Nanopore sequencing of DNA-barcoded probes for highly multiplexed detection of microRNA, proteins and small biomarkers

Caroline Koch[1,4], Benedict Reilly-O'Donnell [1,2,4], Richard Gutierrez[3], Carla Lucarelli[2], Fu Siong Ng [2], Julia Gorelik [2], Aleksandar P. Ivanov [1] ✉ & Joshua B. Edel [1] ✉

There is an unmet need to develop low-cost, rapid and highly multiplexed diagnostic technology platforms for quantitatively detecting blood biomarkers to advance clinical diagnostics beyond the single biomarker model. Here we perform nanopore sequencing of DNA-barcoded molecular probes engineered to recognize a panel of analytes. This allows for highly multiplexed and simultaneous quantitative detection of at least 40 targets, such as microRNAs, proteins and neurotransmitters, on the basis of the translocation dynamics of each probe as it passes through a nanopore. Our workflow is built around a commercially available MinION sequencing device, offering a one-hour turnaround time from sample preparation to results. We also demonstrate that the strategy can directly detect cardiovascular disease-associated microRNA from human serum without extraction or amplification. Due to the modularity of barcoded probes, the number and type of targets detected can be significantly expanded.

The detection of blood serum biomarkers is one of the most common methods for diagnosis, prognosis, predicting future disease and monitoring response to treatment. Biochemicals have been used for over 60 years to aid diagnosis of a wide range of conditions, including cancer[1], pregnancy[2] and cardiac disease[3]. Biomarker tests traditionally rely on the detection of proteins to indicate a condition. However, these tests often lack the specificity to provide clinically useful detail on the pathology[4,5]. To increase the specificity in detecting diseases, multiplexed biomarker assays have been developed. Such strategies have been attempted in several areas, including Alzheimer's disease[6,7], amyotrophic lateral sclerosis[8], cardiovascular disease[9–11], chronic obstructive pulmonary disease[12], infection[13] and cancers[14–17]. Many of the methodologies employed rely upon antibody recognition of an epitope and an associated optical readout for each analyte, for example, ELISA, which can be sensitive to pM levels of analyte. Commercially available systems can detect up to 80 proteins simultaneously[18,19]. Other methods use

modified DNA-based aptamers, which allow the detection of multiple proteins in a single sample with a sensitivity down to 125 fM (refs. 20,21). These techniques offer promising tools for high-throughput protein biomarker detection but are unable to observe multiple molecular species simultaneously.

Clinical diagnostics are now moving beyond protein biomarkers and genetic testing. One example is microRNAs (miRNAs), short non-coding RNAs that regulate gene expression[22,23]. Alterations in miRNA expression have been identified in a wide range of clinical areas, such as cardiology[24], hepatology[25], nephrology[26], neurology[27], oncology[28] and vascular disease[29]. These molecules are detected in the blood primarily through reverse transcription–quantitative PCR (RT–qPCR)[30], which is sensitive down to fM. However, this method requires multiple steps and relies on signal amplification, which may introduce bias in the measurement. Furthermore, the rapid degradation of miRNAs provides a particular problem when considering their

[1]Department of Chemistry, Molecular Science Research Hub, Imperial College London, London, UK. [2]National Heart and Lung Institute, ICTEM, Imperial College London, London, UK. [3]Oxford Nanopore Technologies PLC, Oxford Science Park, Oxford, UK. [4]These authors contributed equally: Caroline Koch, Benedict Reilly-O'Donnell. ✉e-mail: alex.ivanov@imperial.ac.uk; joshua.edel@imperial.ac.uk
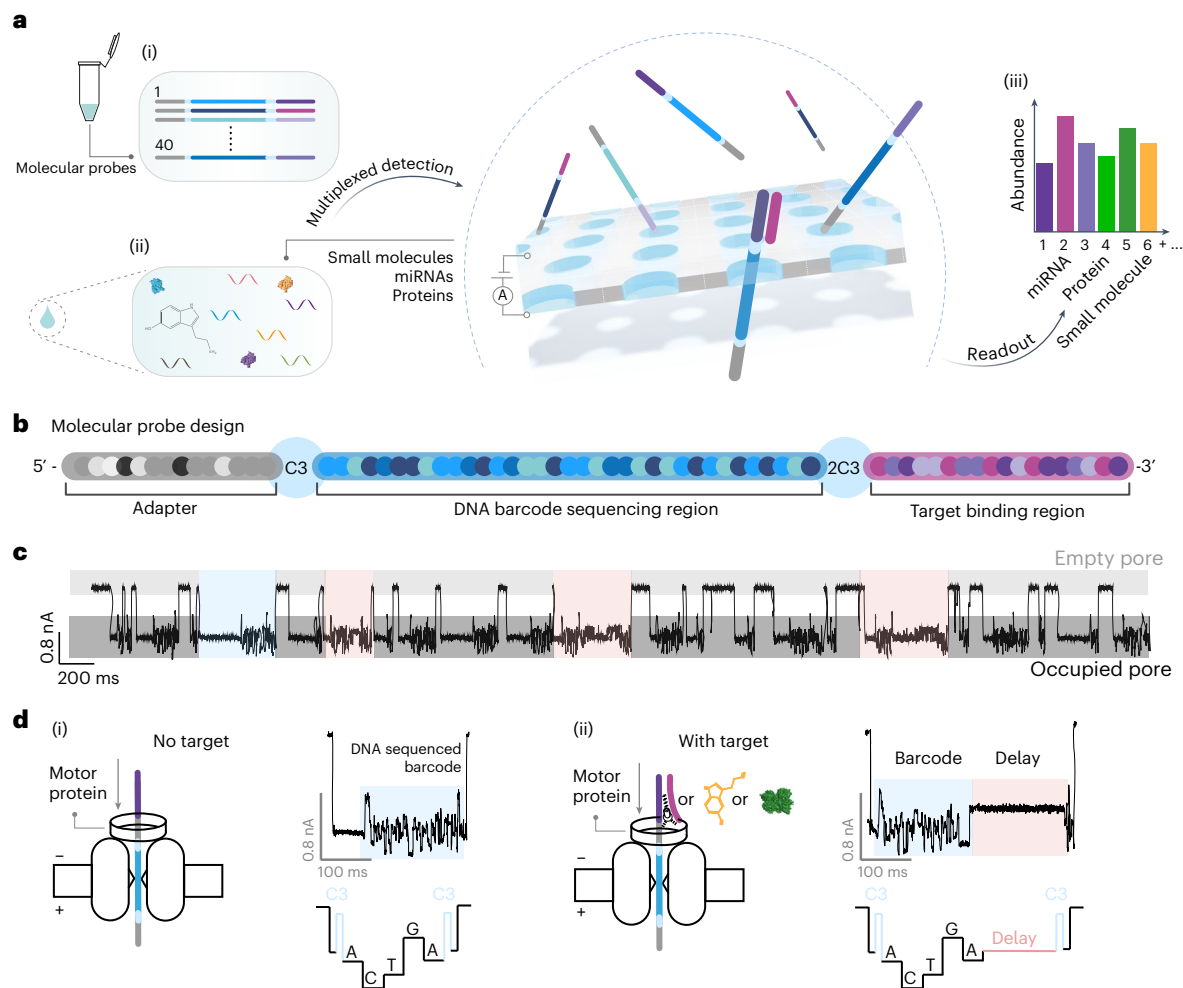
Fig. 1 | Demultiplexing of 40 miRNAs, proteins and small molecules using barcoded probes and nanopore sequencing. a, Workflow for detection of miRNAs, proteins and small molecules. Barcoded probes (i) were incubated with synthetic targets or serum samples from healthy participants (ii). Nanopore sequencing was performed to classify the barcode, and the presence of target analytes (iii) was determined. **b**, Barcoded probe design. The probe consists of a Y-adapter containing the motor protein; the barcode region, containing 35 bases; and the target binding region at the 5′ end of the probe, which includes either a complementary sequence to a miRNA or an aptamer designed to specifically bind a protein or small molecule. **c**, Characteristic current traces. Example events are highlighted for barcoded probe only and barcoded probe events with bound target analytes. **d**, Base calling and analysis of events from barcoded probes result in current traces without delay (i). Base calling and analysis of events from barcoded probes with analyte bound results in current traces with a delay (ii).

use in a clinical setting. A rapid miRNA profiling platform would offer the potential to capture short-lived events and perform frequent longitudinal testing.

