

# DNA-framework-based multidimensional molecular classifiers for cancer diagnosis

Received: 13 May 2022

Accepted: 10 February 2023

Published online: 27 March 2023

 Check for updates

Fangfei Yin<sup>1,8</sup>, Haipei Zhao<sup>2,8</sup>, Shasha Lu<sup>2,3,8</sup>, Juwen Shen<sup>4,8</sup>, Min Li<sup>1</sup>,  
Xiuhai Mao<sup>1</sup>, Fan Li<sup>1</sup>, Jiye Shi<sup>5</sup>, Jiang Li<sup>5,6</sup>, Baijun Dong<sup>1</sup>, Wei Xue<sup>1</sup>,  
Xiaolei Zuo<sup>1,2</sup> ✉, Xiurong Yang<sup>2,7</sup> & Chunhai Fan<sup>1,2</sup>

A molecular classification of diseases that accurately reflects clinical behaviour lays the foundation of precision medicine. The development of *in silico* classifiers coupled with molecular implementation based on DNA reactions marks a key advance in more powerful molecular classification, but it nevertheless remains a challenge to process multiple molecular datatypes. Here we introduce a DNA-encoded molecular classifier that can physically implement the computational classification of multidimensional molecular clinical data. To produce unified electrochemical sensing signals across heterogeneous molecular binding events, we exploit DNA-framework-based programmable atom-like nanoparticles with *n* valence to develop valence-encoded signal reporters that enable linearity in translating virtually any biomolecular binding events to signal gains. Multidimensional molecular information in computational classification is thus precisely assigned weights for bioanalysis. We demonstrate the implementation of a molecular classifier based on programmable atom-like nanoparticles to perform biomarker panel screening and analyse a panel of six biomarkers across three-dimensional datatypes for a near-deterministic molecular taxonomy of prostate cancer patients.

Precision medicine calls for the development of a disease-specific molecular classification method that accurately reflects clinical behaviour<sup>1–4</sup>. A consistent research trend has been to obtain massive amounts of data on multidimensional molecules, including DNA/RNA, proteins and small molecules, which triggers growing interest in using multiple molecular datatypes to better classify diseases<sup>2,5–12</sup>. For example, the World Health Organization incorporated molecular indicators (for example, cyclin-dependent kinase inhibitor 2A/B homozygous

deletion and an isocitrate dehydrogenase mutant) for the classification of tumours of the central nervous system in the 2021 revision of the World Health Organization classification, providing illustrative examples of the new paradigm of integrated molecular classification<sup>13</sup>. Nevertheless, the heterogeneity of data obtained from various types of technologies accordingly increases and raises grand challenges in data integration and interpretation<sup>14–17</sup>. Examples include the heterogeneity in measurement sensitivity between RNA sequencing

<sup>1</sup>Institute of Molecular Medicine, Department of Urology, Shanghai Key Laboratory for Nucleic Acid Chemistry and Nanomedicine, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. <sup>2</sup>Frontiers Science Center for Transformative Molecules, School of Chemistry and Chemical Engineering, Zhangjiang Institute for Advanced Study, and National Center for Translational Medicine, Shanghai Jiao Tong University, Shanghai, China.

<sup>3</sup>School of Materials Science and Engineering, Suzhou University of Science and Technology, Suzhou, China. <sup>4</sup>Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai, China. <sup>5</sup>Division of Physical Biology, CAS Key Laboratory of Interfacial Physics and Technology, Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai, China.

<sup>6</sup>The Interdisciplinary Research Center, Shanghai Synchrotron Radiation Facility, Zhangjiang Laboratory, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, China. <sup>7</sup>State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, China. <sup>8</sup>These authors contributed equally: Fangfei Yin, Haipei Zhao, Shasha Lu, Juwen Shen. ✉e-mail: [zuoxiaolei@sjtu.edu.cn](mailto:zuoxiaolei@sjtu.edu.cn)

and chromatin immunoprecipitation sequencing, which causes significant gene expression variations that cannot be mirrored by chromatin modifications<sup>18</sup>. Hence, extensive computing-intensive data filtering and systematic normalization are indispensable to enable effective multidimensional data integration<sup>19,20</sup>.

Advances in developing *in silico* classifiers coupled with DNA-reaction-based molecular implementation provides a powerful and potentially generalizable means of molecular classification<sup>21,22</sup> (Fig. 1a,b). Seelig and coworkers designed an *in silico* classifier model that could translate parameters and mathematical functions into a class of DNA probe reporters to realize multi-gene classification for the diagnosis of early cancers and respiratory infections<sup>23</sup>. Similarly, Han and coworkers demonstrated a molecular classifier that could analyse different microRNAs (miRNAs) in lung cancer serum samples with a diagnosis precision of 86.4% (ref. <sup>24</sup>). The binding events between a target (DNA/RNA) and multiple single-stranded DNA reporters were uniformly translated to an assignment of weights for *in silico* analysis. However, the extension of this method to the dimensions of proteins or metabolic small molecules is difficult to implement due to the heterogeneous nature of these binding processes. A remaining challenge to realize DNA-based multidimensional molecular classifiers is thus to develop a signal reporter that can translate the heterogeneous, multidimensional molecular information into a unified output signal in a programmable manner (Fig. 1a).

The precision and programmable nature of the Watson–Crick base pairing of DNA delivers a spectrum of valence-controlled programmable atom-like nanostructures (PANs) for colloidal assembly with different compositions, sizes, chiralities and linearities<sup>25,26</sup>. In particular, self-assembled DNA tetrahedral frameworks (DTFs) provide a simple means to fabricate three-dimensional PANs with an ordered structure and versatile modification<sup>27–30</sup>. Here we introduce a PAN-based molecular classifier that can physically implement the computational classification of multidimensional molecular clinical data. The atom-like and programmable nature of a DTF supports the design of valence-controlled PAN signal reporters, resulting in linearity in translating virtually any class of molecular binding to unified electrochemical sensor signals (Supplementary Fig. 1). We demonstrate that the use of a PAN reporter allows precise weight assignment for multidimensional molecular information in computational classification, which is employed to interactively analyse a panel of six biomarkers across three-dimensional datatypes (RNA, protein and metabolic small molecule) for the classification of prostate cancer (PCa) patients. Moreover, we further developed a diagnosis panel screening system using PAN reporters for a classification related to the Gleason score.

## Construction and characterization of PAN reporters

Figure 1c,d shows the general design principle for a DNA-encoded molecular classifier, which physically implements an *in silico* classifier for multidimensional molecular data with an electrochemical sensing system (Fig. 1e,f and Supplementary Fig. 2). To produce unified electrochemical sensing signals across heterogeneous molecular binding events, we designed valence-encoded PAN reporters using DTF-based PANs with  $n$  valence capable of targeting each target molecule across multiple dimensions. More importantly, we envisioned that the use of valence-encoded PAN reporters might encode PANs with a defined number of signal moieties, allowing for the physical implementation of a weight assignment (for example, 1, 2 or  $n$ ) of the *in silico* classifier by anchoring 1, 2 or  $n$  signal moieties on a PAN reporter. Then, the signal gain from each target molecule would be linearly proportional to the number of signal moieties on the PAN reporter, which enables one to weigh each target molecule according to its importance in the *in silico* classifier.

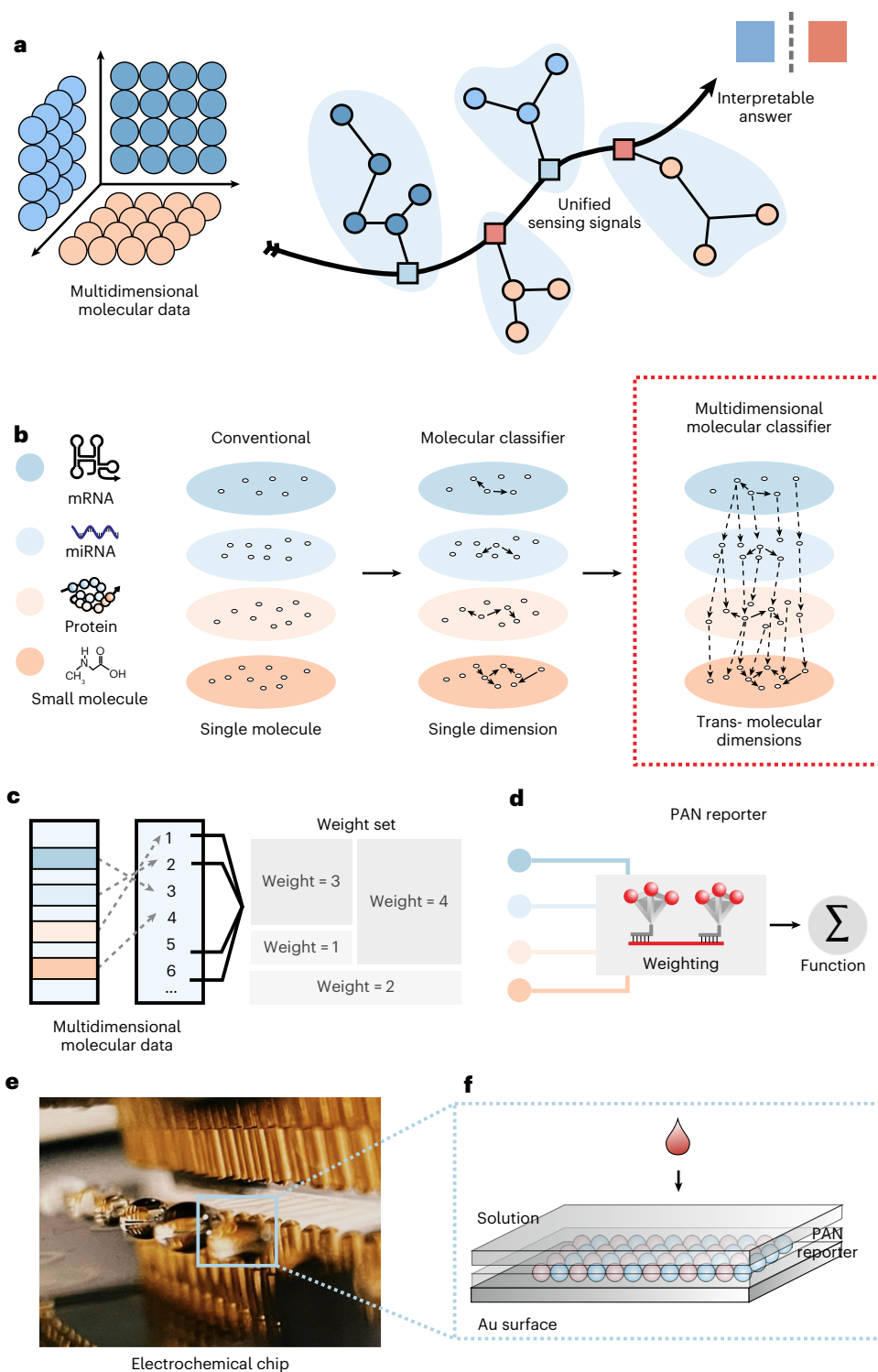
To fabricate DTF-based PAN reporters, we first assembled a DTF containing a handle DNA on a vertex by mixing seven DNA fragments of

