nature nanotechnology

ARTICLES

https://doi.org/10.1038/s41565-020-0699-0

Cancer diagnosis with DNA molecular computation

Chao Zhang[1], Yumeng Zhao[1], Xuemei Xu[1], Rui Xu[1], Haowen Li[1], Xiaoyan Teng[2], Yuzhen Du[2], Yanyan Miao[1], Hsiao-chu Lin[1] and Da Han[1]✉

Early and precise cancer diagnosis substantially improves patient survival. Recent work has revealed that the levels of multiple microRNAs in serum are informative as biomarkers for the diagnosis of cancers. Here, we designed a DNA molecular computation platform for the analysis of miRNA profiles in clinical serum samples. A computational classifier is first trained in silico using miRNA profiles from The Cancer Genome Atlas. This is followed by a computationally powerful but simple molecular implementation scheme using DNA, as well as an effective in situ amplification and transformation method for miRNA enrichment in serum without perturbing the original variety and quantity information. We successfully achieved rapid and accurate cancer diagnosis using clinical serum samples from 22 healthy people (8) and people with lung cancer (14) with an accuracy of 86.4%. We envision that this DNA computational platform will inspire more clinical applications towards inexpensive, non-invasive and rapid disease screening, classification and progress monitoring.

There is an ever-increasing need for rapid, non-invasive, yet accurate methods for cancer diagnosis given that cancer is the most common cause of global disease-related mortality, and the current invasive, unpleasant and inconvenient clinical diagnostic procedures limit their applications[1]. A diagnostic method that can quickly and accurately analyse changes in cancer biomarkers in biological samples will greatly facilitate cancer diagnosis, monitoring, therapy and prognosis[2–4].

Compared to the genomic and proteomic marker molecules, such as DNA, messenger RNA and proteins, non-coding miRNA expression levels have been found to be extremely informative for classifying tumour and normal cells[5–7]. It has also been reported that the levels of miRNA in serum of patients with cancer is consistently and reproducibly different from those of healthy controls[8]. Therefore, miRNAs have become notable tumour markers for rapid and non-invasive cancer diagnosis[9].

However, miRNA levels of individuals exhibit a great complexity in their variety and quantity, and alterations in either factor can lead to different diagnostic results[7]. Current methods for miRNA identification, including quantitative PCR with reverse transcription (RT–qPCR), microarrays and RNA sequencing, can provide quantitative miRNA expression profiles for diagnostic applications; but they remain cost prohibitive, labour intensive and hardware limited under different situations[10]. In addition, the multiple rounds of parallel operations may lead to non-negligible errors in individual miRNA quantitation and subsequently affect the diagnostic results. Furthermore, data analysis for these methods requires expertise for comprehensive understanding, which limits the application scenarios of these methods to specialized laboratory settings. Accordingly, there is still need for diagnostic methods that can accurately analyse multiple miRNA changes correlated to cancers and report results directly and rapidly at the point of care.

DNA molecular computation provides a natural interface between molecular recognition and information processing, because DNA can interact with different molecules, transduce the signals and report the results in a programmable manner[11,12]. With these functions, DNA computation has been implemented for various applications, including solving non-polynomial problems in math[13], mimicking complex systems[14,15], performing logical biosensing[16,17], analysing cell surface receptors[18–22], sorting molecular cargos[23], playing logical games[24–27] and recognizing patterns[28]. However, rationally designed DNA computation systems have seldom been employed in diagnostic applications, where the integration of multiple biomarker recognition and logical information processing is most needed, especially in biological samples towards clinical applications, with only a few successful examples in classifying diseases in synthetic samples mimicking practical analyte composition[29–31].

In our opinion, several existing technical difficulties hinder the clinical applications of DNA molecular computation for cancer diagnostics. First, the incompatibility of complex biomarker inputs in biological samples with simple Boolean values (either high or low) compromises the accuracy of diagnosis. Specifically, it is difficult to develop powerful computation algorithms that allow simple molecular implementation without constraint by the variety or quantity of biomarker inputs. Second, the highly dynamic analyte concentration ranges in biological samples require robust and sensitive design for interfacing the sensing and subsequent computation[32]. In addition, highly specific biomarkers such as miRNA are typically short, unstructured oligonucleotides with sequence similarity between family members, which greatly affect the sequence orthogonality in the computation[7]. Therefore, we sought to resolve those challenges by implementing a DNA-based computational platform that operates on miRNAs in serum and rapidly produces diagnostic results for lung cancer.

The workflow of the DNA-based computational platform includes three main steps, as shown in Fig. 1a. First, an in silico linear classifier is constructed and trained using publicly available miRNA-seq profiles of healthy and non-small cell lung cancer (NSCLC) individuals from The Cancer Genome Atlas (TCGA). The goal of this step is to have a set of miRNA inputs with associated weights and a set of mathematical

[1]Institute of Molecular Medicine, Shanghai Key Laboratory for Nucleic Acid Chemistry and Nanomedicine, State Key Laboratory of Oncogenes and Related Genes, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. [2]Department of Laboratory Medicine, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China. ✉e-mail: dahan@sjtu.edu.cn