There is, therefore, a great need to develop technologies that can perform highly multiplexed detection of various analyte classes, including nucleic acids, proteins and small molecules. Single-molecule nanopore sensing offers the ideal platform for performing this task. Nanopores have previously been shown to enable efficient detection of DNA, RNA, proteins and other molecules[31–35], albeit not in a highly multiplexed configuration. Analyte detection by nanopores depends on measuring current fluctuations as charged molecules are electrophoretically driven through a nanoscale aperture. The translocation of a molecule through a nanopore causes a change in the ionic current, which is dependent on the molecule's charge, size and conformation[36]. However, the method generally lacks selectivity. Strategies to address these limitations include chemical modifications of the pore[37–40], the use of molecular carriers[41] and the use of electro-optical methods[42,43]. These methods are excellent for the detection of molecules at very low concentrations without amplification; however, their throughput and multiplexing ability are limited.

Biological nanopores are advantageous over solid-state devices as they are highly reproducible and can be engineered for specific functionality[44]. In particular, biological nanopores are useful for DNA/RNA sequencing, as shown by Oxford Nanopore Technologies (ONT). ONT has commercialized sequencing devices that use biological nanopore arrays to allow for high-throughput, simultaneous DNA/RNA reads[45] of fragments ≥20 base pairs[46].

This study showcases a multiplexed analyte detection strategy, combining nanopore sequencing with DNA-barcoded molecular probes. The platform allows accurate demultiplexing of events, with simultaneous quantitative detection of at least 40 molecules that can consist of miRNAs, proteins and small molecules, such as neurotransmitters. The presence of each analyte is determined by the translocation dynamics of each probe as it passes through a nanopore. In this study, we selected 40 miRNAs and proteins implicated in cardiac disease. The method established is easily adaptable and scalable, meaning the number of detected biomarkers can be extended to cover multiple diseases. The assay requires a sample volume of less than 30 μl, does not require sample labelling or amplification and costs less than US$100. Furthermore, the technology bears the potential for pooled
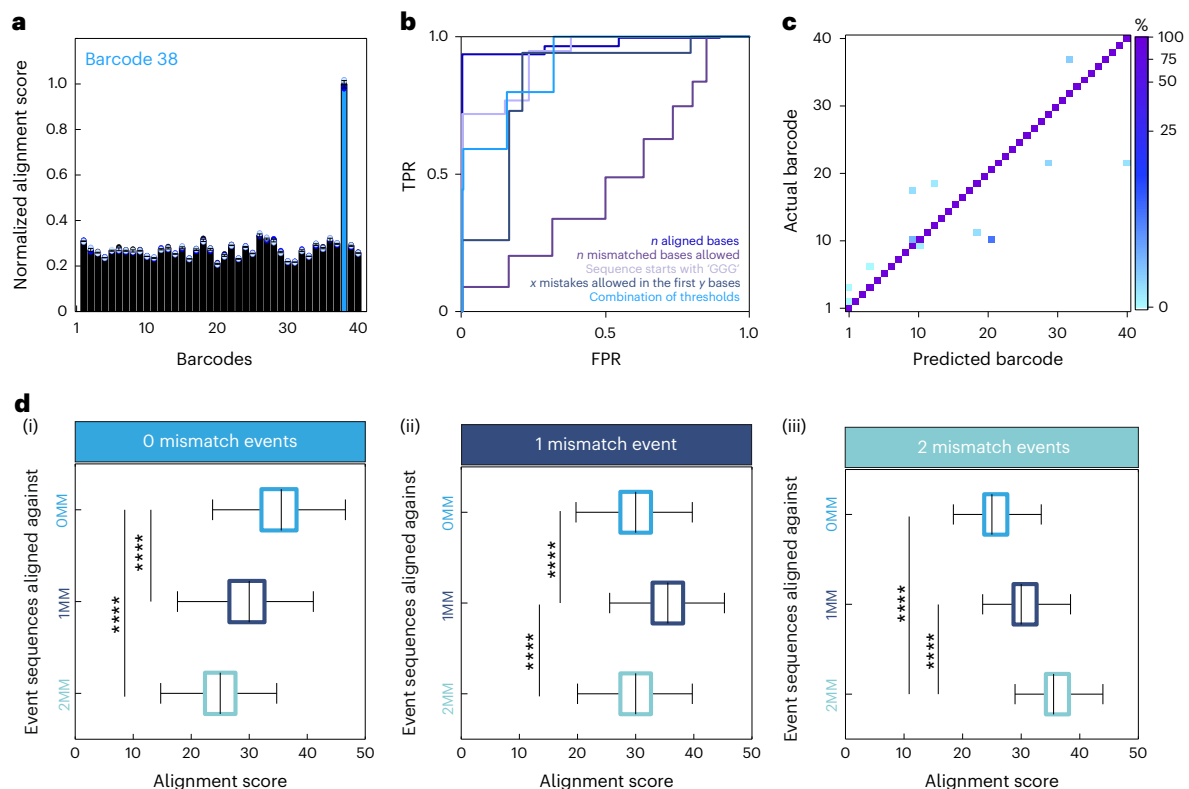
Fig. 2 | Highly accurate detection of nucleic acid barcodes. a, Alignment scores of a barcoded probe 38 experiment. Sequenced events were aligned against a barcode library containing all 40 barcoded probe sequences. The alignment score for barcoded probe 38 was significantly higher (analysis of variance (ANOVA), $F(39, 17,360) = [3,441.41]$, ****$P = 0$, $n = 3$, $n_{\text{total events}} = 454$) than all other alignment scores observed. Data presented as mean ± s.d. **b**, Receiver operating characteristic curve showing the FPR and true-positive rate (TPR) of various alignment thresholds including n aligned bases; n mismatched bases allowed; sequence starts with x; x mismatches allowed in first y bases and a combination ($n = 3$). **c**, Confusion matrix of alignment accuracy of single barcoded probe experiments. Accuracy was >95% for all 40 barcoded probes tested ($n = 3$, $n_{\text{total events}} = 15,489$). **d**, Alignment score of barcoded probe sequences with 1 and 2 mismatched bases. A multiplexed experiment containing three barcoded probes ('0MM', GGGTGCACGAGTGCGTGT; '1MM' GGGT**A**CACGAGTGCGTGT; '2MM' GGGT**A**CA**T**GAGTGCGTGT) shows a significant difference between alignment scores. 0MM events had the highest alignment score for the 0MM sequence (ANOVA, $F(2, 67,524) = [13,657.25]$, ****$P = 0$, $n = 4$, $n_{\text{total events}} = 22,509$) (i). 1MM events had the highest alignment score for the 1MM sequence (ANOVA, $F(2, 33,204) = [2,702.66]$, ****$P = 0$, $n = 4$, $n_{\text{total events}} = 11,069$) (ii). 2MM events had the highest alignment score for the 2MM sequence (ANOVA, $F(2, 33,597) = [7,763.95]$, ****$P = 0$, $n = 4$, $n_{\text{total events}} = 11,200$) (iii). Summary statistics for box plots: centre, median; bounds of box, interquartile range (IQR) 25th and 75th percentile; whiskers, minimum and maximum within 1.5 IQR. Barcoded probe concentration was 30 nM in all conditions. All experiments were performed in 2× sequencing buffer (700 mM KCl, 50 mM HEPES, 100 mM MgCl₂, 100 mM ATP, 4.4 mM EDTA (pH 8.0)).

patient analysis, which could further reduce the cost per test. Due to the simplicity of the experimental protocol, the portability of the platform and the rapid turnaround time for experiments, we are confident that this approach could have an extensive impact on current diagnostics.

Strategy for the multiplexed detection of analytes

A highly multiplexed detection strategy was achieved by combining nanopore sequencing with barcoded molecular probes that selectively bind to target analytes (Fig. 1a and Supplementary Table 1). Barcoded probes were incubated with target analytes (miRNAs, proteins, small molecules), sequenced with the MinION device (ONT) and, subsequently, the presence of target analytes was determined. The barcoded probes consist of three key regions: (1) adapter; (2) barcode; and (3) target binding region (Fig. 1b). The adapter is identical for all probes. The barcode sequence acts as a unique identifier and can have many permutations, with theoretically up to 1.18×10^{21} distinctive arrangements of the nucleotides for the barcode length used in these experiments. The target binding region can either be a complementary sequence (to bind miRNA or DNA) or an aptamer (to bind proteins and small molecules). The translocation of hybridized probes (target is bound) is slowed since the nanopore geometry does not allow the passage of a double-strand or protein/small molecule to feed through the nanopore (Fig. 1c). The slowed translocation can be observed in the event current signal as a 'delay' period. Events can then be subclassified with or without delay (Fig. 1d), revealing at the single-molecule level the presence of an analyte (Supplementary Fig. 1), whilst the sequenced barcode classifies the analyte being targeted.

