

1  
2 **Title:**  
3

4 **Landscaping of Urine Proteome: Unlocking Diagnostic Potential**  
5 **and Overcoming Unique Challenges**  
6  
7

8  
9 **Authors:**

10  
11 Bogdan Budnik<sup>1\*</sup>, Hossein Amirkhani<sup>1</sup>, Klaus Weinberger<sup>1</sup>, Karine Sargsyan<sup>1,2,3</sup>, Mohammad H.  
12 Forouzanfar<sup>1†</sup>, Ashkan Afshin<sup>1\*†</sup>  
13  
14

15  
16  
17 **Affiliations:**

18 <sup>1</sup>Novelna Inc., Palo Alto, California, USA

19 <sup>2</sup>Medical University of Graz, Graz, Austria

20 <sup>3</sup>Cedars-Sinai Medical Center, Los Angeles, USA

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

22 †These authors contributed equally to this work  
23

24 **Abstract:**

25 This study explores the application of deep proteomic profiling to extract disease-specific features from  
26 urine. Early detection of cancer and other chronic disorders is crucial for better outcomes, but traditional  
27 diagnostics as well as emerging genomic-based diagnostics are expensive and invasive. Our research  
28 reveals that a select group of urinary proteins can accurately detect early-stage diseases with high  
29 sensitivity, surpassing current tests. While urine-based protein panels could offer cost-effective and  
30 accurate alternatives to current screening methods, kidney factors and blood urine barrier pathologies  
31 could pose significant challenges. New diagnostic technologies may emerge because of these findings,  
32 ushering in an era of early detection for cancer and chronic diseases.  
33

34 **One-Sentence Summary:**

35 Urine-based protein panels show distinct patterns in early disease detection, promising opportunities for  
36 advancing diagnostic tests

## 37 **Main Text**

38 Urine-based diagnostics are becoming increasingly popular as a non-invasive and cost-effective method  
39 of early detection of various types of diseases.(1-8) In recent years, urine-based cancer diagnostics have  
40 been developed for a variety of different types of cancer, including bladder, prostate, and colorectal  
41 cancer. For example, a urine-based test called UroSEEK has been developed for the early detection of  
42 bladder cancer. The test detects mutations in 11 cancer-associated genes in urine samples and has shown a  
43 sensitivity of 96% and specificity of 88% in early detection of urothelial cancer.(9) EpiCheck is another  
44 example of urine-based test detecting 15 DNA methylation targets in urine samples and has demonstrated  
45 a sensitivity of 68% at the specificity of 88% for bladder cancer.(9) Additionally, a 19-protein urinary  
46 biomarker model was recently developed and exhibited an 87% sensitivity and 65% specificity,  
47 outperforming traditional markers.(10)

48 While urine is seen as an extract and proxy of the serum, reliable measurement of proteins in urine to  
49 develop effective and widely available diagnostic tests is challenging.(11) The low abundance of  
50 biomarkers in urine, the physiology of kidney and urinary system, and the high variability in sample  
51 collection and processing can make it difficult to obtain accurate and consistent results and interpret the  
52 variation across different diseases and healthy populations. Additionally, identifying biomarkers that are  
53 specific and sensitive enough to detect diseases in low levels has been a major challenge for developing  
54 effective urine-based screening tests.(12)

55 Recent advancements in protein measurements, particularly with innovations like the Proximity Extension  
56 Assay, coupled with advanced AI algorithms, hold the transformative potential to reshape the landscape  
57 of urine-based diagnostics. These innovations not only enable the detection of low-abundance biomarkers  
58 in urine with remarkable sensitivity and specificity (13-19) but also open the door to a more  
59 comprehensive diagnostic approach. Rather than concentrating solely on individual biomarkers, we are  
60 now poised to identify the unique patterns associated with each disease within urine samples. This  
61 paradigm shift in diagnostics offers the promise of developing more robust and promising biomarker  
62 panels for the early detection of various diseases.

63 Additionally, the establishment of standardized protocols for urine collection, processing, and storage  
64 represents a critical step in our pursuit of reliable and accurate urine-based diagnostic tests. These  
65 protocols not only reduce variability but also significantly enhance the overall precision and reliability of  
66 the diagnostic process.

67 In this study, leveraging these advancements, we have embarked on landscaping of the urine proteome to  
68 develop novel diagnostics for a range of diseases.