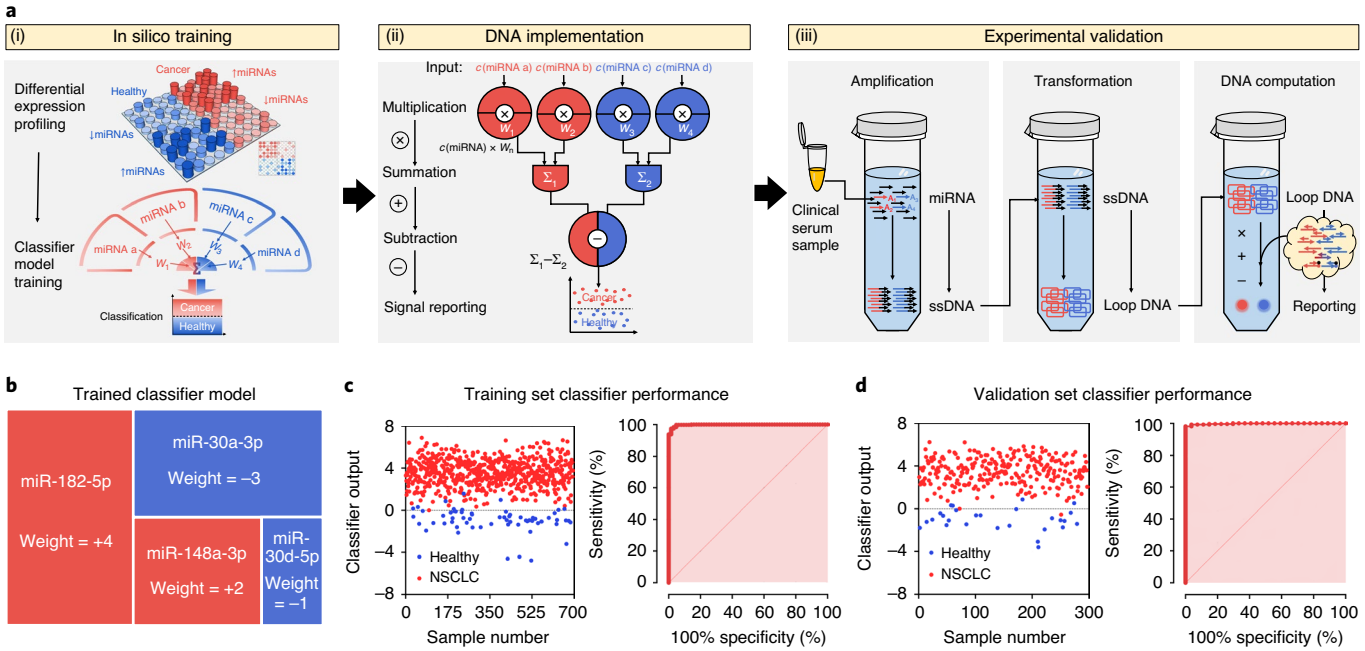

Fig. 1 | DNA computation platform for NSCLC diagnosis. a, The workflow for NSCLC diagnosis with the DNA computation platform: in silico training using publicly available miRNA-seq profiles of healthy and NSCLC individuals from TCGA (i) ; DNA implementation of multiplication, summation and subtraction (ii) and experimental validation on synthetic and clinical serum samples (iii). miRNA **a**–**d** are the miRNA inputs. W_1–W_4 are the weights associated to corresponding miRNA inputs. $\sum_1$ and $\sum_2$ are the mathematical weighted sum of corresponding miRNA inputs. **b**, Selected miRNA combinations and their associated weights for the classifier model. **c**, Performance of the in silico-trained classifier with data from the training set, in which 100 and 85.7% of NSCLC and healthy samples were classified correctly with an area under curve >0.99. The training set includes miRNA-seq data from 630 NSCLC and 63 healthy individuals in TCGA. **d**, Performance of the classifier with data from the validation set, where 99.6 and 85.2% of NSCLC and healthy samples were labelled correctly with area under curve of 0.99. The validation set includes miRNA-seq data from 270 lung cancer and 27 healthy individuals in TCGA.

operations performed over these inputs for optimum classification of healthy and lung cancer individuals. Next, the in silico-trained classifier is decoded to a computational scheme at a molecular level. We found that a DNA-based molecular computer with simplified winner-takes-all computation scheme can experimentally implement classifier trained in silico[28]. Finally, the performance of the DNA-based computation is validated on synthetic and clinical samples, which include miRNA amplification in serum, transformation of the amplified linear single-stranded DNA to circular DNA for more sequence orthogonality and DNA computation with weight multiplication, summation and subtraction followed by signal reporting.

In silico training for model construction

To build an effective classifier model, we selected the support vector machine (SVM) to train the data with the following conditions and steps[33]. First, publicly available miRNA expression data corresponding to 915 NSCLC and 105 healthy individuals from TCGA were used for differential expression analysis with the goal to classify the data into cancer and healthy groups with a desired accuracy. This initial analysis successfully identified 33 up-regulated and 22 down-regulated miRNA candidates with their stable expression levels in the NSCLC group four times more than those of the healthy group (Supplementary Fig. 1). The relatively large differences in miRNA levels between the two groups can ensure the sensitivity and specificity of the classifier by selecting only moderate numbers of inputs to prevent potential crosstalk between large numbers of probes and incorrect targets. Second, more constraints were applied to the 55 miRNA inputs to obtain a minimal set of inputs without compromising the classifier accuracy. For instance, we artificially limited the weights and mathematical operations to integers lower than ten and only to summation, multiplication and subtraction. We also applied a higher misclassification penalty score for the healthy individuals (false negative) than for the patients with lung cancer (false positive) with the intention to prevent false negative diagnosis that may delay treatment. It should be noted that we used the $\log_{10}$ to process the initial miRNA concentration data for the training to make the concentration data consistent with the mechanism of PCR amplification (Supplementary Note). Finally, different models were evaluated for the selection of optimum classification performance (for example, >95% accuracy). The selected model, including four miRNA inputs of miR-148a-3p, miR-182-5p, miR-30d-5p and miR-30a-3p, associated with positive and negative weights ranging from integer values of −4 to +4 can achieve a NSCLC identification sensitivity of 100% (95% confidence interval (CI): 99.2%, 100%), specificity of 85.7% (95% CI: 74.1%, 92.9%) and accuracy of 98.7% (95% CI: 97.5%, 99.4%) on the basis of the training set (630 lung cancer and 63 healthy samples) (Fig. 1b,c and Supplementary Figs. 2 and 3). We also validated the classifier model with 270 lung cancer and 27 healthy samples, and obtained a diagnosis sensitivity of 99.6% (95% CI: 97.6%, 99.9%), specificity of 85.2% (95% CI: 65.4%, 95.1%) and accuracy of 98.3% (95% CI: 96.0%, 99.4%) (Fig. 1d and Supplementary Fig. 3). In comparison, we found that using single miRNA from the dysregulated miRNA pool can only achieve 60–80% diagnostic accuracy for NSCLC in silico (Supplementary Fig. 4). Therefore, with this classifier that shows good sensitivity and specificity, along with simplified mathematical operations, it is feasible to achieve the molecular implementation with a set of carefully designed DNA probes.

miRNA amplification and transformation

One obstacle that limits the miRNA-based computation for clinical applications is the low concentration of miRNA (≤pM) in the