To establish the multiplexed platform, we designed 40 unique barcoded probes to detect miRNAs implicated in cardiac disease. The accuracy of barcode base calling and classification was determined in experiments without target analytes (Fig. 2). Events were sequenced and aligned against a barcode library with known barcode sequences (Supplementary Table 1). For example, the alignment score of a unique barcode, 'barcode 38', is shown in Fig. 2a. The normalized alignment score of events for the true sequence was significantly higher compared to all other barcode sequences in the library (Fig. 2a, Supplementary Fig. 2a and Supplementary Table 2). However, alignment scoring was not suitable in multiplexed experiments due to the occurrence of false positives. To address this, a series of thresholds were tested, including: (1) the number of mismatches; (2) x mismatches in the first y bases; (3) number of aligned bases; and (4) sequence beginning with the bases 'GGG' (Fig. 2b). It was found that reducing the total number of mismatched bases worked poorly on its own (area under the
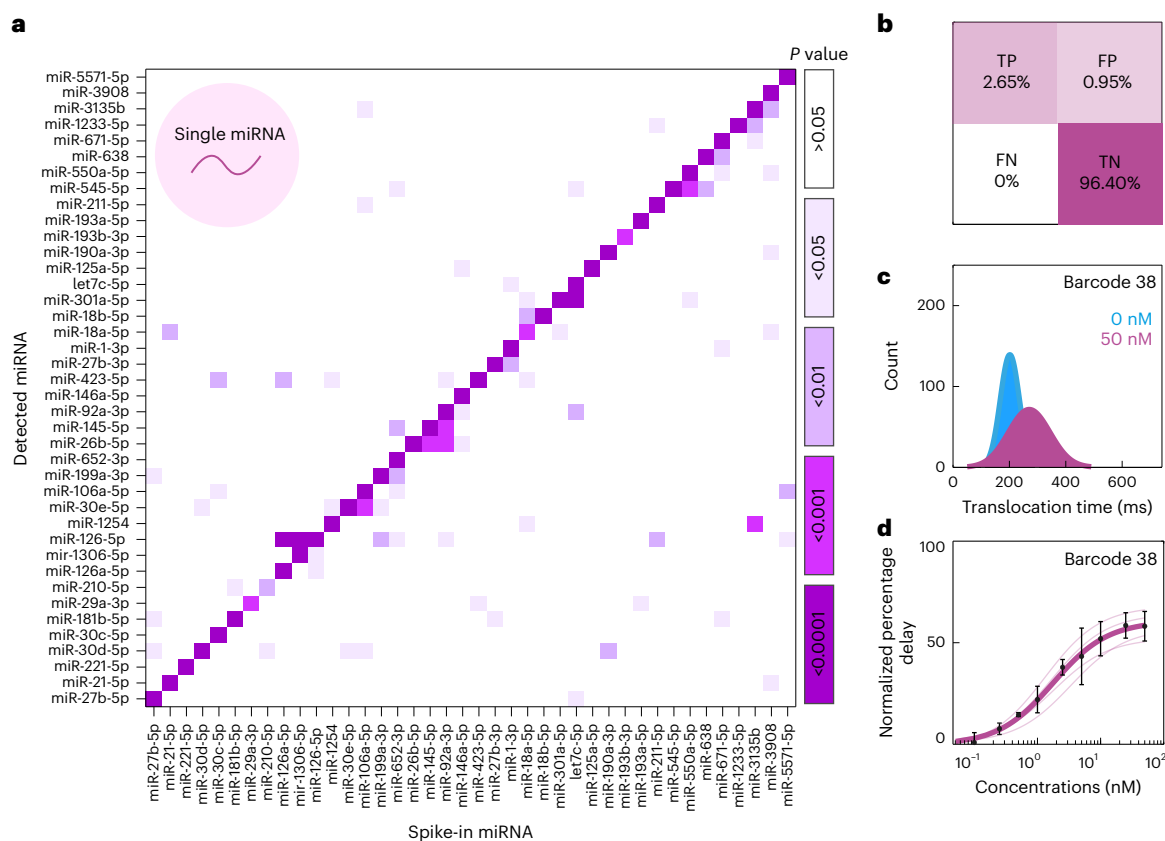
Fig. 3 | Sensitivity and selectivity of single miRNA detection. a, Heat map showing significance for selective miRNA binding. The 40 barcoded probes (30 nM) were incubated with each miRNA (10 nM), respectively. A significant increase (two-tailed t-test, **$P \le 0.01$) in percentage delayed events was observed for all barcoded probes when in the presence of their miRNA target ($n = 3$, $n_{\text{total events}} = 712,400$). **b**, Confusion matrix showing true positive (TP), false positive (FP), false negative (FN) and true negative (TN) occurrence. The accuracy was 99.05%, specificity was 99.02% and sensitivity was 100%. **c**, Translocation time of single barcoded probe and single miRNA experiments ($n = 5$, $n_{\text{total events}} = 54,197$). **d**, Concentration–percentage delay relationship of single miRNA and single barcoded probe experiment (barcoded probe 38, miR-221-5p). Data were fitted using the Hill equation with $n_{\text{H}} = 1$ (dissociation constant (K_{d}) = 1.77 nM and $V_{\max} = 61.03\%$, $R^2 = 96.25$, $n = 5$, $n_{\text{total events}} = 54,197$). Data presented as mean ± s.d. All experiments were performed in sequencing buffer.

curve (AUC) = 0.414). We observed that mismatches at the start of the sequenced event were particularly indicative of poorly resolved events. Consequently, a threshold based on allowing x mismatches in the first y bases was found to be moderately effective in removing unwanted events (AUC = 0.735). Increasing the total number of aligned bases was the second-best applied threshold (AUC = 0.828). Requiring the sequenced events to start with G residues (as this is common among all barcoded probes) proved to be the most efficient at separating true and false events (AUC = 0.846). Finally, we combined each criterion to determine the optimum threshold: ≤5 mismatched bases total; 1 mismatch in the first 10 bases; ≥15 bases aligned in total; sequence starts with 'GGG' (AUC = 0.805). This configuration did not improve sensitivity but markedly reduced the false-positive rate (FPR). The optimized thresholds resulted in an accuracy of >95% in the alignment of base-called events for single barcoded probe experiments (Fig. 2c). To determine the sensitivity of the platform, we tested a barcode against two further probes with one and two mismatched bases. In a multiplexed experiment with all three of the barcoded probes present, the alignment score for the correct barcode was significantly higher than the alignment scores of the other sequences (Fig. 2d).

Hybridization of barcoded probes with miRNA

Many miRNAs share similar sequences[47]. To demonstrate sequence-specific miRNA detection, a mixture of 40 barcoded probes (30 nM) was incubated with each miRNA (10 nM) individually (Fig. 3a). It was found that the percentage delayed events of each barcoded probe

increased significantly ($P < 0.01$) when in the presence of its corresponding target miRNA. We assessed the increase in percentage delayed events by comparing the miRNA sequence homologies (Supplementary Fig. 3a). Using a threshold of ≥90% miRNA sequence similarity (Supplementary Fig. 4a) and a significant increase in percentage delay ($P \le 0.01$), we found that 2.65% of all classifications were true positive, 0.95% false positive, 0% false negative and 96.40% true negative (Fig. 3b), resulting in a platform accuracy of 99.05%, specificity of 99.02% and sensitivity of 100%. Importantly, we observed that the barcode sequence does not influence the percentage delayed events detected for our probes (Supplementary Fig. 4b). In single barcoded probe experiments (for example, barcoded probe 38 and miR-221-5p), the total event time increased when 50 nM of miRNA was present compared to the control (Fig. 3c and Supplementary Fig. 2b). However, it is known that the DNA-controlling motor enzyme does not move the probe through the pore at a consistent rate; therefore, the speed at which the barcoded probes are translocated varies[48]. Hence, rather than relying on the translocation time, the moving standard deviation of the electrical signal was used to identify signal delays. Using this method, it was possible to determine a concentration–percentage delay curve for each barcoded probe and miRNA combination ($R^2 \ge 0.989$) (Fig. 3d and Supplementary Fig. 5).

