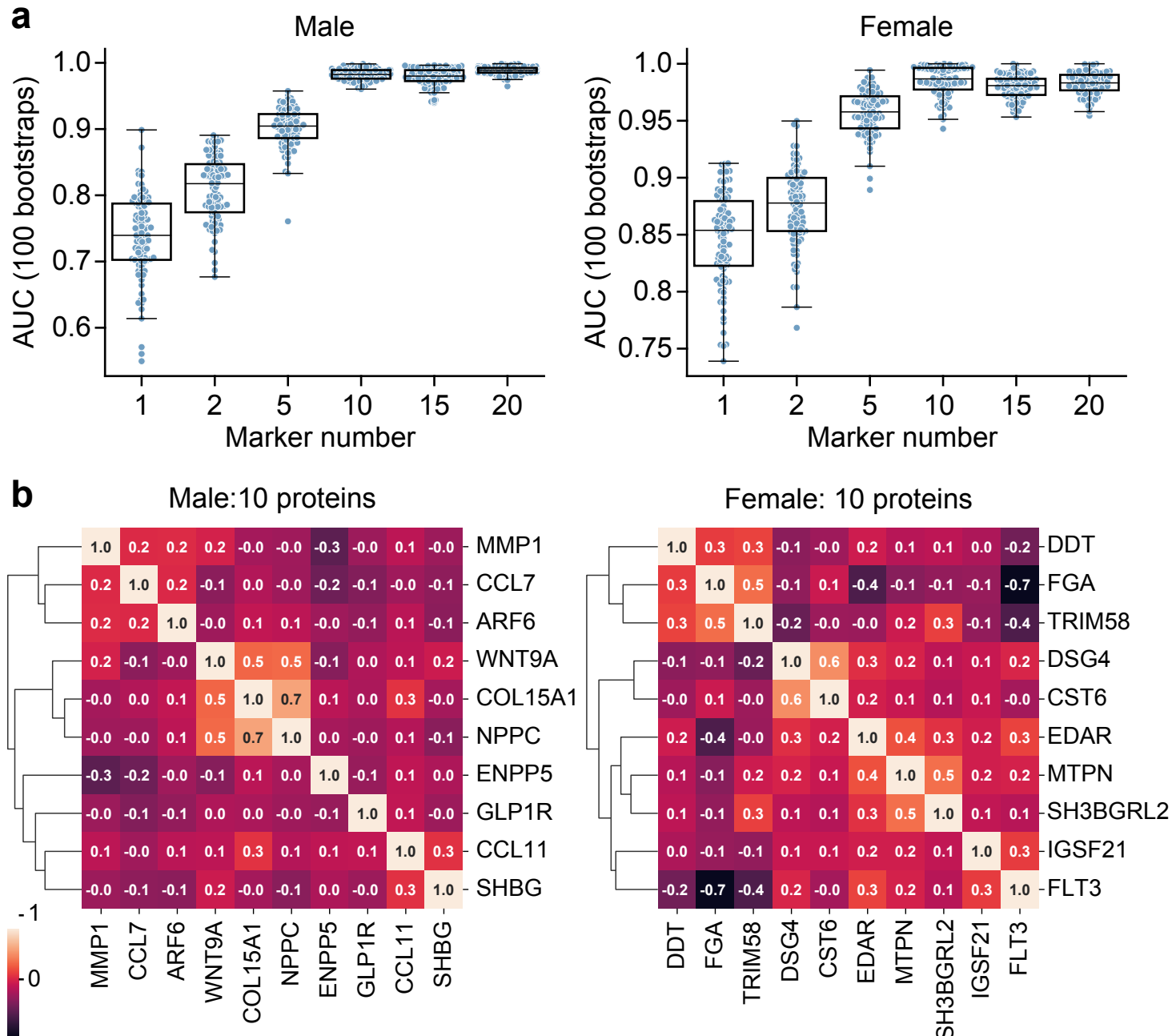
- (a) The Volcano plots showing differential abundances of proteins among both sexes, males, and females. Top 100 proteins based on their p-value highlighted in red.
- (b) The correlation matrix of the top 100 proteins based on p-value among males, females, and both sexes.
- (c) Ranking of the top 10 proteins based on AUC in males and females as well as the ranking of 10 randomly selected proteins in males and females.
- (d) The scatter plot of p-values for each protein among males and females.
- (e) The Venn diagram of top 100 proteins based on volcano plots in males and females.



## Figure 2. Selection of cancer detection panel.

(a) The relationship between the number of proteins included in the protein panel and performance of the panel.

(b) The correlation between proteins included the detection panel by sex.

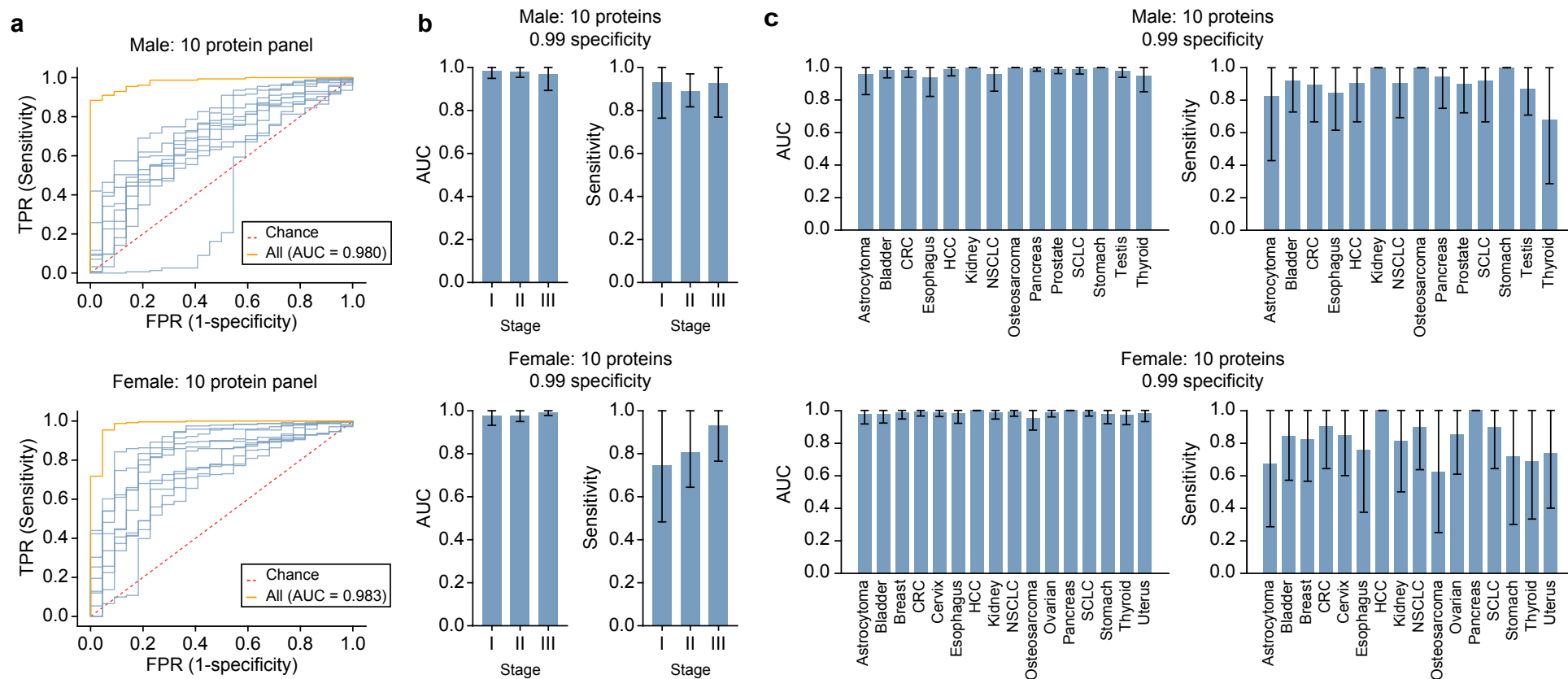


### Figure 3. Performance of the detection panel.

(a) ROC curve (receiver operating characteristic curve) of detection panel for males and females.

(b) Sensitivity of the detection panel at the specificity of 99% for each stage of the cancer for males and females.

(c) Sensitivity of detection panel at the specificity of 99% by stage for males and females.



### Figure 4. The selection and performance of the localization panels across cancers.

- (a) The relationship between the number of proteins included in the localization panel and the overall performance of the panel.  
 (b) The confusion matrix showing the performance of the test in correctly identifying the source of the cancer in the first prediction.  
 (c) The bar chart of the overall performance of the localization panel.