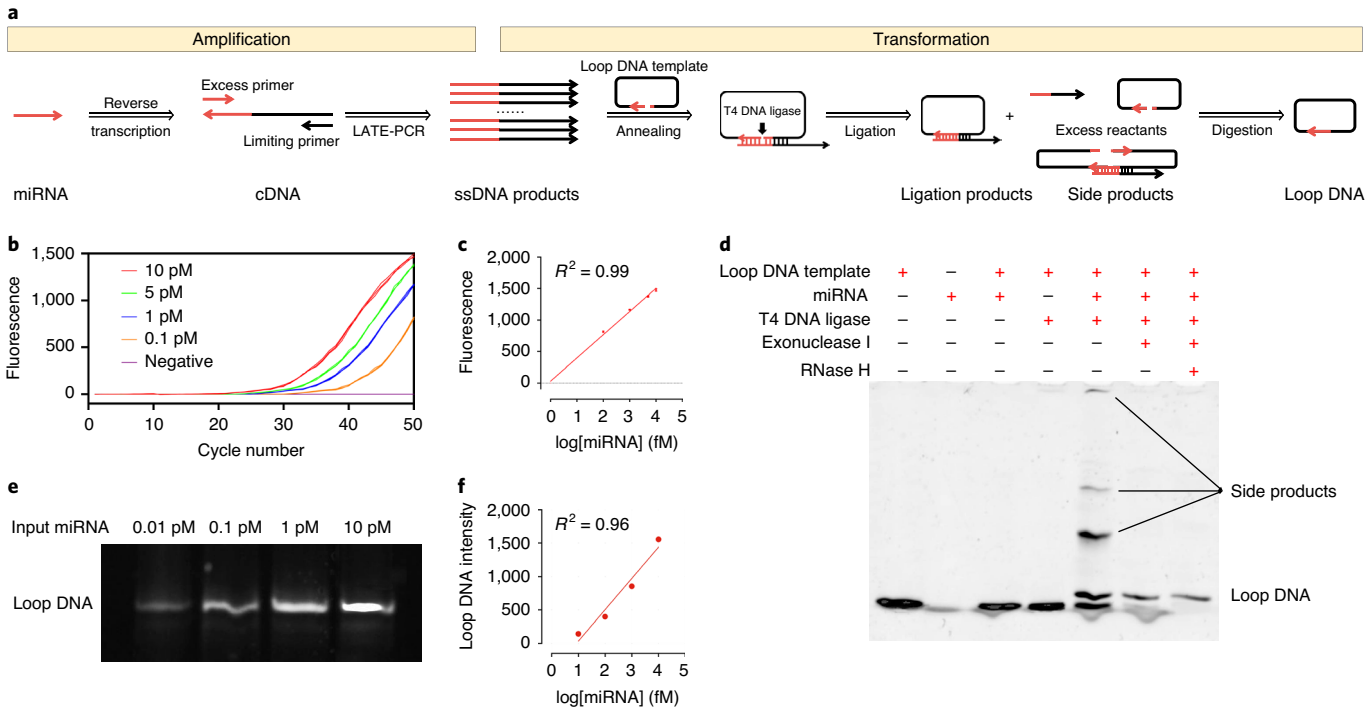

Fig. 2 | Amplification and transformation of miRNAs to loop DNAs. a, Schematic illustration of the amplification and transformation steps for miRNAs. In these two steps, miRNAs are amplified by LATE-PCR and transformed to loop DNA without perturbing the original variety and quantity information. **b**, Detection of LATE-PCR amplification products in samples containing different initial concentrations of miRNAs using a FAM-labelled TaqMan prob. Data are presented with three independent experiments. **c**, Plot of fluorescence of PCR at cycle 50 versus initial logarithmic miRNA concentrations demonstrating, the linear amplification behaviour of LATE-PCR in the concentration ranges of initial miRNA from fM to pM. $R^2 = 0.99$. Data are presented as mean ± s.d., $n = 3$ independent experiments. **d**, Denatured PAGE analysis of products after the transformation step with different treatment combinations. This experiment was repeated three times independently with the same result obtained and only one representative image is shown. **e**, Denatured PAGE analysis of loop DNA products amplified and transformed from different initial concentrations of miRNA. This experiment was repeated three times independently with the same result obtained and only one representative image is shown. **f**, Plots of fluorescence intensity of the loop DNA product bands versus initial logarithmic miRNA concentrations. $R^2 = 0.96$.

biological samples, such as serum, which is undetectable for the subsequent computation reactions. In addition, an amplification method with a linear response is necessary to maintain the initial ratios of each miRNA input so that the amplified sample can still possess the miRNA quantity information that is important for downstream classification analysis. Here, we used an improved asymmetric PCR, linear after the exponential PCR (LATE-PCR), to achieve close-to-linear amplification of miRNA, with respect to their logarithmic initial concentrations in a specific range (0.1–10 pM)[34]. Because the concentrations of miRNA inputs have been already processed with a logarithmic transformation during the in silico training, the LATE-PCR can maintain the initial miRNA quantity information after amplification if the concentrations are in the linear amplification range. As shown in Fig. 2a and Supplementary Fig. 5, the extracted miRNAs were first reversely transcribed into complementary DNA. We studied the RNA extraction efficiency using different commercial kits and found some statistically unneglectable variations in the miRNA distribution (Supplementary Fig. 6). However, this variation did not affect the diagnostic decision using our SVM-trained model (Supplementary Fig. 7). Furthermore, two different types of specially designed primer (excess primer and limiting primer) were used for LATE-PCR reactions. By adjusting the length and nucleotide composition of the limiting and excess primers to assure that $(T_m^{limiting} - T_m^{excess}) > 0$, together with the concentration ratios of those two primers, the exponential phase of the PCR reaction can switch to the linear phase shortly after the limiting primers are exhausted (fewer than ten cycles). This feature diminishes the concentration biases occurring at the exponential amplification step leading to amplified ssDNA products with quantitative ratios similar to those of the original logarithmic concentrations of the miRNA inputs (Supplementary Note)[34,35]. Meanwhile, by controlling the amplification cycle numbers, we can keep the amplified ssDNA products in a suitable concentration range, providing optimum concentrations for subsequent DNA computation (Supplementary Fig. 8). Figure 2b,c exhibits the results of amplified ssDNA products from a series of miRNA initial concentrations ranging from 0.1 to 10 pM using an optimized real-time LATE-PCR. We also combined four different miRNA inputs of miR-148a-3p, miR-182-5p, miR-30d-5p and miR-30a-3p with various initial concentrations in a mixture and amplified them by LATE-PCR. The results showed a linear relationship of the original logarithmic concentrations of the miRNAs and the final concentrations of cDNA products post amplification, confirming the feasibility of this method for simultaneous linear amplification of multiple inputs (Supplementary Figs. 9 and 10). Therefore, this LATE-PCR can help mitigate the bias associated with exponential amplification methods and convert the low concentrations of miRNA inputs (≤pM) to much higher concentrations of ssDNA (≥nM) without perturbing their original variety and quantity information.

Given that the amplified ssDNA products from miRNA inputs do not have sufficient sequence specificity for subsequent computation, we developed a strategy to transform an amplified ssDNA to a loop DNA with T4 ligase, as shown in Fig. 2a. This loop DNA can allow more sites for encoding orthogonal sequences for subsequent DNA computation without disturbing the concentration and sequence information of the transformed ssDNA. Because the

ssDNA templates and unsuccessful ligation products would remain after ligation, possibly causing interference in the subsequent reactions, we used Exonuclease I and RNase H to remove the non-loop DNA sequences post ligation. As shown in Fig. 2d, only pure loop DNA was observed in the products if samples were processed with enzymes. In addition, the final concentrations of loop DNA linearly correlated to the logarithms of the original miRNA concentrations, as shown in Fig. 2e,f. Therefore, in this step, ssDNA amplified by LATE-PCR can be transformed into loop DNA for subsequent DNA computations without disturbing their quantity ratios.

Molecular implementation of in silico model

To have simple (yes or no), but accurate, diagnostic results, a systematic implementation of the in silico-trained classifier with simplified winner-take-all computation scheme was designed, in which the output for a sample reports only the larger weighted sum of inputs of the same kinds[28]. Specifically, the mathematical weighted sum of healthy miRNA inputs is compared with that of cancer miRNA inputs, and only the larger is used to report the corresponding signals for diagnosis of either healthy or cancer. We simplified the mathematical operations of the in silico-trained classifier into three substeps, including multiplication, summation and subtraction, and used cascaded DNA strand displacement to implement the calculations at a molecular level[36].