Multiplexed detection of miRNA, protein and small molecules

Barcoded probes were incubated with synthetic miRNAs to identify whether event delays could be observed in multiplexed conditions
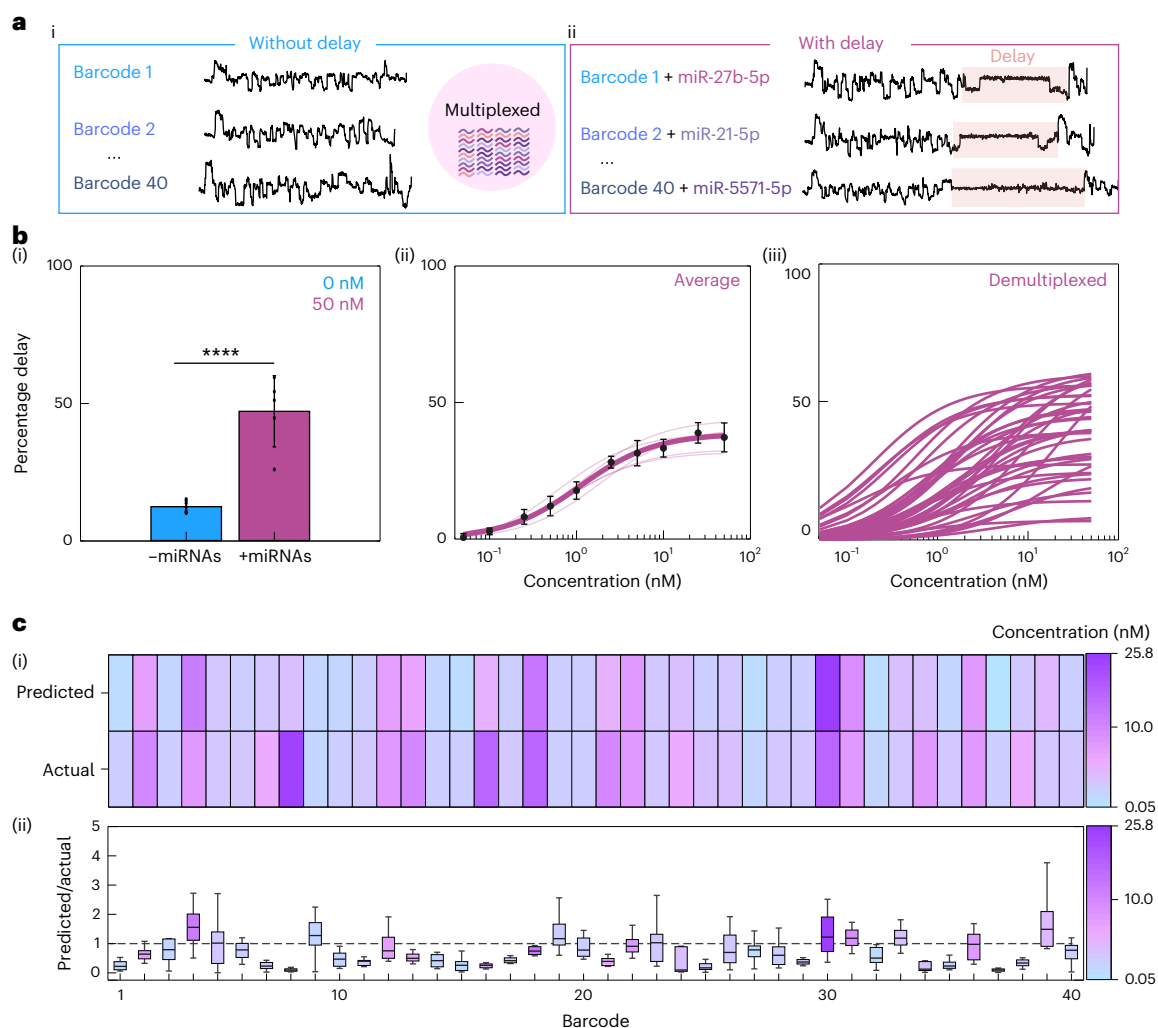
Fig. 4 | Multiplexed detection and quantification of miRNAs. a, Example current traces for barcodes without delay (i) and with delay (ii). **b**, Mean percentage delay ±50 nM miRNAs (two-tailed *t*-test, ****P = 3.58 × 10⁻⁴, n = 5, $n_{\text{total events}}$ = 76,841 (0 nM), 69,983 (50 nM)). Data presented as mean ± s.d. (i). Multiplexed concentration–percentage delay curves of 40 barcoded probes (30 nM) and the corresponding 40 miRNAs. A curve was fitted using the Hill equation with n_H = 1 (K_d = 1.09 nM, $V_{\max}$ = 38.69%, n = 5, $n_{\text{total events}}$ = 1,045,841). Data presented as mean ± s.d. (ii). Concentration–percentage delay curves of 40 individual barcoded probes derived from a multiplexed experiment. Curves

were fitted with the Hill fit (n_H = 1) background subtraction for each curve was performed using the percentage delay value at 0 nM for each curve (n = 5, $n_{\text{total events}}$ = 1,045,841) (iii). **c**, Single-blinded prediction of miRNA concentration and comparison to known concentration. Heat map of concentrations (n = 12, $n_{\text{total events}}$ = 203,812) (i). Analysis of miRNA predictions showing the value for predicted/actual concentration (n = 12, $n_{\text{total events}}$ = 203,812) (ii). Summary statistics for box plot: centre, median; bounds of box, IQR 25th and 75th percentile; whiskers, minimum and maximum within 1.5 IQR. All experiments were performed in sequencing buffer.

(Fig. 4a). It was found that the mean percentage delayed events increased when miRNAs were present (0 nM versus 50 nM, 12.32 ± 2.07% versus 46.99 ± 12.95%, two-tailed *t*-test, ****P = 3.577 × 10⁻⁴, n = 5) (Fig. 4b(i)). Without demultiplexing, it was found that there was a linear increase in percentage delayed events between 0.1 nM and 10 nM (R^2 ≥ 0.988), followed by a plateau (Fig. 4b(ii)). The signal was demultiplexed using our alignment protocol, identifying individual populations of each barcoded probe. We then constructed bespoke concentration–percentage delay curves for each probe (Fig. 4b(iii)), and data points for each curve were normalized with background subtraction. We next tested our ability to quantify multiple miRNAs using a single-blinded test. A total of 40 barcoded probes were incubated with a mix of 40 miRNAs at various concentrations, ranging from 0.25 nM to 20 nM (Fig. 4c). The event files were sent to a blinded researcher who determined the percentage delay of each barcoded probe and interpolated the value with the standard curves to determine the concentration of the miR-NAs. After estimating each miRNA concentration, the experiment was

unblinded and compared to the actual concentration (Fig. 4c(i) and Supplementary Fig. 6). Comparison of each miRNA prediction to its true value can be made by comparing the colour changes vertically. Similar colours mean a close prediction of miRNA concentration to the true value. The ratio of predicted concentration:actual concentration was plotted, allowing comparison across all 40 barcoded probes; values >1 were overpredictions and <1 were underpredicted (Fig. 4c(ii)). The accuracy of all predictions was within one order of magnitude of the true concentration.

This platform, due to its adaptability, has the potential to detect proteins, miRNAs and small molecules in a single experiment. Barcoded probes were designed with aptamers specific for each protein and small molecule (Supplementary Table 3; aptamer sequences were taken from published literature). In single barcoded probe experiments, a significant increase in translocation time could be observed when in the presence of its corresponding protein in pathophysiological con-centrations: thrombin (Fig. 5a, *P = 0.030); B-type natriuretic peptide
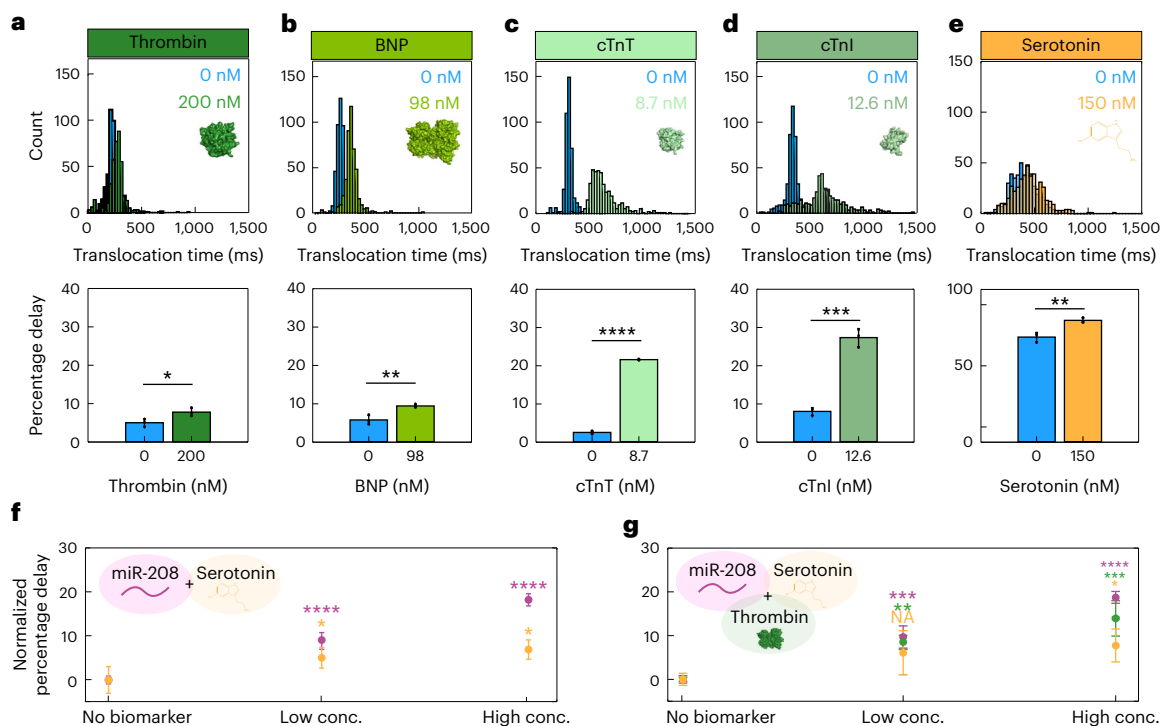
Fig. 5 | Protein and small molecule detection. a–e, Translocation time (top) and percentage delayed events (bottom) of barcoded probes (30 nM) ± each corresponding protein or small molecule: thrombin ($n = 3$, $n_{\text{total events}} = 417$ (0 nM), 1,341 (200 nM); *$P = 0.0303$ (**a**); BNP ($n = 3$, $n_{\text{total events}} = 5,256$ (0 nM), 1,040 (98 nM); **$P = 0.0082$ (**b**); cTnT ($n = 3$, $n_{\text{total events}} = 8,920$ (0 nM), 3,395 (8.7 nM); ****$P = 1.62 \times 10^{-7}$ (**c**); cTnI ($n = 3$, $n_{\text{total events}} = 1,024$ (0 nM), 1,024 (12.6 nM)); ***$P = 1.99 \times 10^{-4}$ (**d**); serotonin ($n = 3$, $n_{\text{total events}} = 2,329$ (0 nM), 1,602 (150 nM)); **$P = 0.005$ (**e**). Significance was determined by two-tailed t-tests. **f,** Multiplexed detection of serotonin and miR-208-5p. Low concentrations (conc.) (10 nM miR-208-5p, 150 nM serotonin) and high concentrations (50 nM miR-208-5p, 750 nM

