DNA NANOTECHNOLOGY

DNA computation improves diagnostic workflows

A molecular reaction network translated from a computer-trained classifier can distinguish lung cancer patients from healthy individuals based on specific microRNAs in the blood.

Pepijn G. Moerman and Rebecca Schulman

DNA is more than just a carrier of genetic information; it can also be used as a building block for molecular computers. In a molecular AND circuit, for example, the logic gate can consist of a partially hybridized DNA complex, made of an output strand sequestered by two sacrificial strands[1]. The inputs are single-stranded DNA oligomers with specific sequences, complementary to the sacrificial strands. Only when both inputs are present, they can bind to the sacrificial strands to release the output strand from the complex; if one input is missing, the reaction cannot proceed. Related principles have also been used to realize circuits that perform other logic operations and even cascades that perform mathematical operations[2] or that can classify complex input mixtures[3].

Computations performed by DNA molecules are not nearly as fast, robust or reproducible as those conducted by electronic chips. Therefore, our laptops will not be replaced by vials of DNA; however, DNA computers have the advantage that they can be easily interfaced with cells and body fluids, which makes them promising diagnostic tools. Computer chips manipulate electric pulses; by contrast, molecular computations act on small strands of DNA or RNA. Often these DNA molecules are lab-made strands with artificial sequences, but they could be viral DNA or microRNA (miRNA) molecules — small RNAs that regulate cellular gene expression.

In 2004, Benenson et al. demonstrated a molecular computation that could recognize specific RNA molecules in vitro and subsequently produce a DNA signal[4]. The authors suggested that such a nanoscopic DNA computer could recognize viral RNA and produce miRNA that activates cellular defence pathways, which would make DNA computations promising tools for at-home or point-of-care diagnosis.

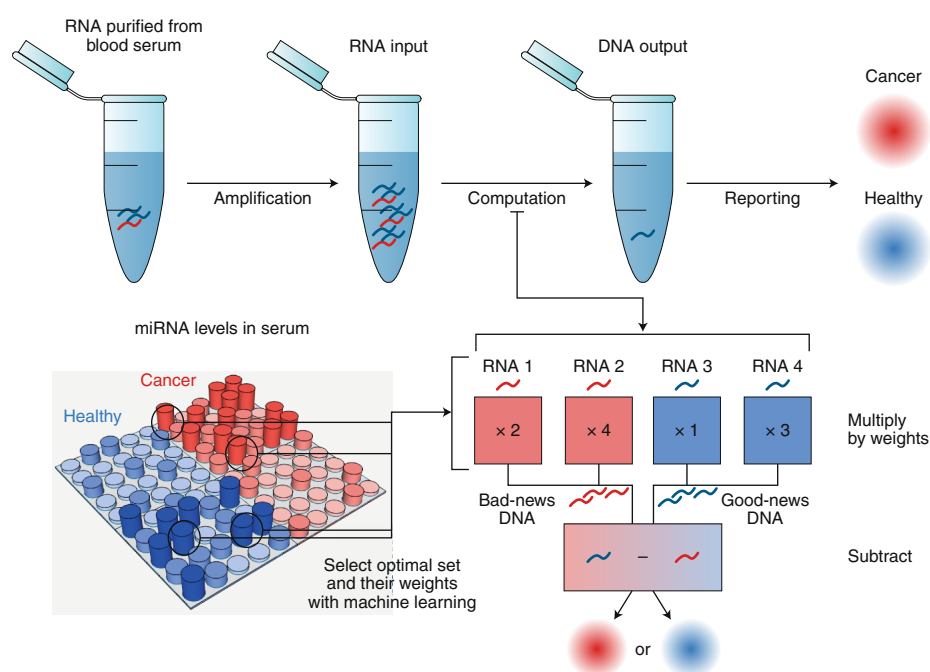Such autonomous diagnostic tools may seem futuristic; however, viral infections

Fig. 1 | Diagnostic workflow for lung cancer diagnosis. The workflow includes purification of RNA from blood serum, amplification of RNA, DNA computation and reporting. The DNA computation takes the concentration of 4 miRNAs as an input, multiplies them with their respective weights, adds them, and subtracts the ones that signal good news from the ones that signal bad news. A computer-based classifier was used to select the optimal set of 4 miRNAs and their weights. Three-dimensional graph (bottom left) reproduced with permission from ref. [7], Springer Nature Ltd.

can already be diagnosed using DNA circuit-based kits; for example, implemented in a paper-based kit for the diagnosis of a Zika virus infection in blood serum[5]. First, RNA from the serum is amplified increasing picomolar RNA concentrations to nanomolar levels, which gives enough signal to be measured, followed by an enzymatic reaction leading to a colour change of the paper when Zika virus RNA is present. This workflow can be applied for viral infections that are readily recognized by the presence of one foreign RNA molecule — that of the virus— but most disease diagnoses

require the recognition of a pattern of many miRNAs, whose concentrations are often only slightly increased or decreased compared to healthy individuals. For example, to diagnose cancer, relative concentrations of multiple miRNAs must be measured for reliable diagnosis. This is where DNA computations come in; Lopez et al. showed that a DNA classifier can pick out RNA concentration patterns indicative of cancer from a slew of miRNAs that mimic those that would be present a blood sample[6]. An output DNA signal is only produced when miRNAs were present in a particular

ratio that is uncommon in healthy cells but typical in cancer cells.

Now, writing in *Nature Nanotechnology*, Chao Zhang et al. have — for the first time — integrated a DNA classifier that recognizes biological miRNA patterns into a complete diagnostic workflow[7]. The researchers were able to diagnose non-small cell lung cancer based on four miRNAs present at picomolar concentrations in blood serum. The procedure takes approximately 6 hours and is 86.4% accurate.

The authors first identified a set of miRNAs that could act as a fingerprint for the recognition of lung cancer. Next, they used publicly available clinical data on the concentrations of these miRNAs in healthy individuals and lung cancer patients to train a computer-based classifier that distinguishes between the two groups. The training process identified an optimal set of four miRNA markers for lung cancer and the relative importance of each miRNA, which was captured in a weight, W. Two miRNAs have positive differential expression in lung cancer patients compared to healthy patients; they signal an increased likelihood of cancer. And two miRNAs have negative differential expression and thus, high concentrations suggest a patient is healthy.

The researchers then translated the computer-trained classifier into a DNA computation that takes the concentrations of the four selected miRNAs as input. The molecular classifier consists of two consecutive computations; first, each input signal i is multiplied by a weight W_i in a process, in which the miRNA molecules interact with a DNA complex such that W_i DNA strands are released for each RNA strand that comes in. The two signals indicating cancer are then subtracted from the two signals indicating health. Pairs of good-news and bad-news DNA strands annihilate and form a waste product; whichever strand is in excess reacts with a fluorescent marker and a signal can be observed (Fig. 1).

These computations require miRNA concentrations three orders of magnitude higher than physiological levels; thus, miRNA levels in blood serum first need to be amplified, which poses a problem, because classification based on relative miRNA concentrations only works if ratios are preserved. RNA amplification is typically done using polymerase chain reaction (PCR), which exponentially increases the concentration of each RNA molecule and distorts their ratios. To resolve this issue, Zhang et al. trained their classifier to act on the logarithm of the RNA concentration rather than on the actual value. Thereby, the weights in the classifier correct for the inherent bias of the amplification step.

Integration of DNA computation into diagnostic tools expedites the interpretation of miRNA levels in blood; however, such tools require specialized equipment and a trained user for sample preparation, RNA amplification and signal readout. Optimization of other steps in the diagnostic pathway could reduce this problem. For example, the produced signal could be a colour change, which is observable by the naked eye. Moreover, a computation process could not only produce a yes-or-no answer, but could distinguish between different types of cancer, producing signals in different colours. Diagnostic kits may eventually even integrate automated RNA purification and amplification. The proof-of-concept diagnostic workflow reported by Zhang et al. is an important step toward developing such point-of-care kits, combining all necessary components — computing, amplification, reporting and machine learning — into a single diagnostic pathway and demonstrating that they can be interfaced with each other. ❐

Pepijn G. Moerman and Rebecca Schulman [ID] ✉

Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA.
✉*e-mail: rschulm3@jhu.edu*

References
1. Zhang, D. Y. & Seelig, G. *Nat. Chem.* **3**, 103–113 (2011).
2. Qian, L. & Winfree, E. *Science* **332**, 1196–1201 (2011).
3. Qian, L., Winfree, E. & Bruck, J. *Nature* **475**, 368–372 (2011).
4. Benenson, Y., Gil, B., Ben-Dor, U., Adar, R. & Shapiro, E. *Nature* **429**, 423–429 (2004).
5. Pardee, K. et al. *Cell* **165**, 1255–1266 (2016).
6. Lopez, R., Wang, R. & Seelig, G. *Nat. Chem.* **10**, 746–754 (2018).
7. Zhang, C. et al. *Nat. Nanotechnol.* https://doi.org/10.1038/s41565-020-0699-0 (2020).