First, different miRNA inputs undergo weighted multiplication with their initial concentrations ($W_n \times c(\text{miRNA}_n) = T_n$), where W_n is the predefined weight, $c(\text{miRNA}_n)$ is the miRNA concentration and T_n is the weighted multiplication, respectively. In the in silico model, we found that miR-148a-3p and miR-182-5p are positively correlated to NSCLC and associated with weights of 2 and 4, respectively; while miR-30d-5p and miR-30a-3p are negatively associated with NSCLC with weights of −1 and −3, respectively. It should be noted that the numerical weights associated with these four miRNAs are practically capturing their levels of significance for determination of the diagnostic results. Using the loop DNA products that are amplified and transformed from the miRNA inputs, we designed a set of multiplication probes (M probe) that can target different sequence regions within each loop DNA (A_n) to implement the weight (W_n) (Fig. 3a and Supplementary Fig. 11). For example, weights $W_n = 4$, 3, 2 or 1 are realized by having 4, 3, 2 or 1 distinct sequence regions that can be targeted by M probes in the same A_n. We used a fluorescence reporting scheme (M reporter in Supplementary Fig. 11) based on a strand displacement reaction to verify the multiplication calculation ($W_n \times c(A_n) = T_n$) by checking the concentrations of generated A_n–M probe complexes (T probe) (Supplementary Fig. 12). As shown in Fig. 3b, the steady-state concentrations of T probes are linearly proportional to the number of M probes bound to A_n of different concentrations, demonstrating the effective molecular implementation of the multiplication operations.

Second, the different weight multiplications (T_n) within the groups of healthy or cancer miRNA inputs are summed ($T_1 + T_2 = \text{E}$; $T_3 + T_4 = \text{F}$). This is achieved by adding a summation probe (S probe) that can react with the T probe resulting from the multiplication step (Fig. 3c). Specifically, this S probe can bind with the identical sequence regions in T probe through a DNA displacement reaction and convert the same type of signals to a common weighted-sum species (E or F probe). We also used a fluorescent reporter probe (S reporter) to test the summed concentrations of T_1 and T_2 (or T_3 and T_4) and found correct signals corresponding to the total concentrations of both complexes (Fig. 3d and Supplementary Figs. 13 and 14). Therefore, we believe this summation scheme can accurately determine the weighted sums of miRNA inputs of the same type.

Finally, the sums of two different types of input are compared by a subtraction operation for reporting the diagnostic results (E − F = diagnostic result, Fig. 3e). In this design, weighted sums of positive and negative output strands (E and F probe) obtained from the previous step can annihilate each other until only the main species is left based on a cooperative hybridization scheme[37,38]. One weighted-sum sequence (E probe) resulted from the previous summation step can bind to a toehold on one side of the annihilator probe (N probe) to reversibly form an intermediate complex, so that no molecules are consumed. However, if another weighted-sum strand (F probe) is also present, it can bind to the N probe on another toehold on the opposite side and break the N probe into two inert products that are not able to report any signal. In this way, the annihilation reaction allows effective subtraction of two weighted sums and exhausts the smaller number of the weighted sums resulting from previous multiplication and summation operations to leave only the larger value (Supplementary Fig. 15).

We used two distinct fluorescent reporters with FAM and ROX fluorophores to report the corresponding outputs associated with positive and negative weights after the subtraction operation. A catalytic entropy-driven amplification restores the offset signals and improves the reporting sensitivity (Supplementary Fig. 16). On reporter calibration, the fluorescence signal from the ROX reporter can be subtracted from the FAM reporter signal to obtain a normalized signal used for classification. If the normalized signal of FAM is observed, samples can be classified as having more positive weighted sums; otherwise, samples with greater ROX signals correspond to negative weighted sums. We verified this scheme by using 36 different weighted-sum (E and F probe) combinations with the initial concentrations ranging from 0 to 50 nM. Figure 3f shows that only the weighted sums with higher initial concentrations result in much higher signals, in agreement with the proposed design (Supplementary Fig. 17).

System validation using synthetic miRNAs

We tested the performance of entire system with samples containing synthetic miRNAs that match the sample profiles from the TCGA database (15 NSCLC and 15 healthy samples) (Fig. 4a). We replicated the 30 samples by adding synthetic miRNAs on the basis of their actual concentrations and ratios calculated from the miRNA-seq data in vitro. After amplification, transformation and DNA computation, fluorescence signals of FAM (NSCLC) and ROX (healthy) were measured to check the diagnostic results of these synthetic samples. Only one healthy sample was mis-diagnosed as lung cancer, and the remaining 15 NSCLC samples and 14 healthy samples were classified correctly with a sensitivity of 100% (95% CI: 74.7%, 100%), specificity of 93.3% (95% CI: 66.0%, 99.7%) and accuracy of 96.7% (95% CI: 81.9%, 99.9%), which are consistent with the results predicted by the in silico SVM model (Fig. 4b,c and Supplementary Fig. 18). We also tested four synthetic samples with four randomly mixed miRNA sequences as the negative control and did not observe obvious reporting fluorescence that can be used for diagnosis (Supplementary Fig. 19).

Lung cancer diagnosis using clinical samples

Finally, we tested our DNA computation-based diagnostic system with clinical serum samples. We studied two groups of serum samples (Supplementary Table 1): (1) serum samples from 14 patients with NSCLC from stage I to IV and (2) eight healthy individuals' serum samples. Using our previously described workflow, total miRNAs in the serum were first extracted from each sample, followed by LATE-PCR amplification. The amplified samples were further transformed to loop DNA and analysed by adding DNA computation probes together with FAM and ROX fluorescent reporters, corresponding to cancer and healthy results, respectively (Fig. 4d). The performance of our method in diagnosing NSCLC is shown in Fig. 4e and Supplementary Fig. 20. It can be seen that 13 out of 14 NSCLC samples were identified correctly with a sensitivity of 92.9% (95% CI: 64.2%, 99.6%), while two of eight healthy samples were mis-diagnosed to cancer with a specificity of 75.0%