serotonin) were compared to the normalized control (no biomarker present) (ANOVA, $F(2,9) = [96.04]$, ****$P = 8.49 \times 10^{-7}$, $F(2,9) = [6.36]$, *$P = 1.90 \times 10^{-2}$, $n = 3$, $n_{\text{total events}} = 65,545$). **g,** Multiplexed detection of miR-208-5p, serotonin and thrombin. Low concentrations (10 nM miR-208-5p, 150 nM serotonin, 300 nM thrombin) and high concentration (50 nM miR-208-5p, 750 nM serotonin, 1,500 nM thrombin) were compared to the normalized control (no biomarker present). Normalized percentage delay of barcoded probes is shown (ANOVA, $F(2,8) = [136.97]$, ****$P = 6.48 \times 10^{-71}$, $F(2,8) = [4.74]$, *$P = 4.38 \times 10^{-2}$, $F(2,8) = [28.57]$, ***$P = 2.00 \times 10^{-4}$, $n = 3$, $n_{\text{total events}} = 44,956$). All data are presented as mean ± s.d. All experiments were performed in sequencing buffer. NA, not applicable.

(Fig. 5b, **$P = 0.008$); cardiac troponin T (Fig. 5c, ****$P = 1.62 \times 10^{-7}$); and cardiac troponin I (Fig. 5d, ****$P = 1.99 \times 10^{-4}$, all two-tailed t-test). This corresponded with a significant increase in the percentage delayed events (Fig. 5a–d), suggesting that the detection of proteins is possible with this method. Moreover, a significant increase in translocation time and percentage delayed events was observed when serotonin (150 nM) was incubated with its corresponding barcoded probe (Fig. 5e, two-tailed t-test, **$P = 0.005$), despite our delay algorithm identifying a high number of delayed events in the control condition.

Multiplexed experiments were conducted by observing analytes of different molecular species. First, a duplex experiment of a small molecule and a miRNA was conducted. Incubation of barcoded probes with 50 nM small molecule and ≥3 nM miRNA significantly increased the percentage delayed events for each probe (Fig. 5f). We also performed a triplex experiment observing a miRNA, small molecule and protein simultaneously. The percentage delayed events for each barcoded probe increased significantly from the controls when miRNA (≥10 nM), small molecule (≥150 nM) and protein (≥50 nM) were present (Fig. 5g).
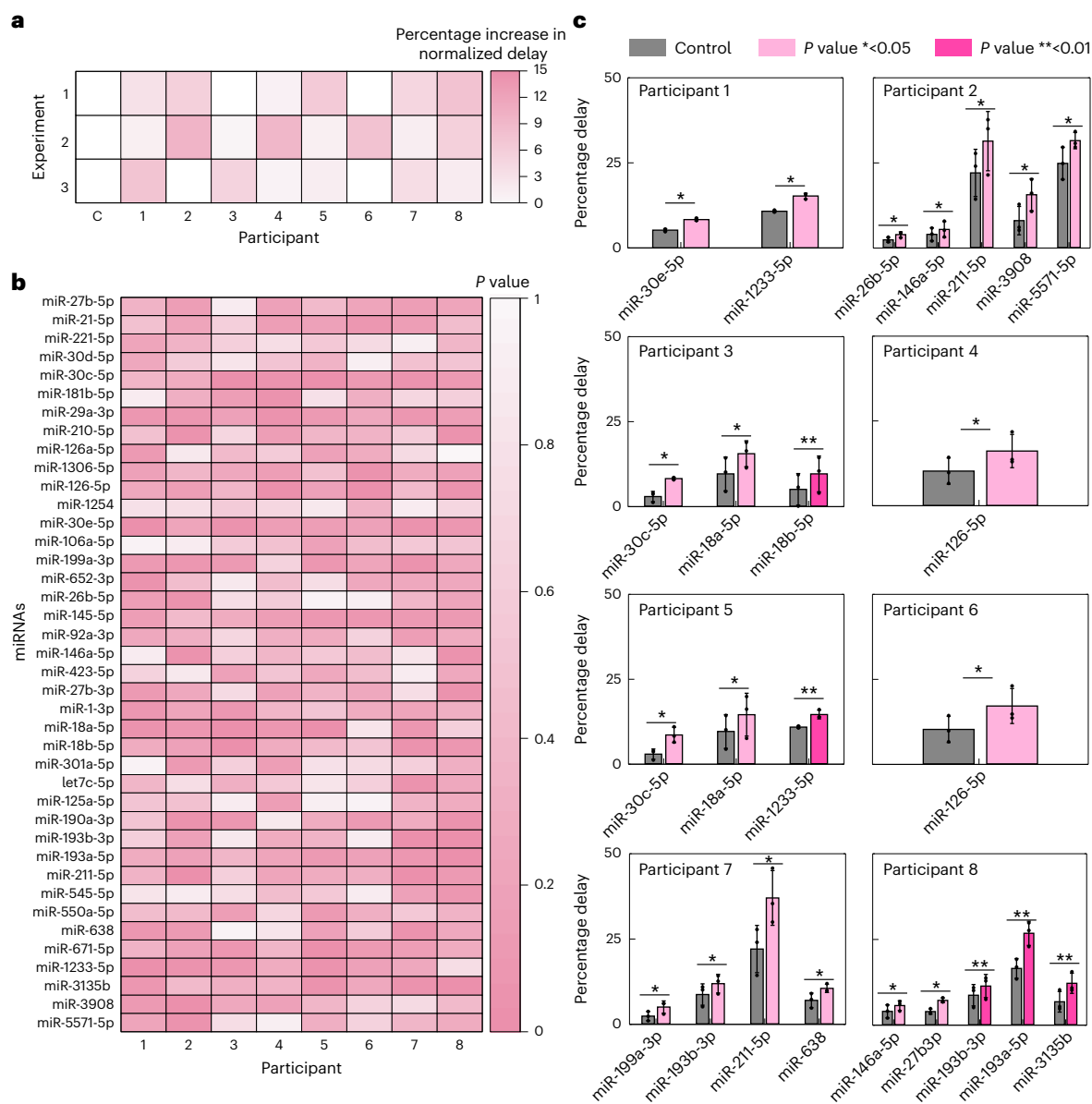
Detection of miRNAs in human serum

Blood serum from eight healthy participants was tested for the presence of 40 miRNAs and compared to the negative control (no serum, only buffer). To reduce pore blockage (Supplementary Fig. 7), serum was centrifuged through a 10 kDa molecular weight cut-off spin filter to remove large proteins. The list of miRNAs selected for detection in this assay have all been previously associated with cardiovascular disease (Supplementary Table 1). The mean percentage delay of all

miRNAs (not demultiplexed) was increased in each of the eight participant samples in three independent experiments; however, no significance was determined (Fig. 6a). Interestingly, when the signal was demultiplexed, significant changes in percentage delay of miRNAs could be observed (Fig. 6b). Of the 40 miRNAs tested, we saw a significant increase in 24 miRNAs across all eight participants (Fig. 6b,c and Supplementary Fig. 8). The most frequently observed miRNAs across all samples were: miR-1233-5p, miR146a-5p, miR211-5p, miR30c-5p, miR18a-5p, miR-126-5p and miR193b-3p. To further verify our results, we compared percentage delayed events with RT–qPCR of miR-29a in four samples. We found that our nanopore-based platform agreed with the RT–qPCR assay (Supplementary Fig. 9).