69

## 70 **Urine proteome landscape in healthy individuals**

71 We utilized the PEA technology to detect a total of 3,072 proteins in urine samples, and successfully  
72 identified 2,850 of these proteins. This subset of 2,850 proteins formed the basis for the exploration of  
73 novel biomarkers in urine. Notably, all proteins were directly assessed from urine samples with a fourfold  
74 dilution, primarily aimed at minimizing the presence of salt in the samples. It is noteworthy that no  
75 further dilution was necessary for the PEA technology analysis, suggesting that the concentrations of the  
76 detected proteins generally remained at lower levels across the entire spectrum of proteins examined.

77 Our analysis indicated that the concentrations of certain proteins did not display significant differences in  
78 relation to the age of the individuals. As illustrated in Figure 1.A, specific proteins, such as VEGFB and

79 IL11, exhibited relatively higher urinary concentrations in older patients, while proteins like BOLA1,  
80 ARHGEF, and NFKB2 showed lower urinary concentrations in older individuals. It is pertinent to  
81 mention that only a limited number of proteins demonstrated a p-value of <0.01, and only one protein  
82 exhibited a p-value of 0.0016, indicating the overall subtle age-related differences in protein  
83 concentration in urine.

84 Furthermore, protein concentrations in urine exhibited variations based on the biological sex of the  
85 individuals. Figure 1.B depicts these differences in protein abundances between healthy males and  
86 females. Notably, Kallikreins, in particular, proved to be highly sensitive to the sex of the samples.  
87 Among these proteins, KLK3 and MSMB displayed significantly higher concentrations in male urine  
88 samples, while proteins KLK8 and KLK13 exhibited significantly higher concentrations in female  
89 samples. In total, our analysis identified more than 20 proteins with a significant p-value of <0.001,  
90 indicating a more pronounced contrast among patients when considering biological sex, in comparison to  
91 age.

92

### 93 **Urine proteome landscape across diseases**

94 The volcano plots, as illustrated in Figure 2, provide an insightful perspective on the differential  
95 expression of proteins in various medical conditions under examination in this study. Notably, it is  
96 apparent that most of these conditions exhibited an asymmetric pattern, characterized by  
97 overrepresentation of proteins in patients compared to normal samples, except for the case of melanoma.  
98 The observations collectively reveal three distinct patterns. Firstly, in the context of melanoma and to  
99 lesser extent endometrial cancer, the volcano plot exhibits a symmetrical pattern without a statistically  
100 significant changes in the great majority of the proteins. Secondly, in cancers of the cervix, ovary, and  
101 prostate as well as in MS, an asymmetrical shape is observed while majority of the protein changes were  
102 non-significant. Thirdly, the third pattern was related to cancers of kidney, bladder, and NASH where we  
103 observed an asymmetrical shape on the volcano plots and significant increase in the great majority of  
104 proteins. These findings underscore the substantial role of the kidneys in protein excretion in urine, which  
105 holds important implications for urine proteomics studies. Overall, our results consistently point to a  
106 distinctive pattern with statistically significant over-presentation of proteins in patients, offering potential  
107 utility for disease detection across a broad spectrum of conditions.

108 The analysis of correlations between different proteins has revealed intriguing patterns (as shown in  
109 Figure S1). To simplify the interpretation of these correlation matrices while ensuring their clarity, we  
110 chose to focus on the top 100 proteins with the lowest p-values in disease-normal comparisons. This  
111 selective approach allowed us to assess changes in the proteomic landscape across different diseases. For  
112 example, when examining the top 100 proteins that distinguish bladder, kidney, and NASH, we observe  
113 notably high correlations in normal samples. This suggests a potential non-differential over-leakage of  
114 these proteins into urine. In contrast, the correlation between proteins that vary between melanoma and  
115 endometrial cancer patients and their respective normal samples is considerably lower, indicating distinct  
116 urinary presentations. Furthermore, when comparing the protein correlations in healthy controls and  
117 patients, we observe significant differences in correlation patterns for the top 100 proteins in the cases of  
118 cervical cancer and prostate cancer. In summary, the volcano plots and correlation matrices together  
119 emphasize the unique urinary patterns associated with various diseases, underscoring their pivotal role in  
120 developing diagnostics for each disease category.

121 Figure S2 illustrates the distribution of over-represented and under-represented proteins across various  
122 diseases, similar to Figure S1, focusing only on the top 100 proteins with the lowest p-values in disease-  
123 normal comparisons. The data underscores the specificity of proteins associated with each disease. As

124 corroborated by the volcano plots, the majority of proteins were over-represented. Notably, among the top  
125 100 proteins with the lowest p-values, only melanoma, prostate cancer, endometrial cancer, and NASH  
126 exhibited under-represented proteins in urine. In diagram S2, the over-represented proteins were uniquely  
127 linked to individual diseases, with no proteins common to more than two conditions, suggesting a disease-  
128 specific protein signature. Figure S2B displays the under-represented proteins identified in the study,  
129 found in fewer diseases. These cases are highly disease-specific, emphasizing that urine proteomes  
130 exhibit distinctive protein changes in response to diseases, thereby underscoring the specificity and  
131 diagnostic potential of urine proteomics in the context of various diseases.

132

### 133 **Optimal disease-specific protein signature**

134 In Figure 3, we present the AUC scores achieved for each of the diseases under examination, with  
135 consideration to the number of proteins utilized. Generally, we observed that an optimal disease detection  
136 performance, characterized by AUC scores exceeding 0.9, required the incorporation of at least 7  
137 proteins. Notably, ovarian cancer was the sole exception, where the maximum AUC score was attained  
138 with a minimum set of 15 proteins. On the other hand, our analysis reveals a distinctive case in our  
139 study—namely, the detection of MS, where the highest AUC score achieved was 0.88, despite the  
140 utilization of 17 proteins. This unique instance suggests that this neurological disease might exhibit  
141 minimal alterations in its protein signature within urine, offering a plausible explanation for the relatively  
142 lower AUC score.

143 Figure 4 illustrates the performance of both individual proteins within the panel and the overall  
144 performance of the panel. As depicted in the figure, the overall performance of the panel surpassed that of  
145 any individual protein in most instances. This underscores the distinct contribution of each protein within  
146 the panel in characterizing the unique proteomic signature of the disease.

147 While the urinary panels exhibited high performance ( $AUC > 0.95$ ) for most diseases in the study, there  
148 was a noticeable trade-off between sensitivity and specificity across diseases. For instance, in the cases of  
149 prostate cancer, NASH, and melanoma, relatively high sensitivity could be achieved at 99% specificity.  
150 However, achieving high sensitivity for panels related to cervical and endometrial cancers proved more  
151 challenging and necessitated a significant reduction in panel specificity.

152 Figure 5 illustrates the quantified significance of individual proteins for various diseases. To determine  
153 the importance of each protein, we employed a random forest classifier trained on each disease using a  
154 feature set composed of concatenated proteins from all panels. The feature importance scores generated  
155 reflect the normalized total reduction in Gini impurities resulting from the utilization of a specific protein  
156 as a feature. Gini impurity serves as a metric to gauge how frequently a randomly selected element from  
157 the dataset would be incorrectly classified if it were randomly labeled according to the distribution of  
158 labels within the subset. The Gini impurity reaches its minimum value when all cases within the node  
159 exclusively belong to a single class. In this context, the scores indicate the influence of each protein in  
160 reducing the mixture or impurity of the samples.

161 The heatmap in Figure 5 effectively portrays the unique significance of protein sets for detecting specific  
162 diseases. The majority of disease detection sets typically comprise seven proteins, with the exception of  
163 the ovarian cancer protein set, which comprises 15 proteins to achieve an acceptable level of specificity. It  
164 is noteworthy that proteins exhibiting the highest predictive specificity for their target diseases are the  
165 most prevalent. Moreover, certain proteins play a crucial role in the detection of multiple diseases, in  
166 addition to their primary target. For instance, the protein VEGFD, included in the prostate cancer protein  
167 set, also proves to be highly significant for bladder cancer. In the case of bladder cancer, protein C9orf40

168 exhibited substantial importance, even though it wasn't selected for the final set. Another example is  
169 protein PPY, which demonstrated equivalent importance for both melanoma and multiple sclerosis  
170 detection.

171

## 172 **Interpretation**

173 Our landscaping of the urine proteome revealed the potential of utilizing low-abundance proteins in urine  
174 for the early detection of cancer, metabolic disorders, and neurological conditions. This finding forms the  
175 basis for the development of a range of non-invasive urine-based screening tests capable of identifying a  
176 variety of diseases. The early diagnosis of conditions such as cancer and metabolic disorders is essential  
177 for the development of effective treatments. Our study has demonstrated that distinct biological signals  
178 can be detected in urine even during the initial stages of diseases.

179 Furthermore, our findings indicate that proteins characterizing the urinary pattern of each disease is  
180 relatively unique and are not affected by the other conditions. This offers a promising avenue for the  
181 development of non-invasive urine-based tests designed to detect disease-specific proteins. These tests, if  
182 implemented, could enable early disease diagnosis, thereby preventing disease progression and  
183 facilitating the development of more effective treatments. In essence, our research has the potential to  
184 significantly impact healthcare by improving early disease detection and advancing public health.

185 Alterations in the urine proteome can be attributed to structural or physiological damage within the  
186 kidneys, potentially affecting the integrity of the blood-urine barrier. It is worth noting that distinguishing  
187 certain diseases that share similarities in protein classes with kidney damage can be challenging.  
188 Therefore, in the development of protein-based tests designed to detect a range of disease types, careful  
189 consideration of these complexities is imperative. Recent insights gleaned from the CKD273 biomarker  
190 panel, specifically tailored for the identification of impaired kidney function, have shed light on the  
191 predominant biomarkers, primarily comprising collagen fragments originating from modified  
192 extracellular matrix turnover.<sup>(9)</sup> This knowledge offers a practical opportunity to either accommodate or  
193 differentiate kidney-related changes when formulating diagnostic tests for other diseases.

194 Our new generation of protein-based urine test has exhibited remarkable sensitivity in the early detection  
195 of a variety of tumors in asymptomatic individuals, positioning it as a strong candidate for widespread  
196 screening—a role currently unattainable with existing methods. The non-invasive and cost-effective  
197 nature of urine testing makes it a practical option for screening large populations for cancer, especially  
198 among individuals with risk-elevating lifestyle factors or family histories. The potential for earlier  
199 detection and subsequent treatment holds promise for substantially improving patient outcomes.

200 In addition to our comprehensive protein measurement coverage and the accuracy of these measurements,  
201 even for proteins present in small quantities; our strengths encompass developing a machine learning  
202 platform for extracting unique features from a wide range of urinary proteins. Furthermore, our approach  
203 encompasses diseases with significant unmet diagnostic needs. There are also limitations to consider,  
204 including the small size of the cohort and the presence of comorbidities. Hence, we need to validate our  
205 protein panels in a larger population cohort before it can be widely accepted across a variety of  
206 populations. Additionally, the test should be evaluated for accuracy and precision in different populations.  
207 Finally, the cost and ease-of-use of the test should be determined.

208 In summary, this study presents several distinct contributions. These include an analysis of the most  
209 extensive proteomics dataset derived from urine, the formulation of a cancer-specific protein signature

210 tailored for early-stage cancers, with an emphasis on the baseline carcinogenic state rather than the later-  
211 stage tumor behavior and human response.

212 **References**

- 213
- 214 1. E. Rodriguez-Suarez, J. Siwy, P. Zurbig, H. Mischak, Urine as a source for clinical proteome  
215 analysis: from discovery to clinical application. *Biochim Biophys Acta* **1844**, 884-898 (2014).
- 216 2. A. Kentsis, Challenges and opportunities for discovery of disease biomarkers using urine  
217 proteomics. *Pediatr Int* **53**, 1-6 (2011).
- 218 3. C. J. Rosser *et al.*, Urinary protein biomarker panel for the detection of recurrent bladder cancer.  
219 *Cancer Epidemiol Biomarkers Prev* **23**, 1340-1345 (2014).
- 220 4. N. Davis *et al.*, A Novel Urine-Based Assay for Bladder Cancer Diagnosis: Multi-Institutional  
221 Validation Study. *Eur Urol Focus* **4**, 388-394 (2018).
- 222 5. M. Frantzi *et al.*, Development and Validation of Urine-based Peptide Biomarker Panels for  
223 Detecting Bladder Cancer in a Multi-center Study. *Clin Cancer Res* **22**, 4077-4086 (2016).
- 224 6. S. Thomas, L. Hao, W. A. Ricke, L. Li, Biomarker discovery in mass spectrometry-based urinary  
225 proteomics. *Proteomics Clin Appl* **10**, 358-370 (2016).
- 226 7. P. Kumar *et al.*, Highly sensitive and specific novel biomarkers for the diagnosis of transitional  
227 bladder carcinoma. *Oncotarget* **6**, 13539-13549 (2015).
- 228 8. W. S. Tan *et al.*, Novel urinary biomarkers for the detection of bladder cancer: A systematic  
229 review. *Cancer Treat Rev* **69**, 39-52 (2018).
- 230 9. A. Argiles *et al.*, CKD273, a new proteomics classifier assessing CKD and its prognosis. *PLoS*  
231 *One* **8**, e62837 (2013).
- 232 10. M. Frantzi *et al.*, Validation of diagnostic nomograms based on CE-MS urinary biomarkers to  
233 detect clinically significant prostate cancer. *World J Urol* **40**, 2195-2203 (2022).
- 234 11. N. Chebotareva *et al.*, Urinary Protein and Peptide Markers in Chronic Kidney Disease. *Int J Mol*  
235 *Sci* **22**, (2021).
- 236 12. Q. U. Ain, S. Muhammad, Y. Hai, L. Peiling, The role of urine and serum biomarkers in the early  
237 detection of ovarian epithelial tumours. *J Obstet Gynaecol* **42**, 3441-3449 (2022).
- 238 13. J. Daza *et al.*, Urine supernatant reveals a signature that predicts survival in clear-cell renal cell  
239 carcinoma. *BJU Int* **132**, 75-83 (2023).
- 240 14. J. Suh *et al.*, Next-generation Proteomics-Based Discovery, Verification, and Validation of Urine  
241 Biomarkers for Bladder Cancer Diagnosis. *Cancer Res Treat* **54**, 882-893 (2022).
- 242 15. T. Sun *et al.*, Diagnostic value of a comprehensive, urothelial carcinoma-specific next-generation  
243 sequencing panel in urine cytology and bladder tumor specimens. *Cancer Cytopathol* **129**, 537-  
244 547 (2021).
- 245 16. L. M. Chen *et al.*, External validation of a multiplex urinary protein panel for the detection of  
246 bladder cancer in a multicenter cohort. *Cancer Epidemiol Biomarkers Prev* **23**, 1804-1812  
247 (2014).
- 248 17. E. Schiffer *et al.*, Prediction of muscle-invasive bladder cancer using urinary proteomics. *Clin*  
249 *Cancer Res* **15**, 4935-4943 (2009).
- 250 18. C. K. Chen, J. Liao, M. S. Li, B. L. Khoo, Urine biopsy technologies: Cancer and beyond.  
251 *Theranostics* **10**, 7872-7888 (2020).
- 252 19. M. Maas, T. Todenhofer, P. C. Black, Urine biomarkers in bladder cancer - current status and  
253 future perspectives. *Nat Rev Urol* **20**, 597-614 (2023).
- 254

255 **Acknowledgments:**

256 **Funding:** *This work was supported by Novelna Inc.*

257 **Author contributions:** *All authors contributed to drafting the overall structure and flow of the*  
258 *manuscript. Subsequently, authors contributed subsections based on their domain of expertise. AA*  
259 *integrated subsections, incorporated comments and performed additional revisions.*

260 **Competing interests:** *The authors have been employed and/or hold shares in Novelna Inc.*

261 **Data and materials availability:** *The datasets generated during and/or analyzed during the current*  
262 *study are available from the corresponding author on reasonable request.*

263  
264  
265  
266

267 **Figures**

268 **Figure 1. Variations in Urine Proteome by Age and Sex.** Volcano plots depicting differential protein  
269 abundances in healthy adults aged above and below 35 years (a), and between adult males and females  
270 (b). Red points indicate proteins with the lowest p-values for statistical differences.

271 **Figure 2. Variations in Urine Proteome across Different Diseases.** Volcano plots showing differential  
272 protein abundances in healthy adults and patients with various conditions: bladder cancer (a), cervical  
273 cancer (b), endometrial cancer (c), kidney cancer (d), melanoma (e), multiple sclerosis (MS) (f),  
274 nonalcoholic steatohepatitis (NASH) (g), ovarian cancer (h), and prostate cancer (i). Red points represent  
275 proteins with the lowest p-values for statistical differences.

276 **Figure 3. The relationship between the size of the protein panel and the performance of panel.** The  
277 X-axis displays the number of proteins in the panel, while the Y-axis measures panel performance using  
278 the AUC for various conditions: bladder cancer (a), cervical cancer (b), endometrial cancer (c), kidney  
279 cancer (d), melanoma (e), multiple sclerosis (f), nonalcoholic steatohepatitis (NASH) (g), ovarian cancer  
280 (h), and prostate cancer (i).

281 **Figure 4. The performance of the selected protein panel.** ROC curves depict the performance of  
282 selected protein panels for each condition: bladder cancer (a), cervical cancer (b), endometrial cancer (c),  
283 kidney cancer (d), melanoma (e), multiple sclerosis (f), nonalcoholic steatohepatitis (NASH) (g), ovarian  
284 cancer (h), and prostate cancer (i). Blue lines represent individual protein performance, while the orange  
285 line represents the overall panel performance.

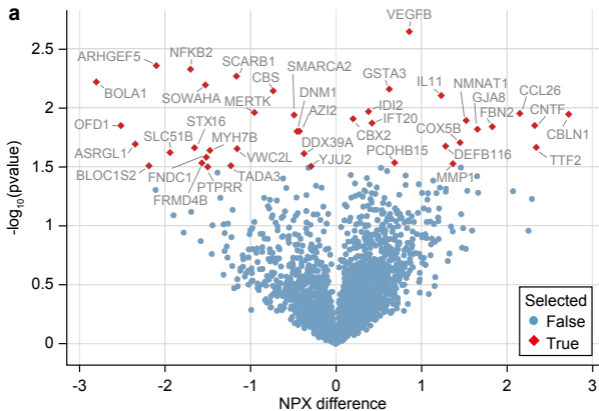
286  
287 **Figure 5. Distinctive Significance of Protein Sets for Disease Detection.** This heatmap illustrates the  
288 distinct importance of each protein of the panel in detecting other diseases.



Figure 1.

### Age-associated proteins

Positive: older than 35 years; negative: younger than 35 years



### Sex-associated proteins

Positive: male; negative: female

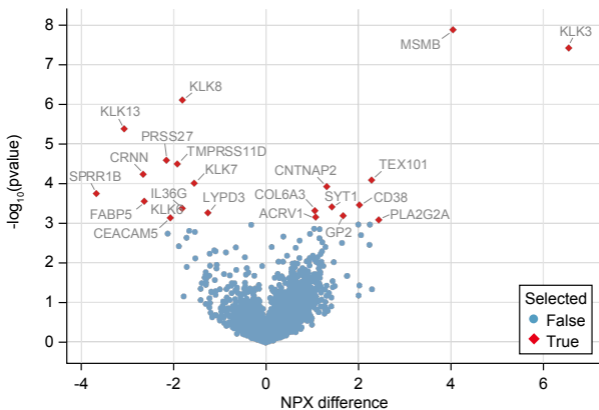


Figure 2.

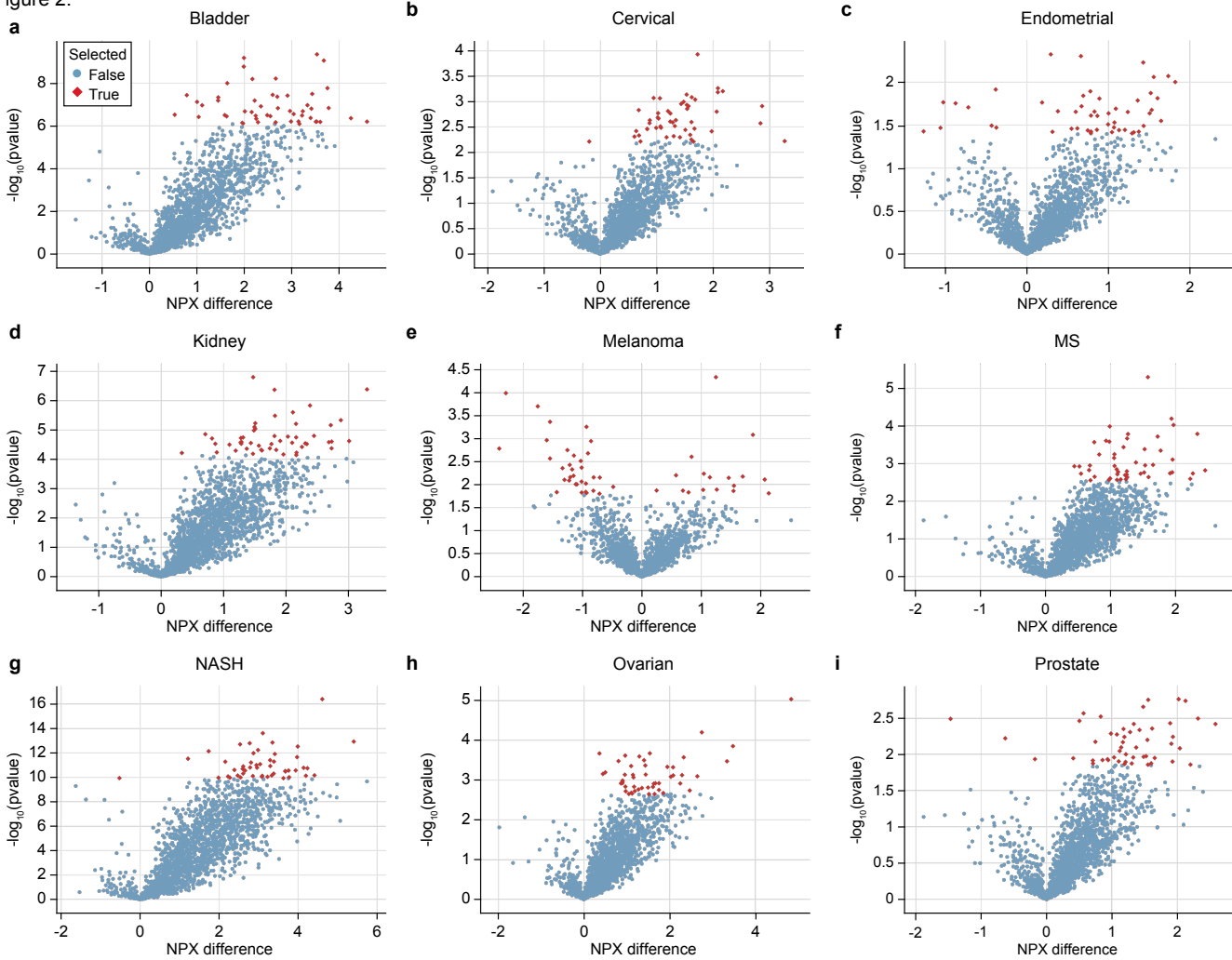
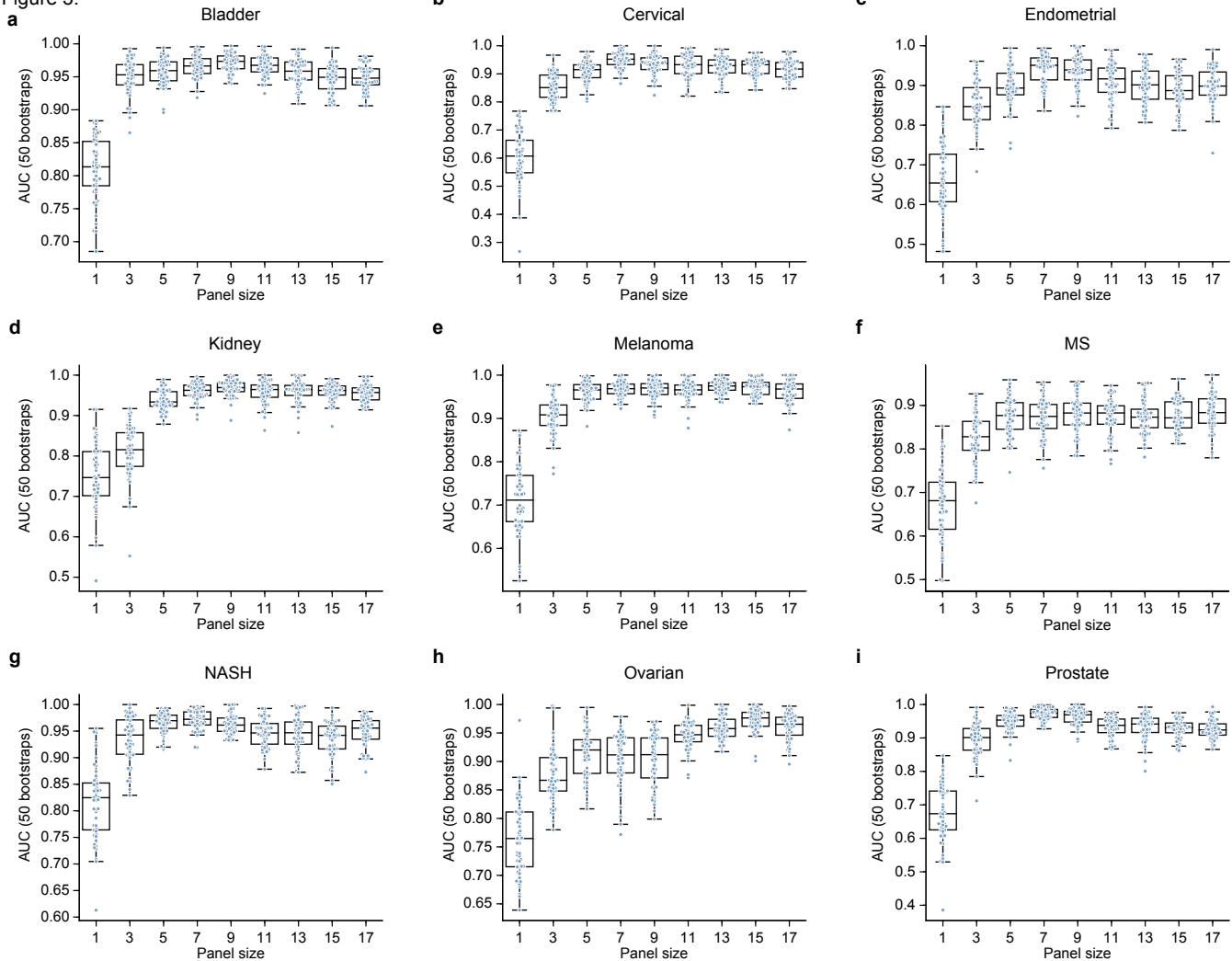
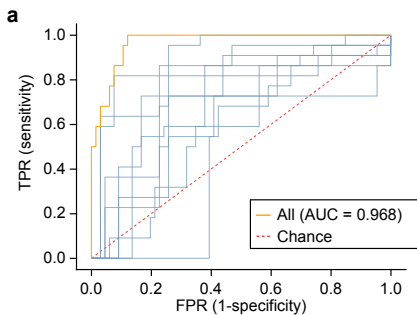
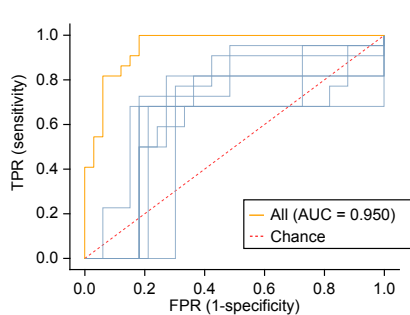
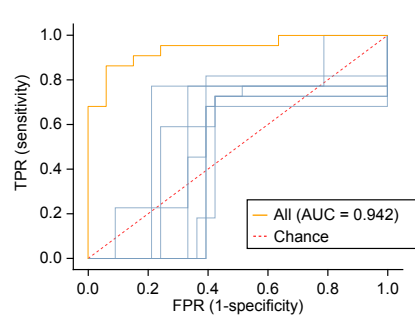
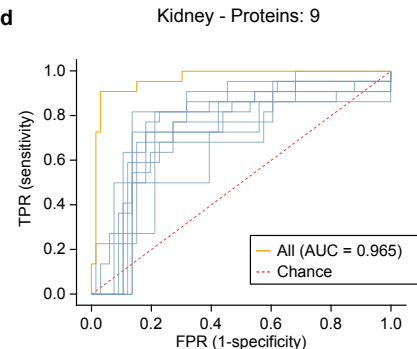
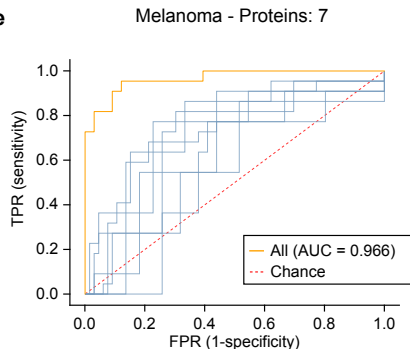
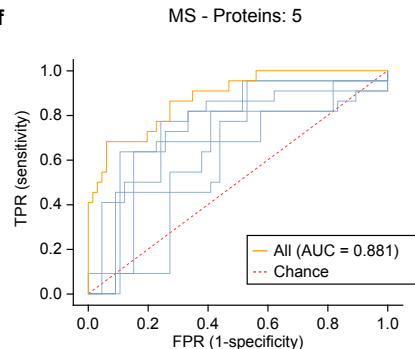
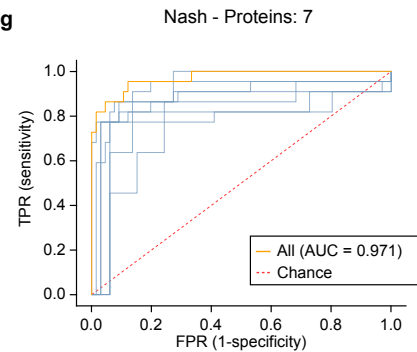
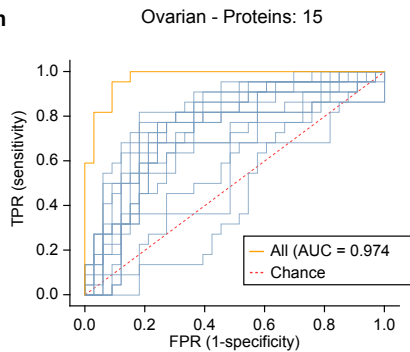


Figure 3.



**Figure 4.****Bladder - Proteins: 9****Cervical - Proteins: 7****Endometrial - Proteins: 7****Kidney - Proteins: 9****Melanoma - Proteins: 7****MS - Proteins: 5****Nash - Proteins: 7****Ovarian - Proteins: 15****Prostate - Proteins: 7**