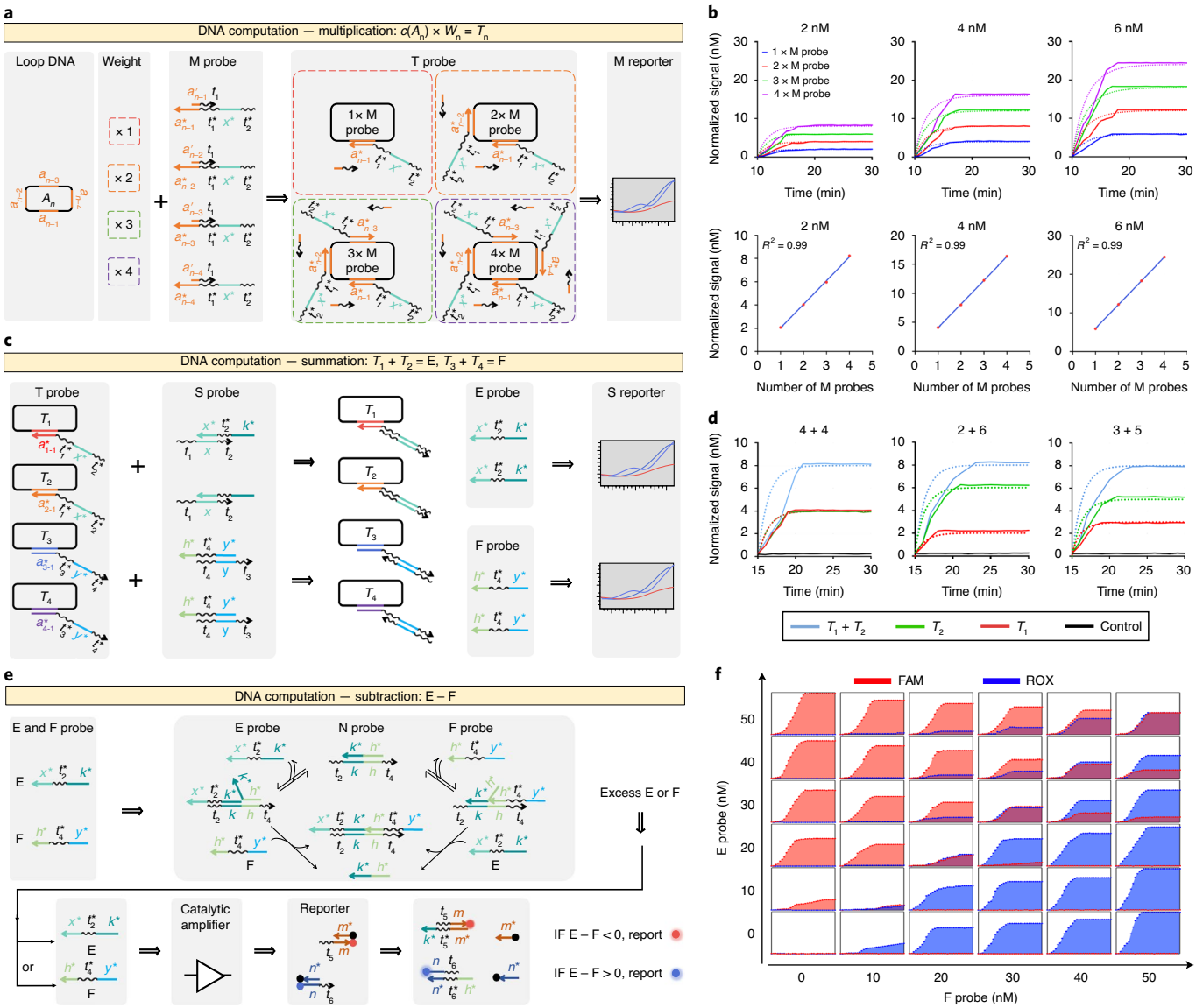

Fig. 3 | Workflow of the DNA computation. a, Scheme for multiplication ($W_n \times c(A_n) = T_n$). Each loop DNA (A_n) is targeted with a number of M probes equivalent to the weight for DNA multiplication. A fluorescent reporting scheme with M reporter is used to confirm the multiplication results of $W_n \times c(A_n) = T_n$ (1). Please refer to Supplementary Fig. 11 for details. **b**, Fluorescence reporting kinetics (solid lines) and simulations (dotted lines) of different concentrations of loop DNA (A_n) with different weights (W_n = 1, 2, 3 or 4). Linear relationships are obtained between the numbers of M probes and the steady-state fluorescence response. The reporter probes were added to the system 10 min after the recording started. Experiments were repeated three times independently. **c**, Scheme for summation ($T_1 + T_2 = E$ (2); $T_3 + T_4 = F$ (3)). Taking $T_1 + T_2 = E$ as an example, this operation is achieved by converting T_1 and T_2 into the same product E by interacting with the S probe through a DNA displacement reaction step. A fluorescent reporting scheme with S reporter is used to confirm the summation results of $T_1 + T_2 = E$. Please refer to Supplementary Fig. 13 for details. **d**, Fluorescence kinetics (solid lines) and simulations (dotted lines) of T_1, T_2 alone and their sums ($T_1 + T_2$) in the presence of equal concentrations of S probes and S reporter (that is, 4 + 4, 2 + 6 and 3 + 5). The reporter probes were added to the system 15 min after the recording started. Experiments were repeated three times independently. **e**, Scheme for subtraction (E – F = diagnostic result (4)). This operation is achieved by using an N probe that can annihilate the sums of different types of input. E and F individually bind preferentially to the N probe in a reversible manner. Equal amounts of E and F will be consumed by N probes and only the remaining type is reported by FAM (corresponding to E) or ROX (corresponding to F) fluorescence with an amplification step for sensitivity improvement. Please refer to Supplementary Fig. 16 for details of signal amplification step. **f**, Two-species winner-takes-all behaviour for different combinations of E and F. The *x* axis of each small graph is time (from 0–60 min) and the *y* axis is the normalized signal (from 0–100 nM). For detailed experimental conditions for all reactions shown in Fig. 3, please see Supplementary Table 3. Zigzag lines indicate toehold domains and straight lines indicate branch-migration domains in DNA strands, with arrowheads marking their 3′ ends. Toehold length of t_1, t_3, t_5 and t_6 is 6 nt, while the length of t_2 and t_4 is 8 nt. Fluorescence signal acquisition frequency is once per minute with the data normalization details described in the section of Methods.

(95% CI: 35.6%, 95.5%). The total accuracy of this method for NSCLC diagnosis in clinical samples was 86.4% (95% CI: 65.8%, 96.1%). It should be noted that the only mis-identified sample of NSCLC was from a patient diagnosed with stage I NSCLC by the pathological section, indicating that this method may be more sensitive to advanced stage NSCLC.