Conclusions

There is a genuine clinical need to develop platforms that can rapidly detect multiple biomarkers in patient samples. By expanding the panel of molecules observed, it is possible that disease subpopulations could be identified and treatment regimens optimized. Detection of biomarkers of different molecular species within the same sample could also massively reduce the analysis pipeline.

Our data show that barcoded probes, in combination with the ONT sequencing platform, offer a highly accurate and sensitive analyte sensing platform with single-molecule resolution. We detected 40 different barcoded probes in the same sample and constructed independent concentration–delay curves. We then used these standard curves to predict miRNA concentrations in a single-blinded experiment. The multiplexed method was applied to multiple molecular species, allowing

Fig. 6 | Multiplexed detection of miRNAs in human serum. a, Mean percentage delay of barcoded probes (not demultiplexed). Serum, after centrifugation with 10 kDa molecular weight cut-off spin filter, was incubated with 40 barcoded probes (30 nM) and compared to zero serum control (one-tailed t-test $P > 0.05$, $n = 3$, $n_{total events} = 194,115$). **b**, Heat map of P values comparing percentage delayed events ±serum (one-tailed t-test, $n = 3$, $n_{total events} = 194,115$). **c**, Significantly upregulated miRNAs per patient (one-tailed t-test *$P \leq 0.05$, **$P \leq 0.01$, $n = 3$, $n_{total events} = 194,115$). Exact P values are available from the source data file for this figure. Data are presented as mean ± s.d. All experiments were performed in sequencing buffer.

us to observe an increase in percentage delayed events due to proteins and small molecules (as well as miRNAs). Finally, we showed that the platform is compatible with human serum, indicating its potential as a biomarker detection platform.

The assay time can be less than 60 minutes in the platform's current configuration. This is a clear improvement compared to other technologies, which take several hours or even days to complete[49]. ONT sequencing software can now spot specific barcode sequences, allowing for live demultiplexing, which will further reduce assay time and provide a critical advantage in the detection of analytes with short half-life. Moreover, since the MinION device is highly portable, assays could be performed away from the clinic. The speed and portability of the platform offer further advantages over other biomarker detection strategies.

The platform can correctly distinguish between barcodes with a single nucleotide change in the sequence. Interaction of probes

was predicted (Supplementary Fig. 3); however, there was no clear relationship between dimer formation and event capture rate. This indicates that any future applications of this technology will not be limited by barcode sequences, but perhaps by the selectivity of the binding regions. Our data suggest that miRNA sequences with a similarity ≥90% have the potential to hybridize with the incorrect probe (Supplementary Fig. 10); strategies to distinguish between such miRNAs must therefore be further developed. There is also potential for the large number of biomarkers assessed to result in 'data blindness', to address this a graphical interface or risk-scoring method must be developed to assist the end user.

Currently, the detection limit of the platform for miRNA is approximately 50 pM. There is potential to push the technology further by reducing the false-negative rate of event alignment. At the moment, there are a large number of events that are rejected due to

the thresholds applied to ensure high accuracy. One way to reduce this event loss and increase events available for delay analysis would be to switch to a different nanopore model.

A particular issue of assaying blood serum is that nanopore lumen are small and are easily blocked by proteins. Pore blockage reduces the barcoded probe capture rate, meaning assay times must be increased or repeated. To reduce the blockage of nanopores by proteins in this study, we added a filtration step, which removed the largest serum proteins (>10 kDa) but prevented simultaneous miRNA and protein detection in serum. In future applications, it would be preferable to develop our methodology further to reduce the sample processing required before detection whilst also reducing the unwanted pore blocking. Limiting proteolysis whilst preventing miRNA degradation is a key hurdle to overcome in the development of an assay that can detect both proteins and miRNAs in serum simultaneously.

The platform shows great potential for use in clinical environments, for example, to offer expansive, longitudinal disease tracking or early disease detection. The assay could be adapted for use with a variety of complex fluids, for example, saliva, urine and cerebrospinal fluid. With further optimization, this strategy could significantly reduce testing time, assay cost and sample volume, whilst increasing the data available to the clinician.

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/s41565-023-01479-z.

References

1. Gold, P. & Freedman, S. O. Demonstration of tumor-specific antigens in human colonic cardinomata by immunological tolerance and absorption techniques. *J. Exp. Med.* **121**, 439–462 (1965).
2. Wide, L., Roos, P. & Gemzell, C. Immunological determination of human pituitary luteinizing hormone (LH). *Acta Endocrinol.* **37**, 445–449 (1961).
3. LaDue, J. S., Wróblewski, F. & Karmen, A. Serum glutamic oxaloacetic transaminase activity in human acute transmural myocardial infarction. *Science* **120**, 497–499 (1954).
4. Raizada, A. et al. Brain type natriuretic peptide (BNP)—as marker of new millennium in diagnosis of congestive heart failure. *Indian J. Clin. Biochem.* **22**, 4–9 (2007).
5. Montalescot, G., Guedeney, P. & Tijssen, J. A multi-biomarker score for a global approach of risk: time for a change?. *J. Am. Coll. Cardiol.* **80**, 898–901 (2022).
6. Hye, A. et al. Proteome-based plasma biomarkers for Alzheimer's disease. *Brain* **129**, 3042–3050 (2006).
7. Ueno, I., Sakai, T., Yamaoka, M., Yoshida, R. & Tsugita, A. Analysis of blood plasma proteins in patients with Alzheimer's disease by two-dimensional electrophoresis, sequence homology and immunodetection. *Electrophoresis* **21**, 1832–1845 (2000).
8. Cao, M. C. et al. Serum biomarkers of neuroinflammation and blood–brain barrier leakage in amyotrophic lateral sclerosis. *BMC Neurol.* **22**, 216–216 (2022).
9. de Ávila, B. E.-F. et al. Multiplexed determination of amino-terminal pro-B-type natriuretic peptide and C-reactive protein cardiac biomarkers in human serum at a disposable electrochemical magnetoimmunosensor. *Electroanalysis* **26**, 254–261 (2014).
10. Sonawane, M. D., Nimse, S. B., Song, K. S. & Kim, T. Multiplex detection of cardiac biomarkers. *Anal. Methods* **9**, 3773–3776 (2017).
11. Zhang, D. et al. Quantitative detection of multiplex cardiac biomarkers with encoded SERS nanotags on a single T line in lateral flow assay. *Sens. Actuators B* **277**, 502–509 (2018).
12. An, B. et al. An antibody-free platform for multiplexed, sensitive quantification of protein biomarkers in complex biomatrices. *J. Chromatogr. A* **1676**, 463261–463261 (2022).
13. Najjar, D. et al. A lab-on-a-chip for the concurrent electrochemical detection of SARS-CoV-2 RNA and anti-SARS-CoV-2 antibodies in saliva and plasma. *Nat. Biomed. Eng.* **6**, 968–978 (2022).
14. Leandersson, P., Åkesson, A., Hedenfalk, I., Malander, S. & Borgfeldt, C. A multiplex biomarker assay improves the diagnostic performance of HE4 and CA125 in ovarian tumor patients. *PLoS ONE* **15**, e0240418 (2020).
15. Ma, S. et al. Multiplexed serum biomarkers for the detection of lung cancer. *eBioMedicine* **11**, 210–218 (2016).
16. Opstal-van Winden, A. et al. A bead-based multiplexed immunoassay to evaluate breast cancer biomarkers for early detection in pre-diagnostic serum. *Int. J. Mol. Sci.* **13**, 13587–13604 (2012).
17. Pan, S. et al. Multiplex targeted proteomic assay for biomarker detection in plasma: a pancreatic cancer biomarker case study. *J. Proteome Res.* **11**, 1937–1948 (2012).
18. Cook, D. B. et al. Multiplexing protein and gene level measurements on a single Luminex platform. *Methods* **158**, 27–32 (2019).
19. Wilson, D. H. et al. The Simoa HD-1 analyzer: a novel fully automated digital immunoassay analyzer with single-molecule sensitivity and multiplexing. *J. Lab. Autom.* **21**, 533–547 (2016).
20. Gold, L. et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **5**, e15004 (2010).
21. Rohloff, J. C. et al. Nucleic acid ligands with protein-like side chains: modified aptamers and their use as diagnostic and therapeutic agents. *Mol. Ther. Nucleic Acids* **3**, e201 (2014).
22. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Front. Endocrinol.* **9**, 402–402 (2018).
23. Romaine, S. P. R., Tomaszewski, M., Condorelli, G. & Samani, N. J. MicroRNAs in cardiovascular disease: an introduction for clinicians. *Heart* **101**, 921–928 (2015).
24. Zhou, S. S. et al. miRNAS in cardiovascular diseases: potential biomarkers, therapeutic targets and challenges. *Acta Pharmacol. Sin.* **39**, 1073–1084 (2018).
25. Schueller, F. et al. The role of miRNAs in the pathophysiology of liver diseases and toxicity. *Int. J. Mol. Sci.* **19**, 261 (2018).
26. Connor, K. L. & Denby, L. MicroRNAs as non-invasive biomarkers of renal disease. *Nephrol. Dial. Transplant.* **36**, 428–429 (2021).
27. Cao, D. D., Li, L. & Chan, W. Y. MicroRNAs: key regulators in the central nervous system and their implication in neurological diseases. *Int. J. Mol. Sci.* **17**, 842 (2016).
28. Peng, Y. & Croce, C. M. The role of MicroRNAs in human cancer. *Signal Transduct. Target. Ther.* **1**, 15004 (2016).
29. Qin, S. & Zhang, C. MicroRNAs in vascular disease. *J. Cardiovasc. Pharmacol.* **57**, 8–12 (2011).
30. Mestdagh, P. et al. Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat. Methods* **11**, 809–815 (2014).
31. Lin, X., Ivanov, A. P. & Edel, J. B. Selective single molecule nanopore sensing of proteins using DNA aptamer-functionalised gold nanoparticles. *Chem. Sci.* **8**, 3905–3912 (2017).
32. Ren, R. et al. Nanopore extended field-effect transistor for selective single-molecule biosensing. *Nat. Commun.* **8**, 586 (2017).
33. Wanunu, M. et al. Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors. *Nat. Nanotechnol.* **5**, 807–814 (2010).
34. Xue, L. et al. Solid-state nanopore sensors. *Nat. Rev. Mater.* **5**, 931–951 (2020).