58 nucleotides (58-nt) and one handle DNA-containing DNA fragment of 81-nt in stoichiometric equivalents in buffer (Supplementary Fig. 3). We heated the mixture to 95 °C and then rapidly cooled it to 4 °C. The DTFs were assembled with a high yield of ~95%, characterized by atomic force microscopy (AFM; Supplementary Fig. 4) and polyacrylamide gel electrophoresis (PAGE; Supplementary Fig. 5). We measured a typical edge length of ~12 nm for DTFs (37 base pairs for each edge), which was consistent with its theoretical length<sup>31</sup>. To form the PAN reporter containing more anchoring sites of signal moieties, we coupled one DTF to another DTF to form DTF dimer structures through the hybridization of a linker DNA and the handle DNA in the two DTFs. The DTF dimer we formed had a dumbbell-shaped structure (with ~95% yield), as shown by AFM and PAGE imaging (Fig. 2a and Supplementary Figs. 5 and 6).

To validate the valence-encoded PAN reporters, which may encode PANs with a defined number of signal moieties, we employed fluorophore labels (for example, cyanine-3 (Cy3)) as signal moieties on PAN reporters and characterized the precise number of signal moieties on the PAN reporters via the single-molecule technique, total internal reflection fluorescence microscopy (TIRFM). PAN reporters containing a defined number of signal moieties ( $n = 1, 2$  or  $n$ ) were realized by anchoring 1, 2 or  $n$  fluorophores on the vertices of a DTF dimer. We observed that the fluorescent intensity of the PAN reporters in bulk solution was linearly proportional to the number of the signal moieties ( $R^2 > 0.986$ ,  $R^2$ , R-squared; Fig. 2b). Moreover, aggregation-caused quenching cannot occur due to the separation of the fluorophores caused by the edge length of ~12 nm of the DTF<sup>32</sup> (Supplementary Fig. 7). Similarly, the fluorescence intensity of a single PAN reporter increased linearly with the increase of the number of Cy3 from one to six ( $R^2 > 0.998$ ) in TIRFM measurements (Fig. 2c and Supplementary Fig. 8a). Moreover, we observed stepwise single-molecule fluorescence photobleaching<sup>33</sup>, as six steps of photobleaching trace the PAN reporter containing six Cy3 labels. One to five steps of photobleaching trace the PAN reporters containing one to five Cy3 labels (Fig. 2d and Supplementary Figs. 8b and 9). Thus, the numbers of the signal moieties on PAN reporters were precisely controlled from one to six.

We next asked whether PAN reporters possess the orthogonality to accommodate programmed multicolour reporters. We anchored two types of fluorophores on PAN reporters with different emissions but without fluorescence resonance energy transfer (Supplementary Fig. 10) on the PAN reporters. To this end, six fluorophores were anchored on single PAN reporters with various number combinations of Alexa Fluor 488 and Cy5. The fluorescence intensity and steps of photobleaching of the PAN reporters were linearly proportional to the numbers of each type of fluorophore, without interference with each other (Fig. 2e and Supplementary Fig. 11). For example, when we anchored one Alexa Fluor 488 fluorophore and five Cy5 fluorophores on a single PAN reporter, we observed one step of a photobleaching trace of Alexa Fluor 488 and five steps of a photobleaching trace of Cy5. Thus, the anchoring sites of the PAN reporter were individually controlled, with a defined number of signal moieties even in the presence of multiple distinct signal moieties.

To demonstrate the generality in labelling multiple types of signal moieties on the PAN reporter, we anchored various signalling moieties, including gold nanoparticles (AuNPs; usually used as a signal moiety for mass or colorimetric output)<sup>34–36</sup> and enzymes (usually used as a signal moiety for fluorescent, colorimetric or electrocatalytic output)<sup>37–40</sup>. We visualized the spatial structure of the PAN reporter anchored with AuNPs via transmission electron microscopy (TEM) with a precisely controlled number from one to six (Fig. 2f and Supplementary Fig. 12). Interestingly, the spatial arrangement of AuNPs coincided with the vertices' arrangement on the DTF dimers, indicating that the signal moieties were well anchored on the PAN reporter. We then used horseradish peroxidase (HRP) as an example to anchor on the PAN reporter. The AFM images showed a precise number of HRP from one to six on the PAN reporter, as shown in Fig. 2g and Supplementary Fig. 13.



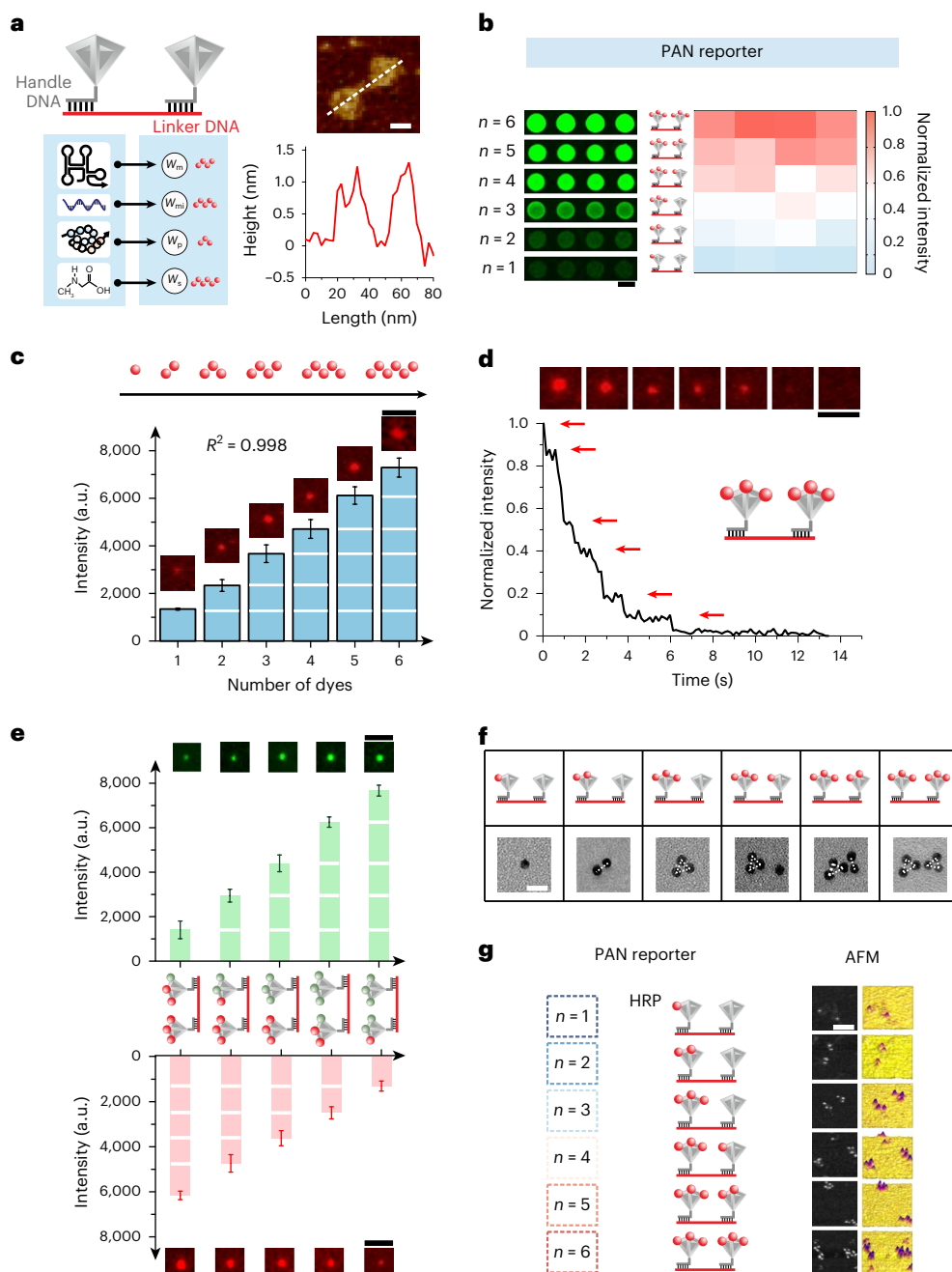
**Fig. 1 | A PAN-reporter-based multidimensional molecular classifier for cancer diagnosis.** **a**, Schematic illustration of multidimensional molecular classifier. The multidimensional molecular information is translated into a unified sensing signal. The classifier then handles the sensing signals to produce the interpretable yes/no answers. **b**, Scheme illustrating a conventional analysis based on one molecule, a single-dimensional molecular classifier and a multidimensional molecular classifier for cancer diagnosis. **c**, Schematic