Accurate diagnosis of diseases with only a droplet of blood is always a goal for clinical medicine. While many pharmaceutical

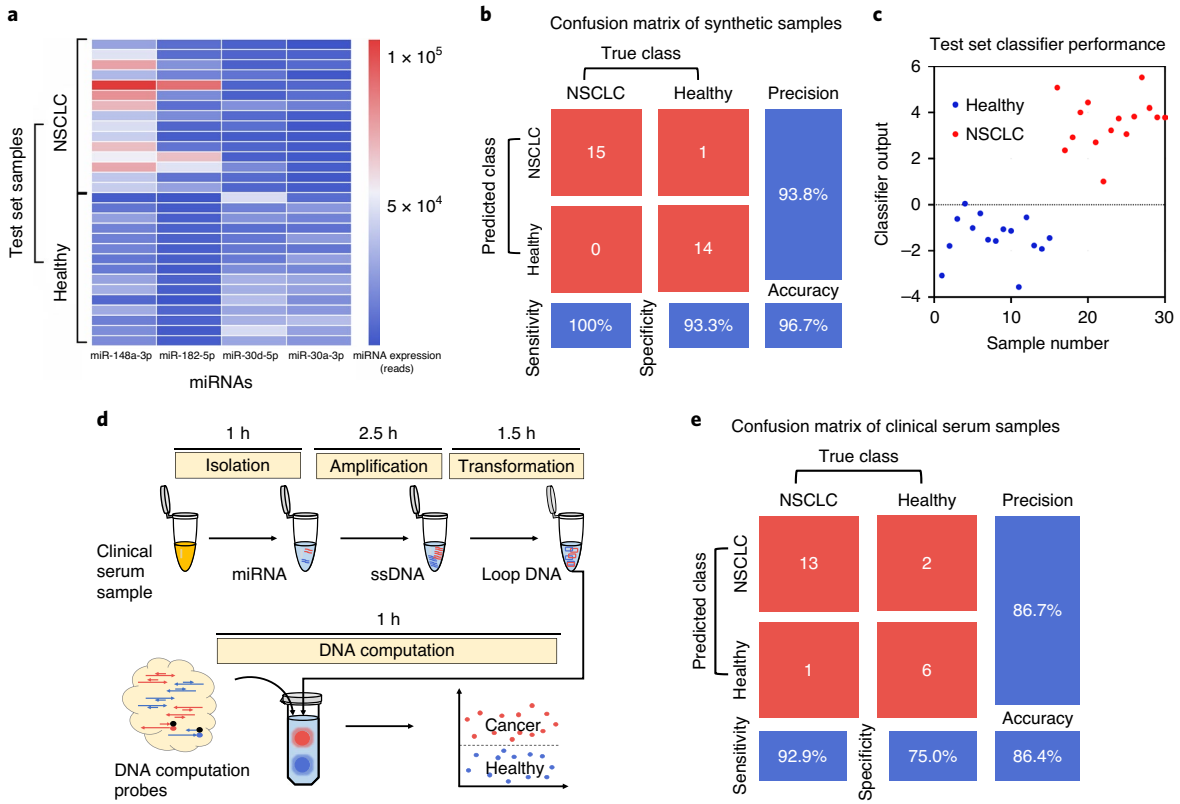

Fig. 4 | Validation of the DNA computation-based diagnostic system with synthetic and clinical samples. a, miRNA expression profiles of 15 lung cancer and 15 healthy samples from TCGA for classifier validation on synthetic samples. **b**, Confusion matrix analysis of the 30 synthetic samples with the DNA computation-based diagnostic system. According to the calculation of the confusion matrix data, we obtained an F1 score of 0.968 (calculated by F1 score $= \frac{2\times \text{precision}\times \text{sensitivity}}{\text{precision}+\text{sensitivity}}$) in NSCLC diagnosis, indicating the good performance of our approach in synthetic samples. **c**, Results of 30 samples predicted by in silico SVM model. **d**, Complete operational process using DNA computation-based system for NSCLC diagnosis with clinical serum samples. **e**, Confusion matrix analysis of 22 clinical samples with 14 NSCLC samples from NSCLC stages I to IV and eight healthy samples with their detailed information shown in Supplementary Table 1. It should be noted that the NSCLC samples were from the patients from the first diagnosis without any treatment. An F1 score of 0.897 indicated good performance of this DNA computation-based system for clinical diagnosis of NSCLC.

entities are investing large amounts of capital for developing automated instruments that couple advanced detection techniques and intelligent data analysis software, our approach provides an alternative route using traditional equipment together with DNA computation to realize rapid and accurate cancer diagnosis. This rationally designed approach combines an in silico classifier trained by publicly available data, a molecular implementation with a DNA-based reaction network and a practical workflow with sample amplification, transformation and computation on clinical samples. Using this approach, we can achieve NSCLC diagnosis using a 2-ml serum sample in 6 h with an accuracy of 86.4%, thus offering unprecedented potential for future clinical translation (Supplementary Table 2 for comparison with other diagnostic methods of NSCLC).

Several factors contribute to the successful applications of this DNA computation system for cancer diagnostics. First, the large amount of diagnostic information encoded in a relatively small numbers of miRNAs enables us to use a modest number of miRNAs to sensitively classify human cancers, while reducing experimental deviations resulting from potential crosstalk between large amounts of probes and incorrect targets. However, the sizes of the classifiers could potentially be further expanded to include more inputs for finer disease classification with sufficient sequence orthogonality. Second, the power of in silico data training together with the molecular implementation of a winner-takes-all computation scheme allows us to perform robust computing-based cancer diagnosis without human intervention. This in situ computation method reduces the experimental deviations caused by multiple-batches of sensing and analysis by the seamless integration of sensing and computing of multiple inputs into one batch operation. Finally, an effective pre-amplification step (LATE-PCR) in the workflow that maintains the original miRNA quantity information to the largest extent also contributes to the robust and sensitive reporting of the diagnostic results with human serum samples. Overall, the idea we used to program this autonomous molecular system provides immediate possibilities in performing in situ diagnosis with advantages such as speed, simplicity, low cost and less tendency for error[39,40].

Despite these promising results, this DNA computation system still needs optimization for clinical use. For instance, the existing method asks for a relatively complicated design scheme that may limit its broader applications in point-of-care diagnosis. A more programmable automation system for user-defined disease diagnostics needs to be further developed. Meanwhile, our system should be improved and validated using a large-scale clinical cohort involving more detailed information, so that its ability to perform early cancer diagnosis can be further explored.

Conclusions

In summary, we designed and tested a DNA computation-based cancer diagnostic method that can accurately differentiate NSCLC and healthy individuals from their serum samples. We envision that the power of DNA computing could inspire more clinical applications

towards low-cost, non-invasive and routine early cancer screening and classification, as well as monitoring cancer recurrence.

Online content

Any methods, additional references, Nature Research 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/s41565-020-0699-0.

Received: 20 January 2020; Accepted: 24 April 2020;
Published online: 25 May 2020