35. Sultan, M. & Kanavarioti, A. Nanopore device-based fingerprinting of RNA oligos and microRNAs enhanced with an osmium tag. *Sci. Rep.* **9**, 14180 (2019).

36. Miles, B. N. et al. Single molecule sensing with solid-state nanopores: novel materials, methods, and applications. *Chem. Soc. Rev.* **42**, 15–28 (2012).

37. Al Sulaiman, D., Cadinu, P., Ivanov, A. P., Edel, J. B. & Ladame, S. Chemically modified hydrogel-filled nanopores: a tunable platform for single-molecule sensing. *Nano Lett.* **18**, 6084–6093 (2018).

38. Ren, R. et al. Selective sensing of proteins using aptamer functionalized nanopore extended field-effect transistors. *Small Methods* **4**, 2000356 (2020).

39. Tan, S. et al. DNA-functionalized silicon nitride nanopores for sequence-specific recognition of DNA biosensor. *Nanoscale Res. Lett.* **10**, 205 (2015).

40. Wei, R., Gatterdam, V., Wieneke, R., Tampé, R. & Rant, U. Stochastic sensing of proteins with receptor-modified solid-state nanopores. *Nat. Nanotechnol.* **7**, 257–263 (2012).

41. Sze, J. Y. Y., Ivanov, A. P., Cass, A. E. G. & Edel, J. B. Single molecule multiplexed nanopore protein screening in human serum using aptamer modified DNA carriers. *Nat. Commun.* **8**, 1552 (2017).

42. Cai, S., Sze, J. Y. Y., Ivanov, A. P. & Edel, J. B. Small molecule electro-optical binding assay using nanopores. *Nat. Commun.* **10**, 1797 (2019).

43. Cai, S. et al. Single-molecule amplification-free multiplexed detection of circulating microRNA cancer biomarkers from serum. *Nat. Commun.* **12**, 3515 (2021).

44. Liu, H. et al. Expression and purification of a novel mycobacterial porin MspA mutant in *E. coli*. *J. Nanosci. Nanotechnol.* **17**, 9125–9129 (2017).

45. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524 (2016).

46. One technology, one platform for all your biology. *Oxford Nanopore Technologies* https://nanoporetech.com/applications/techniques/short-fragment-mode (2023).

47. Kamanu, T. K. K., Radovanovic, A., Archer, J. A. C. & Bajic, V. B. Exploration of miRNA families for hypotheses generation. *Sci. Rep.* **3**, 2940 (2013).

48. Craig, J. M. et al. Revealing dynamics of helicase translocation on single-stranded DNA using high-resolution nanopore tweezers. *Proc. Natl Acad. Sci. USA* **114**, 11932–11937 (2017).

49. Chen, C. et al. Real-time quantification of microRNAs by stem–loop RT–PCR. *Nucleic Acids Res.* **33**, e179 (2005).

Methods

Chemicals and materials

All flow cells (MinION and Flongle) were provided by ONT. All probe sequences were custom designed. Barcoded probes and miRNAs were synthesized by Integrated DNA Technologies (Supplementary Tables 1 and 3). Proteins were obtained from the following: thrombin (Invitrogen, catalogue no. RP-43100); B-type natriuretic peptide (Bachem, catalogue no. 4095916); cardiac troponin I (Genscript Biotech, catalogue no. Z03320); cardiac troponin T (Ray Biotech, catalogue no. 230-00048). Small molecules: serotonin (Sigma-Aldrich, catalogue no. H9523). Ampure XP magnetic beads (for purification of DNA-barcoded probes) were purchased from Beckman Coulter. TA ligase was acquired from New England Biolabs. All chemicals used for buffer preparation were obtained from Sigma-Aldrich, Roche or VWR Chemicals. In all experiments, DNA lo-bind tubes were used (Eppendorf).

Human donor blood serum

Human samples used in this research project were obtained from the Imperial College Healthcare Tissue Bank (ICHTB). ICHTB is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. ICHTB is approved by Wales REC3 to release human material for research (22/WA/0214), and the samples for this project (R22016) were issued from subcollection reference number NHL_FN_021_028.

Venous blood was collected in red-topped vacutainers (Beckton Dickinson) and allowed to clot at room temperature for 15 min before centrifugation at 3,000g, 15 min, 4 °C. The resulting serum was then aliquoted into small volumes and frozen at −80 °C until use. The serum was filtered using a 10 kDa molecular weight cut-off spin filter (Sartorius) before incubation with barcoded probes.

Barcoded probe design

The complete carrier design consists of three sections named 'adapter', 'barcode region' and 'target binding region' (Fig. 1a); when fully assembled, this was called a 'barcoded probe'. The adapter section is identical for all probes. It is the ONT Y-adapter, which consists of (1) a leader (which facilitates threading into the nanopore); (2) a tether (to enhance the capture rate); and (3) a motor protein (to control the translocation of the barcoded probe through a nanopore). The adapter section is ligated to the 5′ end of all barcoded probes, ensuring that translocation events are always in the 5′ to 3′ direction. The barcode region consists of a polynucleotide identifier and is followed by spacer nucleotides (to separate the barcode and target binding regions). The target binding region consists of a DNA aptamer or complementary miRNA sequence, depending on the species of the target analyte.

Barcoded probe preparation

Each probe was incubated with ligation complementary strand in a molar ratio of 1:3 in nuclease-free water at room temperature for 1 h. The resultant mix was combined with 10 nM adapter and an equal volume of TA ligase master mix (New England Biolabs). The mix was centrifuged at 4 °C for 1 min and then incubated at room temperature for 20 min. Probes were purified using the solid phase reversible immobilization method. Ampure XP beads (Beckman Coulter) were added at 1.4 times the total solution volume. The beads (with probes bound) were washed two times with a short fragment buffer (ONT). After the washes, the beads were resuspended in nuclease-free water, causing the probes to be released. A 100 nM tether was then added to the probes along with sequencing buffer. The 2× sequencing buffer contained 700 mM KCl, 50 mM HEPES, 100 mM MgCl$_2$, 100 mM ATP, 4.4 mM EDTA (pH 8.0). Barcoded probes were incubated with the target analyte for 30 min at room temperature before loading into flow cells. The concentration of the barcoded probes in all experiments was 30 nM, as it was more sensitive than 100 nM (Supplementary Fig. 11).

Preprepared mixes of barcoded probes are resilient to multiple freeze–thaw cycles and long-term storage at −20 °C (Supplementary Fig. 12). Hybridization dynamics of miRNAs with barcoded probes were also investigated, revealing that incubation of 5–10 minutes is sufficient to reach equilibrium for probe–analyte binding (Supplementary Fig. 13). This can potentially reduce the sample preparation time and assay variability significantly. A gel binding assay was performed to confirm interactions of barcoded probes and their corresponding miRNAs (Supplementary Fig. 14).