showing how an in silico classifier was trained and validated with publicly available data. Each of the multidimensional molecular targets was assigned a weight to represent its importance. **d**, The  $n$ -valence PAN enabled the construction of a valency-controlled signal reporter for multidimensional molecules. **e, f**, Image (**e**) and schematic illustration (**f**) of the multidimensional molecular classifier coupled with an electrochemical array with 16 Au electrodes. The arrow represents a biological sample being applied.

### Molecular implementation of weight assignment

An in silico classifier realizes data classification via the assignment of a numerical weight to each piece of data that represents its

importance, and then summing the weighted result<sup>41</sup>. Analogously, a multidimensional molecular classifier translates each molecular input with a weighted sensing signal representing its importance

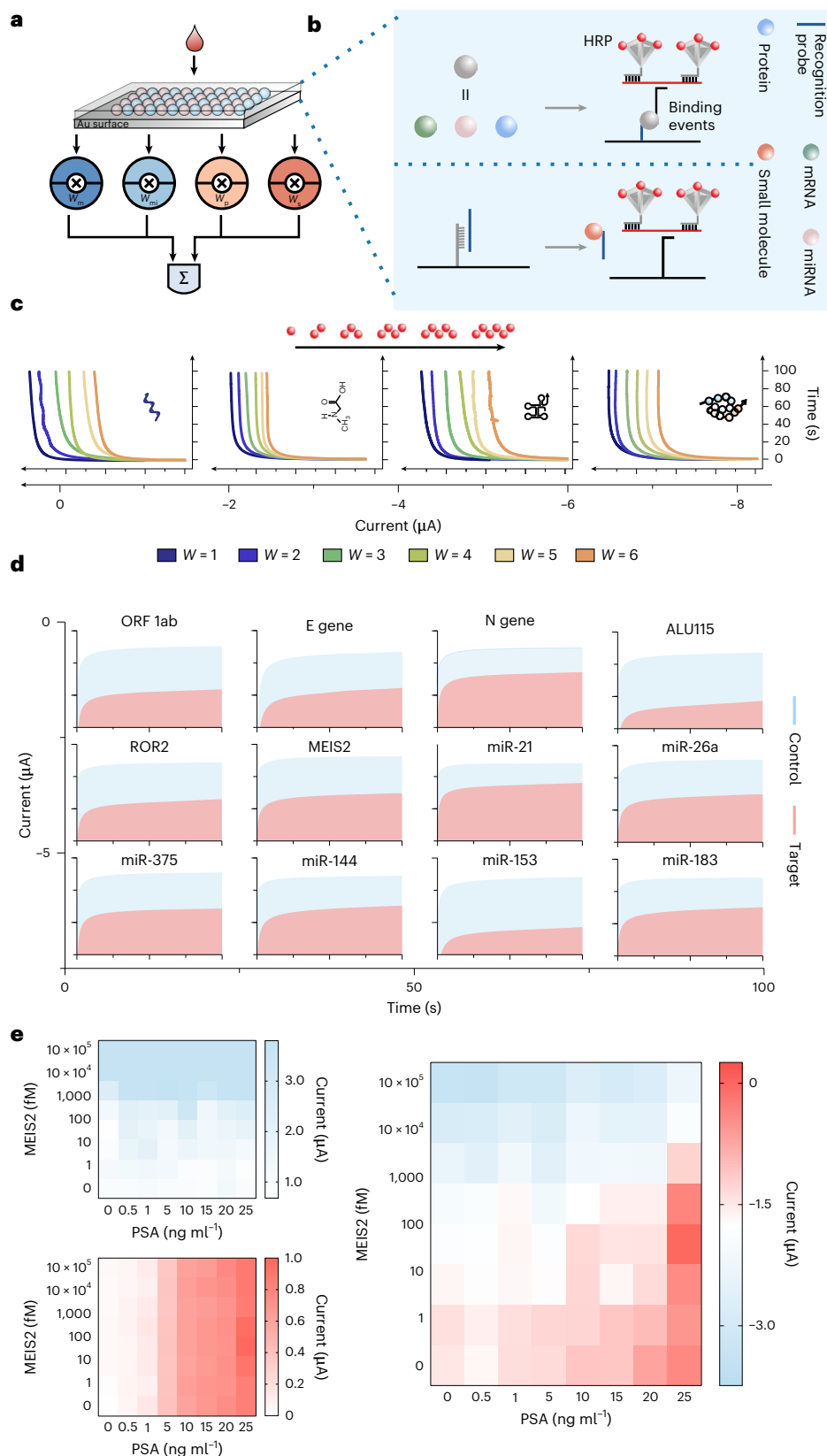


**Fig. 2 | The design of valence-encoded signal reporters using PANs.** **a**, In the multidimensional molecular classifier, a signal reporter was generated to translate each molecule target into a unified output signal and linearly programmed signal gain. We employed a DTF-based PAN reporter to implement the information translation (left).  $W_m$ ,  $W_{mi}$ ,  $W_p$  and  $W_s$  were defined as the weight for different biomarkers. The structure of the PAN reporter was confirmed with an AFM image (right, top). The height–length measurement of the PAN reporter (right, bottom) indicates the two DTFs were successfully coupled through the hybridization of the linker DNA and handle DNA. Scale bar, 10 nm. The AFM experiments were repeated three times, and one representative image is shown. **b**, Fluorescence mapping (left) and intensity (right) of the PAN reporter with different numbers ( $n$ ) of Cy3. The center schematics show the structures. Scale bar, 2  $\mu\text{m}$ . **c**, Correlation between the number and fluorescence intensity of each defined number of Cy3 dyes (shown in the diagram along the top) on the PAN reporter. The images of

fluorescence dots were shown on the top. Scale bar, 500 nm. Error bars indicate standard deviations (mean  $\pm$  s.d.,  $n = 20$ ). **d**, Six steps of photobleaching traces (red arrows) were observed when the number of Cy3 dyes was six on a single PAN reporter (inset schematic). The images of fluorescence dots modified with different number of dyes were shown on the top. Scale bar, 500 nm. **e**, Correlation between the number and fluorescence intensity of each controlled number, with two types of fluorophores, Alexa Fluor 488 and Cy5, on the PAN reporter (shown on the diagrams in the center). Error bars indicate standard deviations (mean  $\pm$  s.d.,  $n = 20$ ). The images of fluorescence dots modified with different number of dyes were shown on the top (Alexa Fluor 488) and bottom (Cy5). Scale bars, 500 nm. **f**, TEM images illustrate the precisely controlled number of AuNPs on the PAN reporter, connected with white dashed lines and also shown in the schematics at the top. Scale bar, 10 nm. **g**, The formation of the PAN reporter anchored with different numbers of HRP and confirmed with AFM. Scale bar, 100 nm.

by designing a valence-encoded PAN reporter to program the unified electrochemical sensing signal for the multidimensional molecules.

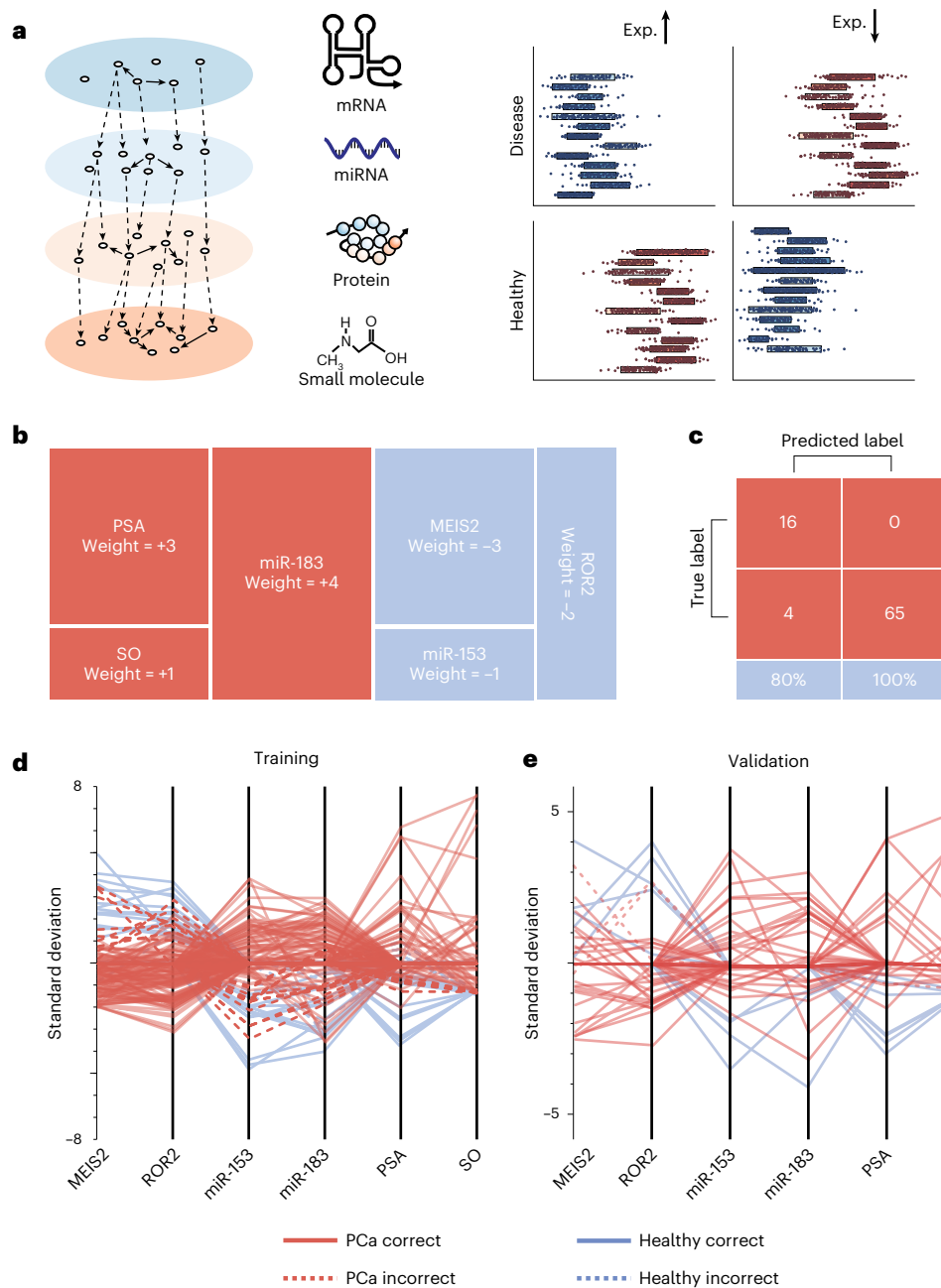
We developed the weighting system for multidimensional molecules with PAN reporters (Fig. 3a). The essential role of the system was to facilitate the binding event between the probe and



**Fig. 3 | A PAN reporter-based weighting system for multidimensional molecules.**

**a**, Scheme of the weighting system for analysing multidimensional molecules on a gold electrode. **b**, Scheme illustration of the recognition and weighting for multidimensional molecules. The system facilitated a recognition binding event between the probe and target molecule, which triggered a weighted electrochemical signal. The top shows mRNA, miRNA or protein, while the bottom shows a small molecule. **c**, Electrochemical signals (shown as current versus time curves) of four

types of target (miRNA, small molecule, mRNA and protein, illustrated in the insets) with precise weight assignment ( $W$ ). The long axis is an isometric zoom of the short axis to show the values more clearly. **d**, Generality of our weighting system for multiple biomarkers. **e**, Validation of two-dimensional molecular classifier. The PSA and MEIS2 gene were used as biomarkers to verify the multidimensional molecular classifier. Left: the individual PSA or MEIS2 gene analysis. Right: the classification based on the linear classifier model for PSA and MEIS2.

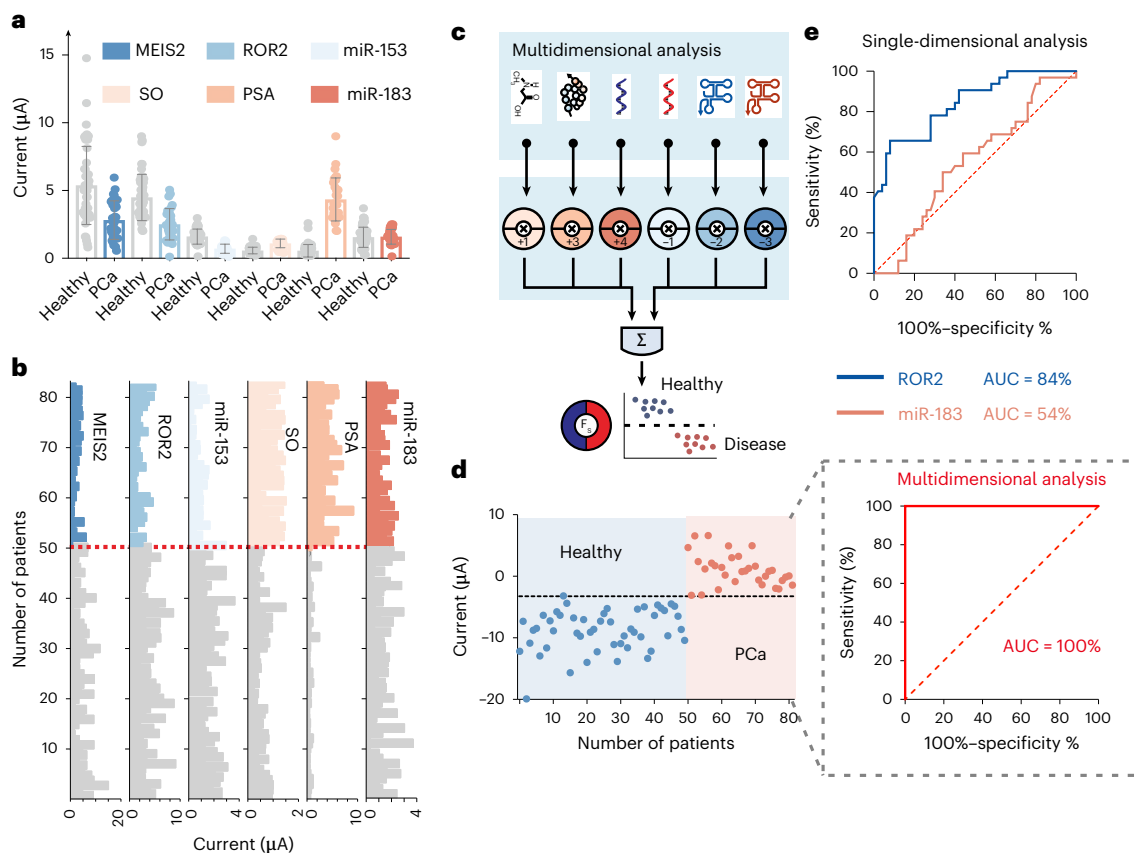


**Fig. 4 | In silico training of linear molecular classifier to discriminate PCa patients and healthy individuals. a**, Left: scheme of computational training using publicly available multidimensional molecular information. Right: the expression (Exp.) information of the multidimensional molecular was used for the computational training. **b**, Selected combinations and their associated

weights as a linear classifier model. **c**, Confusion matrix analysis of 85 samples with a multidimensional classifier. According to the calculation of the analysis data, we obtained a precise score of 100% with the linear classifier model. **d,e**, Standard deviation analysis of training (**d**) and validation sets (**e**) indicate the performance of the multidimensional classifier for the diagnosis of PCa.

target molecule to trigger a weighted electrochemical signal. We used DTFs to pattern recognition probes on the electrode surface according to our previous reports<sup>42</sup>, leading to a uniform biorecognition layer. We employed a sandwich configuration to translate the molecular binding event into the recruitment of the PAN reporter on the electrode for RNAs and proteins. For example, for RNAs (messenger RNA (mRNA) or miRNA), a single-stranded DNA probe was used as the recognition probe, where base-pairing interactions capture the target RNAs on the electrode surface (Supplementary Fig. 14a,b). The PAN reporter then specifically recognized the overhang portion of the probe–target complex and translated the presence of the target RNAs into a weighted electrochemical signal with HRP as the

signalling molecule (Fig. 3b). For proteins, a specific monoclonal antibody was used to capture the target protein on the electrode. Another antibody was then used to form an antibody–protein–antibody sandwich for the target protein (Fig. 3b and Supplementary Fig. 14c). For small molecules, we used an aptamer–DNA duplex as the recognition probe. The small-molecule-to-aptamer binding triggered the release of DNA on the electrode surface, which recruits the PAN reporter via a hybridization between the released DNA and the DNA linker on the PAN reporter (Fig. 3b and Supplementary Fig. 14d). Thus, we designed the weighting system for all the major dimensions of biologically relevant molecules, indicating the generality of our PAN reporter for the weight assignment in multidimensional molecules (Fig. 3a,b).



**Fig. 5 | Multidimensional molecular classifier for PCa diagnosis.** **a**, Mean distribution of six biomarkers in samples from 32 PCa patients and 50 healthy individuals. Error bars indicate standard deviations from  $n = 32$  for PCa group and  $n = 50$  for healthy group (mean  $\pm$  s.d.). The grey boxplots are for healthy individuals and the coloured boxplots are for PCa patients. **b**, Signal translation of six biomarkers' content information with weights in clinical samples from PCa patients and healthy individuals. The grey means the signal translation of six biomarkers' content information of healthy individuals. And the coloured data means the signal translation of six biomarkers' content information of PCa patients. The red dashed line is used to demarcate the data between healthy people and cancer patients. **c**, Scheme of precise PCa diagnosis using

a multidimensional molecular classifier. The  $F_s$  is defined as a symbol for classification of PCa patients and healthy individuals. **d**, Left: diagnosis results of the 82 samples using a multidimensional molecular classifier. The black dashed line is used to demarcate the data between healthy people and cancer patients. Right: ROC analysis for the discrimination between PCa patients and healthy individuals using a multidimensional molecular classifier. **e**, ROC analysis for the discrimination between PCa patients and healthy individuals using a single-dimensional molecule analysis. In **d** and **e**, '100%–specificity%' is the x-axis legend of ROC curve. In **e**, AUC is defined as the area under the ROC curve. The red dashed line is the boundary for calculating the area, which is common to both ROC curves (including blue and orange).

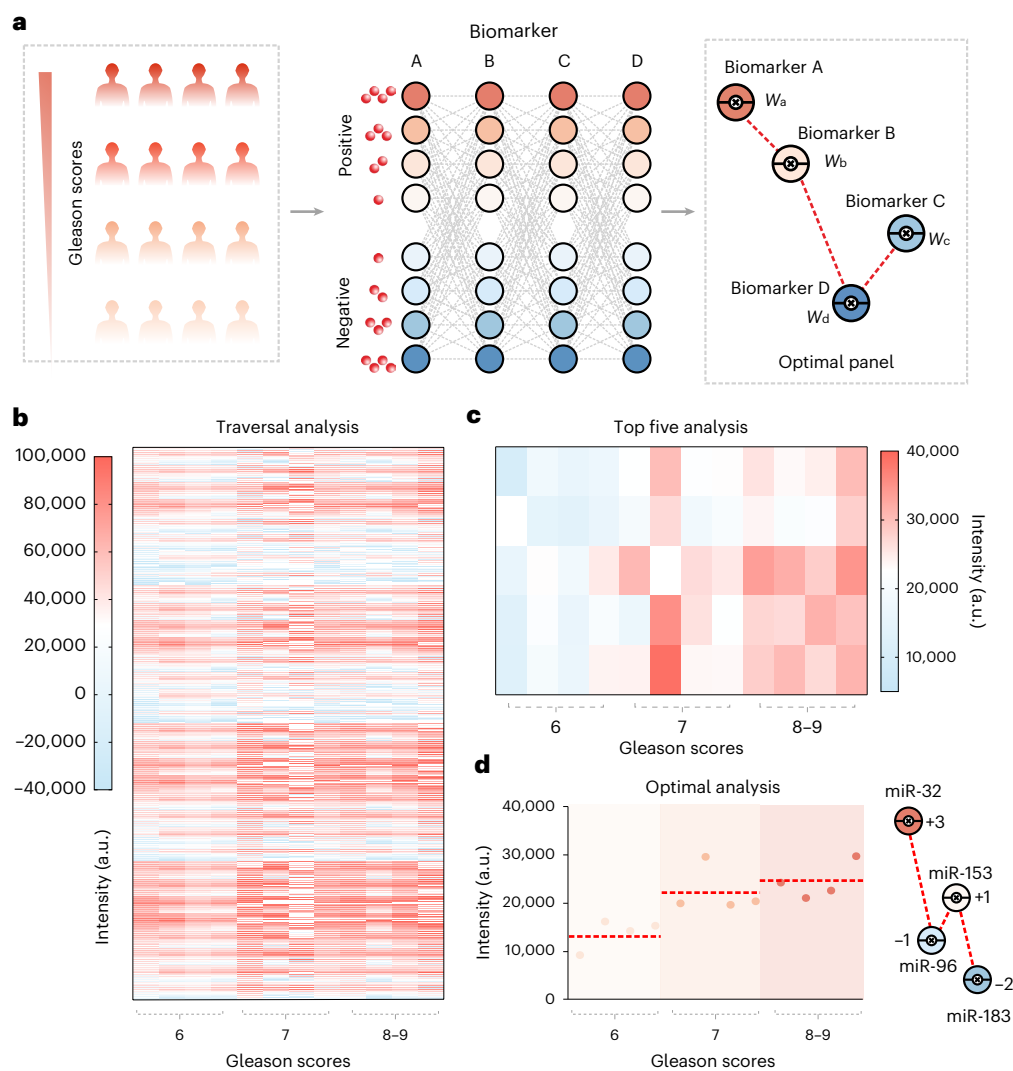
We experimentally implemented this weighting system by designing a weight assignment with one to six HRP's using a PAN reporter for multidimensional molecules (for example, miRNA, mRNA, proteins and small molecules). The electrochemical signal corresponding to the weight assignments was recorded after the addition of the targets until a steady electrochemical signal was achieved. We observed that the signals were linearly proportional to the weights that were realized through controlling the number of HRP on the PAN reporter ( $R^2 > 0.997$ ) for an RNA of 78-nt, a miRNA of 22-nt, an antigen of ~30,000 daltons and a small molecule with 13 atoms. Thus, this system was suitable for assigning an integer-valued weight to different targets (Fig. 3c).

To further demonstrate the generality of the design, we applied the weighting system to 12 additional biomarkers, including COVID-19 biomarkers (including Open Read Framework 1ab (ORF 1ab), envelope gene (E gene) and nucleus gene (N gene))<sup>43</sup>; cancer biomarkers (mRNA ROR2, mRNA MEIS2 and circulating tumor DNA ALU115)<sup>44</sup>; and disease-related miRNAs (miR-21, miR-26a, miR-375, miR-144, miR-153 and miR-183)<sup>45</sup>. We achieved a signal gain of 3.35  $\mu\text{A}$  for ORF 1ab at a concentration of 1,000 copies  $\mu\text{l}^{-1}$  (~1.66 fM), indicating successful signal translation (Fig. 3d). Analogously, we observed remarkable signal gains of 3.75  $\mu\text{A}$  for ALU115 with a concentration of 1 fM.

We further explored the implementation of the weighting system in complicated and biologically relevant matrices, including four types of different diluent of human body fluids (sweat, serum, urine and saliva) and five types of mouse tissue homogenates (heart, kidney, lung, stomach and liver). We observed efficient signal translation and achieved a remarkable signal gain for target molecules, so our weighting system was suitable for complicated biological samples (Supplementary Figs. 15 and 16).

### Validation of the two-dimensional molecular classifier

To experimentally validate a two-dimensional molecular classifier, we employed prostate-specific antigen (PSA), a biomarker in PCa diagnosis, and MEIS2, an mRNA biomarker related to PCa, as the target biomarkers (Fig. 3e)<sup>46</sup>. We assigned a positive weight of +3 to PSA and a negative weight of –3 or –1 to MEIS2. A positive weight represents the positive correlation and a negative weight represents the negative correlation to disease, while their values indicate their importance. We prepared 64 mimetic samples through mixing these two biomarkers with different concentration combinations (Supplementary Table 1) and measured these biomarkers using our PAN reporter (Fig. 3e, left).



**Fig. 6 | Diagnosis panel screening using PAN reporters for PCa. a**, Schematic illustration of diagnosis panel screening using PAN reporters. We tested different biomarkers (miR-32, miR-96, miR-153 and miR-183, labelled A–D) in a series of clinical samples with different scores. Several weights (1, 2, 3 and 4) were assigned to the different markers for information transducing, and all combinations of weights were traversed to finally obtain the optimal diagnostic panel. **b**, The

heatmap of the total traversal analysis for 2,048 ( $=8^4/2$ ). **c**, The heatmap of the top five correlation analysis. **d**, The clustering analysis with the optimal weight combination (miR-32 with weight of +3, miR-96 with weight of -1, miR-153 with weight of +1 and miR-183 with weight of -2). The red dashed line represents the average of the group with different Gleason scores. Significant clustering related to the Gleason score was observed.

After analysing the data via a mathematical function ( $\text{Result} = 3C_{\text{PSA}} - 3C_{\text{MEIS2}}$ ;  $C$ , concentration), we found that the 64 samples were classified into two groups, in agreement with our classifier design (Fig. 3e, right). Moreover, when we changed the weight of MEIS2 from -3 to -1 via a mathematical function ( $\text{Result} = 3C_{\text{PSA}} - 1C_{\text{MEIS2}}$ ), those samples were also classified into two groups but with a different thresholding boundary compared with  $\text{Result} = 3C_{\text{PSA}} - 3C_{\text{MEIS2}}$  (Supplementary Fig. 17).

### In silico training for PCa diagnosis

Next, we attempted to scale up our molecular classifier and employ multidimensional data to classify PCa patients. The workflow is illustrated in Supplementary Fig. 18. To obtain an in silico classifier model for PCa patients' classification, we used publicly available gene and miRNA profiling data from Gene Expression Omnibus, as well as PSA and sarcosine measurement data from previous works<sup>47</sup>, for classifier training (Fig. 4a). We analysed the distributions of the multidimensional molecules between the healthy individuals and PCa patients, and the selected molecules were distinguishable between

these two groups (Supplementary Figs. 19–22). We further investigated the classification models with our classifier, and the robust validation capabilities were confirmed (Supplementary Figs. 23–25).

We integrated the three datasets into a large dataset to evaluate the application for multidimensional molecules and searched the weight combinations by using several logistic regression models with different optimized emphases (Supplementary Fig. 26). We then selected the precision-optimized model to avoid overtreatment (Fig. 4b,c). The optimal weights obtained included miR-153 (weight = -1), miR-183 (weight = +4), ROR2 (weight = -2), MEIS2 (weight = -3), PSA (weight = +3) and sarcosine (SO; weight = +1). With this set of weights, we achieved a recognition sensitivity of 80%, specificity of 100%, F1-score of 97%, receiver operating characteristic (ROC) curve of 97%, precision of 100% and accuracy of 95% for the validation set (Fig. 4c and Supplementary Fig. 26c; the parameters are presented in Supplementary Table 2). Further, we compared the training and validation sets using standard deviation analysis of the multidimensional targets for PCa diagnosis (Fig. 4d,e). The classifier showed



excellent specificity and sensitivity, and it was feasible to achieve molecular implementation.

## PCa diagnosis using multidimensional molecular classifier

We first validated the signal-translating performance of the PAN reporter for six biomarkers of PCa. The electrochemical signal of miRNA exhibited a concentration-dependent linear response with a dynamic range of four orders of magnitude. The detection limit for miRNAs was estimated as 100 fM, allowing for the direct analysis of miRNAs for real samples<sup>48</sup> (Supplementary Fig. 27). Similarly, we achieved the sensitive detection of mRNA, PSA and SO with dynamic ranges of three to five orders of magnitude. The detection limits were down to 1 pM for mRNA, 0.05 ng ml<sup>-1</sup> for PSA and 10 nM for SO (Supplementary Figs. 28–30). The electrochemical signals were also positively correlated to the weights for each biomarker, in agreement with the trends in Fig. 3c. Thus, we successfully established the weight assignment for the six biomarkers (miR-153, miR-183, ROR2, MEIS2, PSA and SO; Supplementary Figs. 31–34).

We then implemented the molecular classifier for the classification of real clinical samples from 32 PCa patients and 50 healthy individuals (the sample information is summarized in Supplementary Table 3). The workflow for clinical sample classification is presented in Supplementary Fig. 35. As shown in Fig. 5a,b, we successfully employed the PAN reporter to convert the six biomarkers into weighted electrochemical signals using the optimized weight sets (Fig. 5c). We realized an accurate classification between PCa patients and healthy individuals with our molecular classifier (*P* value < 0.01; Fig. 5d). The ROC curve indicated a high predictive power with an area under the curve (AUC) of 100% using our molecular classifier (Fig. 5d). We obtained a specificity of 100% and sensitivity of 100%, with the optimal cut-off value. By contrast, we obtained an AUC of only 54% with a single miRNA (miR-183) and an AUC of 84% with a single mRNA (ROR2; Fig. 5e and Supplementary Fig. 36).

## Biomarker panel screening using molecular classifier

Biomarker panels have the potential to distinguish between patients in various disease processes<sup>49</sup> (for example, patients with various Gleason scores for PCa). The rational design of biomarker panels with optimal weighting more accurately reflects the multiple disease processes of cancer. However, the screening of the optimal weighting of each biomarker is challenging. We used serum samples from 12 patients to screen the optimal weighting of the biomarker panel. Samples included four samples with a Gleason score of 6, four samples with a Gleason score of 7 and four samples with a Gleason score of 8 or 9. We used a panel of miRNAs (miR-32, miR-96, miR-153, miR-183) as a model system and assigned weights 1, 2, 3 and 4 to each miRNA using our PAN reporter's weighting system. The weighted signals from the miRNAs with different weight combinations were obtained as 2,048 combinations. The results were used for clustering analysis to screen the optimal weighting set of the biomarker panel (Fig. 6a,b). As shown in Fig. 6c, top five correlation analysis allowed for the classification of three groups according to the Gleason scores, with the optimal weighted result given as  $\text{Result} = 3C_{\text{miR-32}} - C_{\text{miR-96}} + C_{\text{miR-153}} - 2C_{\text{miR-183}}$ , as shown in Fig. 6d, indicating the ability of our molecular classifier to perform the biomarker panel screening.

## Conclusions

In summary, we developed valence-encoded PAN signal reporters by exploiting DNA frameworks to realize multidimensional molecular classification, which resulted in precise PCa diagnosis (an AUC of 100%) with six biomarkers across three-dimensional datatypes (Supplementary Information). Given the ever-increasing amount of molecular information from the gene, RNA, protein and metabolomic profiling

of diseases, our multidimensional molecular classifiers for analysing multidimensional molecular biomarkers sheds light on precision diagnosis and therapy.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41565-023-01348-9>.

## References

- Collins, F. S. & Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).
- Thomasian, N. M., Kamel, I. R. & Bai, H. X. Machine intelligence in non-invasive endocrine cancer diagnostics. *Nat. Rev. Endocrinol.* **18**, 81–95 (2022).
- Vargas, A. J. & Harris, C. C. Biomarker development in the precision medicine era: lung cancer as a case study. *Nat. Rev. Cancer* **16**, 525–537 (2016).
- Nassiri, F. et al. Detection and discrimination of intracranial tumors using plasma cell-free DNA methylomes. *Nat. Med.* **26**, 1044–1047 (2020).
- Krzywinski, M. & Savig, E. Multidimensional data. *Nat. Methods* **10**, 595 (2013).
- Luo, Y. et al. A multidimensional precision medicine approach identifies an autism subtype characterized by dyslipidemia. *Nat. Med.* **26**, 1375–1379 (2020).
- Larance, M. & Lamond, A. I. Multidimensional proteomics for cell biology. *Nat. Rev. Mol. Cell Biol.* **16**, 269–280 (2015).
- Cohen, J. D. et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).
- Berger, B., Peng, J. & Singh, M. Computational solutions for omics data. *Nat. Rev. Genet.* **14**, 333–346 (2013).
- Crichton, D. J. et al. Cancer biomarkers and big data: a planetary science approach. *Cancer Cell* **38**, 757–760 (2020).
- Liang, H. et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat. Med.* **25**, 433–438 (2019).
- Kristensen, V. N. et al. Principles and methods of integrative genomic analyses in cancer. *Nat. Rev. Cancer* **14**, 299–313 (2014).
- Komori, T. The 2021 WHO classification of tumors, 5th edition, central nervous system tumors: the 10 basic principles. *Brain Tumor Pathol.* **39**, 47–50 (2022).
- Blanc, T., El Beheiry, M., Caporal, C., Masson, J. B. & Hajj, B. Genuage: visualize and analyze multidimensional single-molecule point cloud data in virtual reality. *Nat. Methods* **17**, 1100–1102 (2020).
- Adamcova, M. & Šimko, F. Multiplex biomarker approach to cardiovascular diseases. *Acta Pharmacol. Sin.* **39**, 1068–1072 (2018).
- Subramanian, I., Verma, S., Kumar, S., Jere, A. & Anamika, K. Multi-omics data integration, interpretation, and its application. *Bioinf. Biol. Insights* <https://doi.org/10.1177/1177932219899051> (2020).
- Montaner, J. et al. Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke. *Nat. Rev. Neurol.* **16**, 247–264 (2020).
- Tarazona, S., Arzalluz-Luque, A. & Conesa, A. Undisclosed, unmet and neglected challenges in multi-omics studies. *Nat. Comput. Sci.* **1**, 395–402 (2021).
- Tarazona, S. et al. Harmonization of quality metrics and power calculation in multi-omic studies. *Nat. Commun.* **11**, 3092 (2020).

20. Lopez de Maturana, E. et al. Challenges in the integration of omics and non-omics data. *Genes* **10**, 238 (2019).
21. Benenson, Y., Gil, B., Ben-Dor, U., Adar, R. & Shapiro, E. An autonomous molecular computer for logical control of gene expression. *Nature* **429**, 423–429 (2004).
22. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-free nucleic acid logic circuits. *Science* **314**, 1585–1588 (2006).
23. Lopez, R., Wang, R. & Seelig, G. A molecular multi-gene classifier for disease diagnostics. *Nat. Chem.* **10**, 746–754 (2018).
24. Zhang, C. et al. Cancer diagnosis with DNA molecular computation. *Nat. Nanotechnol.* **15**, 709–715 (2020).
25. Yao, G. et al. Meta-DNA structures. *Nat. Chem.* **12**, 1067–1075 (2020).
26. Yao, G. et al. Programming nanoparticle valence bonds with single-stranded DNA encoders. *Nat. Mater.* **19**, 781–788 (2020).
27. Li, J. et al. Encoding quantized fluorescence states with fractal DNA frameworks. *Nat. Commun.* **11**, 2185 (2020).
28. Wiraja, C. et al. Framework nucleic acids as programmable carrier for transdermal drug delivery. *Nat. Commun.* **10**, 1147 (2019).
29. Zhang, T. et al. Design, fabrication and applications of tetrahedral DNA nanostructure-based multifunctional complexes in drug delivery and biomedical treatment. *Nat. Protoc.* **15**, 2728–2757 (2020).
30. Song, P. et al. Programming bulk enzyme heterojunctions for biosensor development with tetrahedral DNA framework. *Nat. Commun.* **11**, 838 (2020).
31. Lin, M. et al. Programmable engineering of a biosensing interface with tetrahedral DNA nanostructures for ultrasensitive DNA detection. *Angew. Chem. Int. Ed.* **54**, 2151–2155 (2015).
32. Woehrstein, J. B. et al. 100-nm metafluorophores with digitally tunable optical properties self-assembled from DNA. *Sci. Adv.* **3**, e1602128 (2017).
33. Ulbrich, M. H. & Isacoff, E. Y. Subunit counting in membrane-bound proteins. *Nat. Methods* **4**, 319–321 (2007).
34. Hearty, S., Leonard, P. & O’Kennedy, R. Barcodes check out prostate cancer. *Nat. Nanotechnol.* **5**, 9–10 (2010).
35. Hill, H. D. & Mirkin, C. A. The bio-barcode assay for the detection of protein and nucleic acid targets using DTT-induced ligand exchange. *Nat. Protoc.* **1**, 324–336 (2006).
36. Nam, J.-M., Thaxton, C. S. & Mirkin, C. A. Nanoparticle-based bio-bar codes for the ultrasensitive detection of proteins. *Science* **301**, 1884–1886 (2003).
37. Zebda, A. et al. Mediatorless high-power glucose biofuel cells based on compressed carbon nanotube-enzyme electrodes. *Nat. Commun.* **2**, 370 (2011).
38. de Jong, O. G. et al. A CRISPR-Cas9-based reporter system for single-cell detection of extracellular vesicle-mediated functional transfer of RNA. *Nat. Commun.* **11**, 1113 (2020).
39. Zhao, Z. et al. Nanocaged enzymes with enhanced catalytic activity and increased stability against protease digestion. *Nat. Commun.* **7**, 10619 (2016).
40. He, L. et al. Transducing complex biomolecular interactions by temperature-output artificial DNA signaling networks. *J. Am. Chem. Soc.* **142**, 14234–14239 (2020).
41. Li, H., Brouwer, C. R. & Luo, W. A universal deep neural network for in-depth cleaning of single-cell RNA-Seq data. *Nat. Commun.* **13**, 1901 (2022).
42. Lin, M. et al. Electrochemical detection of nucleic acids, proteins, small molecules and cells using a DNA-nanostructure-based universal biosensing platform. *Nat. Protoc.* **11**, 1244–1263 (2016).
43. Gorog, D. A. et al. Current and novel biomarkers of thrombotic risk in COVID-19: a Consensus Statement from the International COVID-19 Thrombosis Biomarkers Colloquium. *Nat. Rev. Cardiol.* **19**, 475–495 (2022).
44. Schwarzenbach, H., Hoon, D. S. B. & Pantel, K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat. Rev. Cancer* **11**, 426–437 (2011).
45. Xiao, B. et al. Plasma microRNA panel is a novel biomarker for focal segmental glomerulosclerosis and associated with podocyte apoptosis. *Cell Death Dis.* **9**, 533 (2018).
46. Bhanvadia, R. R. et al. MEIS1 and MEIS2 expression and prostate cancer progression: a role for HOXB13 binding partners in metastatic disease. *Clin. Cancer Res.* **24**, 3668–3680 (2018).
47. Kumar, D., Gupta, A., Mandhani, A. & Sankhwar, S. N. Metabolomics-derived prostate cancer biomarkers: fact or fiction? *J. Proteome Res.* **14**, 1455–1464 (2015).
48. Rajakumar, T. et al. A blood-based miRNA signature with prognostic value for overall survival in advanced stage non-small cell lung cancer treated with immunotherapy. *npj Precis. Oncol.* **6**, 19 (2022).
49. Nassiri, F. et al. A clinically applicable integrative molecular classification of meningiomas. *Nature* **597**, 119–125 (2021).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

## Methods

The study was approved by the Ethics Committee at Renji Hospital, School of Medicine, Shanghai Jiao Tong University. All methods were performed in accordance with these approved guidelines.

### Workflow

The workflow for the classification of real clinical samples is presented in Supplementary Fig. 35. Recognition probes for each target were first modified on the electrode. The read-out of the electrochemical signal of the multidimensional target was performed by weighting the capture of the recognition probe with the PAN reporter. The final classification of clinical samples was achieved by a diagnostic function. The cost for a patient is only US\$6.3 (Supplementary Table 4)

### Data availability and simulations

**The miRNA data.** The miRNA data for PCa patient analysis was from [GPL8227](#) (Agilent-019118 Human miRNA Microarray 2.0 G4470B; miRNA ID version). This dataset included 113 prostate patients and 28 healthy individuals. For every single person, there were 881 miRNA described, such as miR-183. According to the *t*-test result, 171 miRNA described were selected with a high significant difference between patients and healthy ones. Tree-based feature selection from sklearn (the function library of tree-based feature) was used to select the top related miRNAs (miR-183 and miR-153).

**The mRNA data.** The mRNA data for the PCa patient analysis was from [GPL10264](#) (Human Exon 1.0 ST Array; CDF file, HuEx\_1\_0\_st\_v2\_main\_A20071112\_EP.cdf) and recorded the Affymetrix gene expression of 150 PCa patients and 29 healthy individuals. The descriptors were dimensionality-reduced from 43,419 to 6,148 by a *t*-test, and remained two items ([NM\\_170675](#) (MEIS2) and [NM\\_004560](#) (ROR2)) by tree-based feature selection.

**Clinical dataset.** The clinical dataset was from the literature<sup>47</sup>. It contained 70 PCa patients and 32 healthy individuals. The most important features were the PSA and SO after being treated similarly, as mentioned earlier.

From gene expression data ([GPL10264](#)), we obtained 150 PCa patients and 29 healthy individuals and employed a tree-based feature selection method to screen for the two most related aberrant expressed genes. We selected ROR2 and MEIS2 from 43,419 items (dataset 1). Similarly, we analysed the miRNA profiling ([GPL8227](#)) with 113 PCa patients and 28 healthy individuals, and selected two important miRNAs (miR-153 and miR-183) by feature selection (dataset 2). In addition, PSA and SO were selected as protein and small molecule biomarkers, respectively, from the clinical data (70 PCa patients and 32 healthy individuals; dataset 3).

The missing data were replaced by the average of each descriptor among the same group. In all, the combined dataset had 422 samples; among these were 333 PCa patients and 89 healthy individuals. Each individual was described by six selected descriptors.

**Software.** Tree-based feature selection from sklearn was used for feature selection, and the logistic regression module from sklearn was applied to classify the two-category model. To find the integer weights of each descriptor, an exhaustive search method hunted through the whole integer parameter space from  $-4$  to  $4$ . The accuracy, precision, recall and F1-score of every model were calculated and recorded. The classification analysis was implemented by the Classification learning app in MATLAB (R2020b).

**Benchmarking.** The concentration and weight correlation of the molecular classifier were calibrated with the standard samples for different targets before diagnosis applications. The concentration of the standard was quantified by the UV absorbance at 260 nm by

the Shanghai Institute of Measurement and Testing Technology. (The certificate of the standard samples is provided in Supplementary Information and Supplementary Table 5.)

**Synthesis and purification of DTF-based PAN reporter.** All DNA strands were mixed in TM buffer to synthesize the DTF structures (the proportions are illustrated in Supplementary Tables 6–19). The mixture was heated to 95 °C for 15 min, and cooled to 4 °C for at least 20 min by using a PTC-200 thermal cycler DNA engine (MJ Research, USA). We purified the synthesized DTF structures according to the method reported in the literature<sup>50</sup>. Our PAN reporter is simple to prepare and can be successfully synthesized even by undergraduate students without any knowledge in this field (Supplementary Fig. 37 and Supplementary Table 20). Moreover, we were able to achieve millilitre-level (7.5 ml) synthesis using a metal blocker for the bulk preparation of PAN (Supplementary Fig. 38). The PAN was tested and characterized through PAGE after being stored in buffer solution or serum for 1, 3, 7 and 15 days. As shown in Supplementary Fig. 39, PAN remained stable in the buffer solution even after 15 days and stable in the serum for at least a day. Thus, PAN reporters can be prepared in bulk and preserved for long periods, with potential for practical clinical applications (Supplementary Table 21). The stability of the DTF at the interface was also examined, to adapt it to interfacial applications. After being modified with both Cy3 and Cy5 on the same edge of the DTF, we found that the DTF can be stable at the interface for up to five days, as determined with fluorescence resonance energy transfer and dual fluorescence co-localization (Supplementary Fig. 40).

**Weighting system for miRNA information translation.** The purified DTF for short-strand RNA interface capture (1  $\mu$ M, 6  $\mu$ l) was incubated on the cleaned electrode overnight at room temperature. The electrodes were then passivated by methylcyclohexane (2 mM), polyethylene glycol 2000 (2 mM) and 10% bovine serum albumin. After that, the electrodes were washed with phosphate-buffered saline (PBS) and dried with nitrogen. Next, the samples were dropped on the electrode surface and incubated for 2 h at 25 °C. The PAN reporter (50 nM) was added on the electrode surface and incubated for 2 h at 25 °C. Finally, 4  $\mu$ l of avidin-HRP was added on the electrode surface for 15 min to bind to the biotin in the molecular reporter. After being washed thoroughly, the electrodes were immersed in TMB solution buffer for electrochemical measurements.

**Weighting system for mRNA information translation.** The purified DTF dimer for long-strand RNA interface capture (1  $\mu$ M, 6  $\mu$ l) was incubated on the cleaned electrode overnight at room temperature. The processes of sealing and content information transformation for mRNA were the same as those of miRNA.

**Weighting system for PSA information translation.** The purified DTF for PSA interface capture (1  $\mu$ M, 6  $\mu$ l) was incubated on the cleaned electrode overnight at room temperature. After being washing twice with PBST buffer and once with PBS buffer, anti-PSA monoclonal antibody (coating; monoclonal antibody is a highly uniform antibody and only specific to a specific epitope) (L1; 100  $\mu$ M, 6  $\mu$ l) was dropped on the chip electrode and incubated at room temperature for 2 h to form the fixed probe, and then the electrode was washed twice with PBST buffer and once with PBS buffer. Subsequently, a series of PSA samples in PBS buffer (6  $\mu$ l) at variable concentrations were dropped on the chip electrodes and incubated at 37 °C for 1 h. After washing, anti-PSA monoclonal antibody (labelling) (L2; 100  $\mu$ M, 6  $\mu$ l) was dropped on the chip electrode and incubated at room temperature for 2 h to form the capture probe. The chip electrode was washed twice with PBST buffer and once with PBS buffer. After that, the PAN reporter (100  $\mu$ M, 6  $\mu$ l) was dropped on the chip electrode and incubated at room temperature for 2 h to form the weighting probe. Finally, excess avidin-HRP was

dropped onto the electrode and incubated at room temperature for 15 min. After washing twice with PBST buffer and once with PBS buffer, electrochemical testing was performed immediately. The sequences of L1 and L2 were shown in Supplementary Tables.

**Weighting system for SO information translation.** The purified DTF for SO interface capture (1  $\mu\text{M}$ , 6  $\mu\text{l}$ ) was incubated on the cleaned electrode overnight at room temperature. The electrodes were passivated with 0.13% methylcyclohexane, 20  $\text{mg ml}^{-1}$  polyethylene glycol and 1% casein in sequence for 1 h. The diluted sample solution was incubated on the 16-channel electrodes for 2 h at room temperature (6  $\mu\text{l}$ ). After 2 h, the 16-channel electrodes were washed with the washing buffer. The PAN reporter was incubated on the electrodes for 2 h and then washed with PBS buffer. Finally, excess avidin-HRP was dropped onto the electrode and incubated at room temperature for 15 min. After washing twice with PBST buffer and once with PBS buffer, electrochemical testing was performed immediately.

**Electrochemical measurements.** All electrochemical measurements were done on a Model 1040C (CH Instruments). The working gold 16-channel electrode, the auxiliary electrode and the reference electrode, integrated in the chip, were used. Cyclic voltammetry was carried out at a scan rate of 100  $\text{mV s}^{-1}$ . The current was recorded at  $-100\text{ mV}$  after the steady state of the HRP catalytic reaction was reached<sup>30</sup>.

**Biomarker panel screening using molecular classifier.** In the experiments of the biomarker panel screening, we used the fluorescent signal chip system. The 500 nM miRNA capture probes (Supplementary Tables 16 and 17) were printed by microarray robot (Nano-Plotter NP2.1). After incubating overnight, the chip was then blocked by 2 mM polyethylene glycol 2000 for 45 min and 2% bovine serum albumin for 1 h. The diluent for clinical samples was added on the chip and incubated for 2 h at 25 °C. The PAN reporter with different weights (50 nM) was then added on the chip and incubated for 2 h at 25 °C. Then the chip was imaged by a GenePix 4100A microarray scanner. We obtained signals of four targets with four weights for 12 patients. By adding and subtracting combinations of them, 2,048 ( $8^4/2$ ) diagnostic formulas were obtained. Finally, the optimal formulas were filtered by cluster analysis.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request. Furthermore, the miRNA, mRNA, PSA and SO data used in

this study are available in ref.<sup>47</sup> and the National Center for Biotechnology Information database, <https://www.ncbi.nlm.nih.gov/genome>.

### References

50. Li, F. et al. Ultrafast DNA sensors with DNA framework-bridged hybridization reactions. *J. Am. Chem. Soc.* **142**, 9975–9981 (2020).

### Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (T2188102, 22025404, 22001168); National Key R&D Program of China (2021YFF1200300); China National Postdoctoral Program for Innovative Talents (BX2021190) by the China Postdoctoral Science Foundation; Innovative Research Team of High-Level Local Universities in Shanghai (SHSMU-ZLCX20212602); 2022 Shanghai ‘Science and Technology Innovation Action Plan’ Fundamental Research Project (22JC1401202); Shanghai Jiao Tong University Scientific and Technological Innovation Funds (21X010202096) and Shanghai Municipal Health Commission (2022JC027).

### Author contributions

X.Z., C.F. and F.Y. conceived the study. F.Y., H.Z. and S.L. performed the experiments. F.Y. performed the TIRFM imaging and nucleic acid information translation. H.Z. performed the TEM imaging and SO information translation. S.L. performed the AFM imaging and PSA information translation. J. Shen performed the target screen and data training. B.D. and W.X. provided samples and analysed the clinical data. F.Y., H.Z. and S.L. performed the clinical sample detection. F.Y., J. Shi, M.L., X.M., F.L. and J.L. carried out the assays and analysed the results. X.Z. and C.F. directed the research. X.Z., C.F. and F.Y. wrote the paper. X.Z., C.F. and X.Y. supervised the project. All authors read the paper and provided comments.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41565-023-01348-9>.

**Correspondence and requests for materials** should be addressed to Xiaolei Zuo.

**Peer review information** *Nature Nanotechnology* thanks Hao Yan and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a                                 | Confirmed  |
|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated  |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

- |                 |   |
|-----------------|---|
| Data collection | The miRNA-seq, mRNA-seq, PSA and sarcosine data used in this study are available on ref. 47 and NCBI website: <a href="https://www.ncbi.nlm.nih.gov/genome">https://www.ncbi.nlm.nih.gov/genome</a> . Data collected and analyzed using CHI 1040C commercial software for all electrochemical measurements.   |
| Data analysis   | Tree-based feature selection from sklearn was used for feature selection and the LogisticRegression module from sklearn was applied to classify two categories model. The classification analysis was implemented by the Classification learning app in MATLAB (R2020b). Prism 8 and Origin 2019 for plotting data as figures included in manuscript. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request. Furthermore, the miRNA-seq, mRNA-seq, PSA and sarcosine data used in this study are available on NCBI database and ref 47.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	Since our study was focusing on prostate cancer (PCa), the samples were all from males.
Population characteristics	Study participants include PCa patients and healthy individuals (including 32 PCa patients and 50 healthy individuals). The PCa samples were from the patients of the first diagnosis without any treatment. Healthy samples were from individuals with no issue of prostate reported in physical examination screening reports. The age, gender and cancer stage of all samples were not limited. Detailed clinical sample information is provided in Supplementary Table 3.
Recruitment	All subjects were enrolled with Institutional Review Board-approved protocols and all serum samples were collected with informed consent. PCa patients were recruited based on clinical diagnosis of PCa with confirmed pathological biopsy results. Healthy individuals were recruited based on their health screening reports with no issue of the prostate. The serum samples were collected by the Renji Hospital, School of Medicine, Shanghai Jiao Tong University.
Ethics oversight	All samples analyzed in this study were collected with informed consent from subjects and approved protocols that complied with all relevant ethical regulations at Renji Hospital, School of Medicine, Shanghai Jiao Tong University.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation during experiment design. For in silico training, sample size was determined by the number of cases available in the NCBI database and mined. For the evaluation of diagnostic accuracy of the multidimensional molecular classifier, we used 82 clinical serum samples from 50 healthy and 32 PCa individuals. Herein, the sample size was largely determined by sample availability.
Data exclusions	None clinical sample was excluded.
Replication	We replicated PAGE, AFM and TIRFM characterization at least 3 times and electrochemical measurements at least 2 times across different conditions. Any case was not found where the data was not able to be replicated.
Randomization	PCa patients and healthy controls were enrolled in the study based on either confirmed diagnosis of prostate cancer or health screening reports with no issue of the prostate. Therefore, randomization of individuals in different groups is not applicable
Blinding	No blinded experiments. For In silico training, we obtained all the available data from NCBI database and ref.47. For the clinical samples, we collected known types of samples from Renji hospital according to the experimental needs to verify the accuracy of multidimensional molecular classifier.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Included in the study   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Antibodies                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology          |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                          |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern           |

## Methods

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Included in the study                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

## Antibodies

Antibodies used

Anti-PSA McAb (coating) (Shanghai Linc-Bio Science Co. LTD, catalog: L1C00401, clone numbers:PS001 ); Anti-PSA McAb (labeling) (Shanghai Linc-Bio Science Co. LTD, catalog: L1C00402, clone numbers:PS002)

Validation

Anti-PSA McAb (coating) Reactivity: Mouse; Hybridization of P3-X63-Ag8.653 .myeloma cells with spleen cells from Balb/c mice. Application: PSA capture. Anti-PSA McAb (labeling) Reactivity: Mouse; Hybridization of P3-X63-Ag8.653 .myeloma cells with spleen cells from Balb/c mice. Application: PSA detection.

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

BALA/C-nu mice (3-4 weeks age, from Jiesijie Laboratory Animal Center) was used in the study. The mice were housed five per cage in a specific pathogen-free environment (18–22°C, 50–70% relative humidity, 12h light–dark cycle) and access to water and food freely.

Wild animals

No wild animals were used in the study.

Reporting on sex

The animal was used only to provide tissue homogenates. Sex had no effect on the experiment

Field-collected samples

No field collected samples were used in the study.

Ethics oversight

All the experiments were complied with international guidelines and animal experiments were performed under the guidelines evaluated and approved by the ethics committee of the Institutional Animal Care and Use Committee of Shanghai Jiao Tong University.

Note that full information on the approval of the study protocol must also be provided in the manuscript.