References

1. Allemani, C. et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* **391**, 1023–1075 (2018).
2. Rotunno, M. et al. A gene expression signature from peripheral whole blood for stage I lung adenocarcinoma. *Cancer Prev. Res.* **4**, 1599–1608 (2011).
3. Liu, C. et al. Low-cost thermophoretic profiling of extracellular-vesicle surface proteins for the early detection and classification of cancers. *Nat. Biomed. Eng.* **3**, 183–193 (2019).
4. Gines, G. et al. Isothermal digital detection of microRNAs using background-free molecular circuit. *Sci. Adv.* **6**, eaay5952 (2020).
5. Esquela-Kerscher, A. & Slack, F. J. Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer* **6**, 259–269 (2006).
6. Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. *Nat. Rev. Cancer* **6**, 857–866 (2006).
7. Lu, J. et al. MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).
8. Chen, X. et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* **18**, 997–1006 (2008).
9. Madhavan, D., Cuk, K., Burwinkel, B. & Yang, R. Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures. *Front. Genet.* **4**, 116 (2013).
10. Best, M. G. et al. RNA-seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. *Cancer Cell* **28**, 666–676 (2015).
11. Liu, Q. et al. DNA computing on surfaces. *Nature* **403**, 175–179 (2000).
12. Adleman, L. M. Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021–1024 (1994).
13. Braich, R. S., Chelyapov, N., Johnson, C., Rothemund, P. W. & Adleman, L. Solution of a 20-variable 3-SAT problem on a DNA computer. *Science* **296**, 499–502 (2002).
14. Han, D. et al. A cascade reaction network mimicking the basic functional steps of adaptive immune response. *Nat. Chem.* **7**, 835–841 (2015).
15. Srinivas, N., Parkin, J., Seelig, G., Winfree, E. & Soloveichik, D. Enzyme-free nucleic acid dynamical systems. *Science* **358**, eaal2052 (2017).
16. Green, A. A. et al. Complex cellular logic computation using ribocomputing devices. *Nature* **548**, 117–121 (2017).
17. Joesaar, A. et al. DNA-based communication in populations of synthetic protocells. *Nat. Nanotechnol.* **14**, 369–378 (2019).
18. Han, D. et al. Engineering a cell-surface aptamer circuit for targeted and amplified photodynamic cancer therapy. *ACS Nano* **7**, 2312–2319 (2013).
19. You, M., Zhu, G., Chen, T., Donovan, M. J. & Tan, W. Programmable and multiparameter DNA-based logic platform for cancer recognition and targeted therapy. *J. Am. Chem. Soc.* **137**, 667–674 (2015).
20. Rudchenko, M. et al. Autonomous molecular cascades for evaluation of cell surfaces. *Nat. Nanotechnol.* **8**, 580–586 (2013).
21. Chang, X. et al. Construction of a multiple-aptamer-based DNA logic device on live cell membranes via associative toehold activation for accurate cancer cell identification. *J. Am. Chem. Soc.* **141**, 12738–12743 (2019).
22. Douglas, S. M., Bachelet, I. & Church, G. M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831–834 (2012).
23. Thubagere, A. J. et al. A cargo-sorting DNA robot. *Science* **357**, eaan6558 (2017).
24. Pei, R., Matamoros, E., Liu, M., Stefanovic, D. & Stojanovic, M. N. Training a molecular automaton to play a game. *Nat. Nanotechnol.* **5**, 773–777 (2010).
25. Chao, J. et al. Solving mazes with single-molecule DNA navigators. *Nat. Mater.* **18**, 273–279 (2019).
26. Song, J. et al. Reconfiguration of DNA molecular arrays driven by information relay. *Science* **357**, eaan3377 (2017).
27. Qian, L., Winfree, E. & Bruck, J. Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368–372 (2011).
28. Cherry, K. M. & Qian, L. Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks. *Nature* **559**, 370–376 (2018).
29. Lopez, R., Wang, R. & Seelig, G. A molecular multi-gene classifier for disease diagnostics. *Nat. Chem.* **10**, 746–754 (2018).
30. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-free nucleic acid logic circuits. *Science* **314**, 1585–1588 (2006).
31. 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).
32. Koscielny, S. Why most gene expression signatures of tumors have not been useful in the clinic. *Sci. Transl. Med.* **2**, 14ps12 (2010).
33. Brown, M. P. et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl Acad. Sci. USA* **97**, 262–267 (2000).
34. Rice, J. E. et al. Monoplex/multiplex linear-after-the-exponential-PCR assays combined with PrimeSafe and Dilute-'N'-Go sequencing. *Nat. Protoc.* **2**, 2429–2438 (2007).
35. Pierce, K. E., Sanchez, J. A., Rice, J. E. & Wangh, L. J. Linear-after-the-exponential (LATE)-PCR: primer design criteria for high yields of specific single-stranded DNA and improved real-time detection. *Proc. Natl Acad. Sci. USA* **102**, 8609–8614 (2005).
36. Simmel, F. C., Yurke, B. & Singh, H. R. Principles and applications of nucleic acid strand displacement reactions. *Chem. Rev.* **119**, 6326–6369 (2019).
37. Zhang, D. Y. Cooperative hybridization of oligonucleotides. *J. Am. Chem. Soc.* **133**, 1077–1086 (2011).
38. Chen, Y. J. et al. Programmable chemical controllers made from DNA. *Nat. Nanotechnol.* **8**, 755–762 (2013).
39. Broza, Y. Y. et al. Disease detection with molecular biomarkers: from chemistry of body fluids to nature-inspired chemical sensors. *Chem. Rev.* **119**, 11761–11817 (2019).
40. Tregubov, A. A., Nikitin, P. I. & Nikitin, M. P. Advanced smart nanomaterials with integrated logic-gating and biocomputing: dawn of theranostic nanorobots. *Chem. Rev.* **118**, 10294–10348 (2018).

Methods

SVM training and validation. To train the SVM classifier, we obtained the miRNA-seq data of 915 patients with NSCLC and 105 healthy individuals from TCGA. Among them, 900 NSCLC and 90 healthy samples were randomly divided into a training set and a validation set with a 7:3 ratio for construction of in silico model, and the remaining 15 NSCLC and 15 healthy samples were saved for experimental testing with synthetic miRNA profiles replicated from those samples. Differential expression analysis was first used to identify the primary up and down-regulated miRNAs in NSCLC and healthy groups. Second, the selected expression data of miRNA was converted mathematically by changing all concentration values to their $\log_{10}$ values. A linear SVM classifier was then trained on the basis of these logarithmic data (training set including 630 NSCLC and 63 healthy samples) to classify the cancer and healthy groups using LIBSVM[41,42] by MATLAB R2018a. This classifier was validated using a validation set (including 270 NSCLC and 27 healthy samples).

DNA and RNA oligonucleotides and reagents. All DNA and RNA strands were prepared and purified (high-performance liquid chromatography) by Sangon. All DNA sequences are listed in Supplementary Table 4. DNA probes were suspended at 100 μM in 1× TE buffer and stored at 4 °C or −20 °C. RNA probes were stored in nuclease-free water at −80 °C until needed. All chemicals were of analytical grade and were purchased from Sigma-Aldrich. Buffers were prepared according to standard laboratory procedures.

DNA purification. All DNA duplexes were prepared by mixing top and bottom strands in a 1:1.2 ratio with specific concentrations detailed in Supplementary Table 3. Complexes were annealed in a thermal cycler (Bio-Rad) by heating to 98 °C for 5 min and then cooling to 25 °C at a rate of 1 °C per 1 min. Probes were purified using 12% polyacrylamide gel electrophoresis (PAGE). Double-stranded bands were cut out and eluted into 1× TE buffer with 12.5 mM Mg^{2+}. Concentrations were determined by NanoDrop (Thermo Fisher).

RNA extraction in serum samples. All serum samples were collected from Shanghai Jiao Tong University Affiliated Sixth People's Hospital (Shanghai, China) with informed consent. 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. Total RNA in serum samples was extracted using the miRNeasy Serum/Plasma Kit from Qiagen of Canada if not specifically mentioned. All samples were purified according to the manufacturer's instructions without recombinant DNase digest, and the eluted RNAs were stored in nuclease-free water at −80 °C until needed. For the studies of RNA extraction efficiency, we also used the RNA isolation kit (Direct-zol RNA MiniPrep kit R2050) from Zymo Research as a reference.

LATE-PCR and its efficiency. Synthetic or extracted miRNAs were first reverse transcribed into cDNA by using a reverse transcription kit in a 20-μl system including reverse transcription enzyme and reaction buffer at 37 °C for 60 min on a Roche Light Cycler 480 Real-Time PCR System using 96-well plates. Please refer to Supplementary Fig. 5 for the mechanism of reverse transcription. LATE-PCR reactions were performed in a 20-μl system including template cDNA, excess primer, limiting primer, TaqMan probe, Taq DNA polymerase and PCR reaction buffer (Sangon). The LATE-PCR efficiency was studied by spiking four synthetic miRNAs (miR-148a-3p, miR-182-5p, miR-30d-5p and miR-30a-3p) of 10 pM each to four separated PCR tubes and added the four corresponding primer combinations (limiting and excess primer), enzymes and reporting probes (Taqman). The amplification efficiencies were studied by comparing the amplification curve and fluorescence intensities at the cycle number of 50. Please refer to Supplementary Table 3 for detailed experimental conditions.

Ligation and enzymatic digestion. LATE-PCR amplicons were mixed with loop DNA templates for annealing and ligation using a T4 DNA Ligase Kit (New England BioLabs) according to the manufacturer's instructions. We then processed the ligation products with Exonuclease I and RNase H (New England Biolabs) to eliminate excess amplicons, loop DNA templates and other side products from the ligation step, leaving only the loop DNA in the system for DNA computation. Please refer to Supplementary Table 3 for the concentrations and other experimental conditions used for ligation and enzymatic digestion.

Fluorescence kinetic measurements. Fluorescence kinetics data were collected by using a HORIBA FluoroMax 4 spectrofluorometer for single measurement and a Synergy H1MF Hybrid Multi-Mode Microplate Reader (BioTek) for high-throughput measurement. The reactions (see Supplementary Table 3 for concentrations) were carried out in 1× TE buffer with 12.5 mM Mg^{2+}.

Data normalization. Arbitrary fluorescence units were normalized to concentrations using a standard curve for each reporter complex. To generate a standard curve, annealed reporter complex was suspended in 1× TE buffer with 12.5 mM Mg^{2+}, and an initial baseline fluorescence signal was recorded. This was followed by addition of a series of known concentrations of reporter initiator strands. The linear standard curve was constructed using the steady-state fluorescence of different concentrations of reporters.

Reporting Summary. Further information on research design is available in the Nature Research 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-seq data used in this study are available on TCGA database https://portal.gdc.cancer.gov.

Code availability

The SVM training and validation code used in this study is from ref. [41] and available on https://www.csie.ntu.edu.tw/~cjlin/libsvm/.

References

41. Chang, C.-C. & Lin, C.-J. LIBSVM: a library for support vector machines. *ACM T. Intel. Syst. Tec.* **2**, 27 (2011).
42. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (grant nos. 21974087 and 31871009), Shanghai Municipal Education Commission—Gaofeng Clinical Medicine Grant Support (no. 20181709), Innovative Research Team of High-Level Local Universities in Shanghai, Faculty Start-up Funding Support from the Institute of Molecular Medicine of Shanghai Jiao Tong University and the Recruitment Programme of Global Youth Experts of China. We thank M. Zhang for the help on constructing kinetic models and data simulations. We thank J. Sun for the helpful discussion.

Author contributions

C.Z. and D.H. conceived and designed the experiments. C.Z. carried out the assays and analysed the results. C.Z., Y.Z., X.X., H.L., R.X., Y.M. and D.H. supported the optimization of assays and prepared the data. X.T. and Y.D. collected the specimens. D.H., H.-C.L. and C.Z. wrote the manuscript. D.H. supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41565-020-0699-0.

Correspondence and requests for materials should be addressed to D.H.

Peer review information *Nature Nanotechnology* thanks Tom de Greef 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.

nature research

Corresponding author(s): Da Han

Last updated by author(s): Apr 10, 2020

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	
☐	☒	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	The Support Vector Machines (SVM) training and validation code used in this study is available on https://www.csie.ntu.edu.tw/~cjlin/libsvm/. The miRNA-seq data used in this study are available on TCGA database https://portal.gdc.cancer.gov.
Data analysis	The SVM training and validation was done using Matlab_R2018a. Data analysis was done using IBM SPSS Statistics V26 and Prism 8.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on 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-seq data used in this study are available on TCGA database https://portal.gdc.cancer.gov.

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 statistical methods were used to predetermine sample size. Sample size was determined by the number of cases available in the TCGA database mined. For the in silica SVM training and validation, we obtained all the available miRNA-seq data from TCGA database with a total of 915 NSCLC and 105 healthy samples. Among them, 900 NSCLC and 90 healthy samples were randomly divided into a training set and a validation set with a 7:3 ratio for construction of in silica model, and the remaining 15 NSCLC and 15 healthy samples were saved for experimental testing with synthetic miRNA profiles replicated from those samples. For the evaluation of diagnostic accuracy of the DNA computation-based method, we used 22 clinical serum samples from 8 healthy and 14 NSCLC individuals. Herein, the sample size was largely determined by sample availability.
Data exclusions	No data was excluded from the analysis.
Replication	We replicated mi RNA extraction, PCR, and fluorescent reporting at least 3 times across different conditions. We did not find any case where the data was not able to be replicated.
Randomization	NSCLC patients and healthy controls were enrolled in the study based on either confirmed diagnosis of non-small-cell lung cancer or health screening reports with no issue of the lungs. Therefore, randomization of individuals in different groups is not applicable.
Blinding	No blinding was performed. For training and validation of SVM models, we obtained all the available miRNA-seq data from TCGA database . For the clinical samples, we collected known types of samples from hospital according to the experimental needs to verify the accuracy of DNA molecular computation.

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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	Study participants include non-small cell lung cancer (NSCLC) patients and healthy individuals. NSCLC samples were from the patients of the first diagnosis without any treatment. Healthy samples were from individuals with no issue of lungs reported in physical examination. screening reports The age, gender and cancer stage of all samples were not limited. They were collected by the hospital within last one year. Detailed clinical sample information is provided in Supplementary Table 1.
Recruitment	All subjects were enrolled with Institutional Review Board-approved protocols and all serum samples were collected with informed consent. NSCLC patients were recruited based on clinical diagnosis of NSCLC with confirmed pathological biopsy results. Healthy individuals were recruited based on their health screening reports with no issue of the lungs. In order to avoid bias that may affect the results, all NSCLC samples were taken from patients who have not yet undergone any medical treatment.
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.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com