Sequencing experiments and data acquisition

All sequencing experiments were performed at 34 °C using either the MinION or Flongle sequencing device (ONT). The MinION/Flongle was connected through a USB 3.0 port to a PC with a minimum of 16 GB RAM. A membrane check was performed before each run to determine the integrity of the membrane and to identify how many nanopores were active. Before each experiment, flow cells were flushed with 2× 500 µl (MinION) or 1× 150 µl (Flongle) sequencing buffer. The volume for each sequencing experiment was determined by the flow cell used: 150 µl (MinION) or 30 µl (Flongle). After each experiment, the MinION flow cell was washed with the flow cell wash kit (ONT) according to the manufacturer's protocol.

Data collection was performed using the proprietary software MinKNOW (ONT). Base calling was either performed in real time (MinKNOW) or offline within a custom-written MATLAB script, 'The Nanopore App', previously published by our group[43].

Event analysis

Barcoded probe translocations were identified and analysed using the following workflow: (1) event identification; (2) event base calling; (3) event alignment; (4) event delay analysis.

Event identification included the tracking and then subtraction of the baseline signal. A cut-off threshold was then determined on the basis of the background noise (30–40 standard deviations above mean noise level). 'Peakfinder' function in MATLAB was used to spot events. Any events identified that were shorter than 100 ms were excluded at this stage.

Event base calling was performed with the Guppy base caller using the high accuracy flip-flop model (ONT).

Each base-called event was aligned against a library of barcode sequences. Each event was attributed to one of the library sequences on the basis of an alignment scoring method. We found that the accuracy of the Guppy base caller was much more accurate at the 5′ end of the barcode sequence. Further thresholds were applied to ensure that only true positive barcode events were retained for further analysis. These thresholds were: (1) sequence starts with 'GGG', (2) ≥15 bases aligned to the library sequence; (3) one mismatched base in the first ten bases; and (4) ≤5 mismatched bases in the entire sequence. Further information is in Supplementary Methods ('Event selection pipeline').

To distinguish between analyte-bound and unbound barcoded probe events, we performed analysis of each event. When an analyte is bound to the barcoded probe, there is a 'quiet' sojourn in the electrical signal, which persists until the analyte is dissociated from the probe, at which point it can complete its translocation. An event is defined as delayed if the moving standard deviation of the signal is less than the threshold of 0.003 for a period greater than 10 bins (each event signal is separated into a total of 75 bins). All other events are defined as having no delay.

Statistical tests

Where possible, a Kolmogorov–Smirnov test was performed to determine whether our data were normally distributed at the 5% significance level.

One-sample t-tests, were performed with $n − 1$ degrees of freedom unless otherwise stated. Details are indicated in the figure legends. No statistical methods were used to predetermine sample sizes.

The experiments performed to generate the data in Fig. 4c were single-blinded (the analyst was blinded to sample contents). The experiments performed to generate the data in Fig. 6 and Supplementary Figs. 8 and 9 were fully blinded to all researchers.

Reporting summary

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

Data availability

Source data are provided with this paper. Further data supporting the plots within this paper and other study findings are available in the supplementary files or by contacting the corresponding authors.

Code availability

The Nanopore App used to analyse the data is available at https://www.imperial.ac.uk/people/joshua.edel or by request to the corresponding authors.

Author contributions

C.K., B.R.-O'D., A.P.I. and J.B.E. conceived and designed experiments. C.K. and B.R.-O'D. performed experiments and wrote the paper. J.B.E. wrote and developed the Nanopore App analysis software. C.L. collected and prepared human serum samples. B.R.-O'D., F.S.N. and J.G. obtained the ethical approval for the collection and storage of human serum samples. R.G., J.G., A.P.I. and J.B.E. were responsible for the conception and supervision of the project. All authors edited the paper.

Competing interests

R.G. is an employee of Oxford Nanopore Technologies PLC. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41565-023-01479-z.

Correspondence and requests for materials should be addressed to Aleksandar P. Ivanov or Joshua B. Edel.

Peer review information *Nature Nanotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at 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.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

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	
☐	☑	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☑	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☑	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☑	☐	A description of all covariates tested
☐	☑	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☑	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)
☐	☑	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☑	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☑	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☑	☐	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	All data was collected using "MinKNOW" software v4.2 (or higher), available from Oxford Nanopore Technologies
Data analysis	All data was analysed with "The nanopore app" v7 (or higher), which runs on Matlab 2021A (or higher), available from corresponding authors upon request. The nanopore app used "Guppy" v5 for sequencing and basecalling, available from Oxford Nanopore Technologies.

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 for figure 4 were supplied as part of the submission. Other findings of this study are available from the corresponding author upon reasonable request. Source data for the figures and supplementary figures are provided as a Source Data file.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](). See also policy information about [sex, gender (identity/presentation), and sexual orientation]() and [race, ethnicity and racism]().

Reporting on sex and gender	We do not report the sex or gender of any donors within our study. We did not consider sex or gender within the study design. The study did not aim to determine any clinically relevant information but was to esablish a techology therefore consideration of gender or sex at this stage was irrelevant.
Reporting on race, ethnicity, or other socially relevant groupings	We did not report the race, ethnicity, or other socially relevant groupings within our study, for reasons detailed above
Population characteristics	Age: 26-35 years; Gender: male/female,50/50;
Recruitment	Participants were recruited from Students and Staff of Imperial College London via word of mouth and posters. All participants identified as 'healthy', using the study inclusion criteria. Participant inclusion criteria were: >18 years old; No known cardiovascular disease; no COVID-19 symptoms; no blood borne disease; not pregnant/ breast feeding; not prescribed diuretics/beta-blockers/ACE inhibitors/ Ca channel blockers/Digoxin
Ethics oversight	Human samples used in this research project were obtained from the Imperial College Healthcare Tissue Bank (ICHTB). ICHTB is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. ICHTB is approved by Wales REC3 to release human material for research (22/WA/0214), and the samples for this project (R22016) were issued from subcollection reference number NHL_FN_021_028.

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]()

Life sciences study design

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

Sample size	No sample size calculation was performed, there was no comparison between study groups. The use of human samples was to establish the technology, rather than observe clinical characteristics
Data exclusions	No data was excluded from the study
Replication	All experiements were performed with technical triplicates successfully
Randomization	There was no randomisation as the study was observational (there was no treatment performed on any samples)
Blinding	Blinding was not relevant to the human sample arm of the study as there were no groups allocated. We indicate in the manuscript where experiments have been blinded

Behavioural & social sciences study design

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

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing	
Data exclusions	
Non-participation	
Randomization	

Ecological, evolutionary & environmental sciences study design

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

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing and spatial scale	
Data exclusions	
Reproducibility	
Randomization	
Blinding	

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	n/a
Location	n/a
Access & import/export	n/a
Disturbance	n/a

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	Involved in the study
☑	☐ Antibodies
☑	☐ Eukaryotic cell lines
☑	☐ Palaeontology and archaeology
☑	☐ Animals and other organisms
☑	☐ Clinical data
☑	☐ Dual use research of concern
☑	☐ Plants

Methods

n/a	Involved in the study
☑	☐ ChIP-seq
☑	☐ Flow cytometry
☑	☐ MRI-based neuroimaging

Antibodies

Antibodies used	
Validation	

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

Authentication

Mycoplasma contamination

Commonly misidentified lines
(See [ICLAC](#) register)

Palaeontology and Archaeology

Specimen provenance

Specimen deposition

Dating methods

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

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

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

Wild animals

Reporting on sex

Field-collected samples

Ethics oversight

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

Clinical data

Policy information about [clinical studies](#)
All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Study protocol

Data collection

Outcomes

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

	No	Yes	
	☐	☐	Public health
	☐	☐	National security
	☐	☐	Crops and/or livestock
	☐	☐	Ecosystems
	☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

	No	Yes	
	☐	☐	Demonstrate how to render a vaccine ineffective
	☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
	☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
	☐	☐	Increase transmissibility of a pathogen
	☐	☐	Alter the host range of a pathogen
	☐	☐	Enable evasion of diagnostic/detection modalities
	☐	☐	Enable the weaponization of a biological agent or toxin
	☐	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks

Novel plant genotypes

Authentication

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

Files in database submission

Genome browser session
(e.g. UCSC)

Methodology

Replicates

Sequencing depth

Antibodies

Peak calling parameters

Data quality

Software

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Instrument

Software

Cell population abundance

Gating strategy

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type

Design specifications

Behavioral performance measures

Imaging type(s)

Field strength

Sequence & imaging parameters

Area of acquisition

Diffusion MRI   ☐ Used   ☐ Not used

Preprocessing

Preprocessing software

Normalization

Normalization template

Noise and artifact removal

Volume censoring

Statistical modeling & inference

Model type and settings

Effect(s) tested

Specify type of analysis: ☐ Whole brain   ☐ ROI-based   ☐ Both

Statistic type for inference

(See Eklund et al. 2016)

Correction

Models & analysis

n/a	Involved in the study
☐ ☐	Functional and/or effective connectivity
☐ ☐	Graph analysis
☐ ☐	Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis