BioChip J. (2013) 7(1): 29-37
DOI 10.1007/s13206-013-7105-z

Original Research

Rule-based *in vitro* molecular classification and visualization

Soo-Yong Shin[1,†], Kyung-Ae Yang[2,†], In-Hee Lee[2], Seung Hwan Lee[3], Tai Hyun Park[3] & Byoung-Tak Zhang[2]

Received: 15 November 2012 / Accepted: 17 January 2013 / Published online: 20 March 2013

Abstract Molecular computing using programmable nucleic acids has been attracting attention for use in autonomous sensing systems and information processing systems by interacting with a biological environment. Here, we introduce a rule-based *in vitro* molecular classification system that can classify disease patterns using several microRNA (miRNA) markers via the assembly of programmed DNA strands. The classification rules were derived by analyzing large-scale miRNA expression data obtained from a public database, and the identified rules were converted into DNA sequences. Classification was performed via the detection of miRNA markers in the rules. The classification results were reported as a binary output pattern according to their hybridization to the rule sequences, which can be conveniently visualized using gold nanoparticle aggregation. Our results demonstrate the utility of *in vitro* molecular classification by illustrating one of the ways in which molecular computing can be used in future biological and medical applications.

Keywords: *In vitro* classification, Molecular classification, DNA computing, Nanoparticle self-assembly, Rule-based system

[1]Department of Biomedical Informatics, Asan Medical Center, and University of Ulsan College of Medicine, 43 Olympicro, Songpa-gu, Seoul 138-736, Korea
[2]Department of Computer Science and Engineering, Seoul National University, 599 Gwanangro, Gwanak-gu, Seoul 151-742, Korea
[3]School of Chemical and Biological Engineering, Bio-MAX Institute, Seoul National University, 599 Gwanangro, Gwanak-gu, Seoul, 151-742, Korea
[†]These authors contributed equally to this work.
Correspondence and requests for materials should be addressed to B.-T. Zhang (✉btzhang@snu.ac.kr)

Introduction

DNA is a macromolecule that can be easily engineered and has the unique ability to carry information as well as the specificity of Watson-Crick base pairing. The DNA computing field has been spawned based on these features[1] and has received considerable attention as a bridge technology between information processing, materials science and biological systems[2-4].

Following pioneering work accomplished by a computer scientist demonstrating silicomimetic implementation using standard molecular techniques and encoded DNAs[1], early DNA computing work focused primarily on tackling computationally difficult problems by using the massive parallelism of DNA computing[1-3,5]. However, because the kinetics and fidelity of hybridization place inherent limitations upon the size and complexity of computing, the idea of parallel computing by DNA molecules has not yet been fully exploited. Alternatively, DNA-based computing may be useful in other promising areas such as biotechnology and nanotechnology, in which rationally designed DNA systems can perform information processing autonomously *in vitro* or *in vivo* with minimal human interference[6-17]. Programmed DNA molecules implemented a computing algorithm by releasing a piece of DNA designed to act as a drug depending on the expression of particular genes[6,7]. Several DNA computing algorithms carried out information processing based on the detection of microRNAs[8,9]. Moreover, systemically engineered DNA-based logic gates implemented network communication with higher complexity[10,11] and performed a silicomimetic games[12-14]. Additionally, a self-managed DNA robot has been developed[15-17].

Previous studies have indicated that a carefully designed DNA system can store information and execute

an algorithm in fluid autonomously; thus, such a system can be very useful for intelligent molecular sensing applications (e.g., disease diagnosis, biological sample analysis or pattern recognition), handling multiple pieces of target information directly and integrating the sensing procedure. Therefore, DNA computing can lead to intelligent and efficient diagnosis methods in medical applications.

In this paper, we describe the implementation of a DNA-based systematic analysis method using a rule-based classification algorithm. The proposed method can be applied for the classification of microRNA (miRNA)-dependent cancer types, is capable of classifying molecular patterns *in vitro* and visualizing the classification results through visible color changes via gold nanoparticle aggregation in a simple way.

Results and Discussion

Verification of the proposed molecular classification method

First, we verified our method using simple test rules consisting of one variable (X=T and X=F). The test scenarios are shown in Figure 1A, and the sequence information is provided in Table 1. Four test scenarios correspond to potential situations during hybridization between rule sequences and input molecules. Case 1 occurs when a rule is X=T and the actual value for X was T (present). In this case, the input satisfies the rule, and gold nanoparticles (Au-NPs) attached to the rule sequences aggregate showing a color change (Figure 1B). Case 2 occurs when the rule is X=T but the actual value for X was F (absent). In this case, the rule is not satisfied by the input and no aggregation is observed. Case 3 occurs when the rule is X=F and

Table 1. Sequence sets that demonstrate the concept of molecular classification.

Rule sequences (5′→3′)		
S	S1	GTTGCCAT-thiol (C3)
	S2	GTTCGTTT-thiol (C3)
	S3	AAACGAACGAACGACT
Rule (X=T)	L1	ATGGCAACTTACAGC
	L2	GTTGGTACAGTCGTTC
Rule (X=F)	$\overline{\text{Input}}$	TTACAGCGTTGGTAC
	L1	ATGGCAACGTACCAA
	L2	CGCTGTAAAGTCGTTC
Input data sequences (5′→3′)		
Input		GTACCAACGCTGTAA

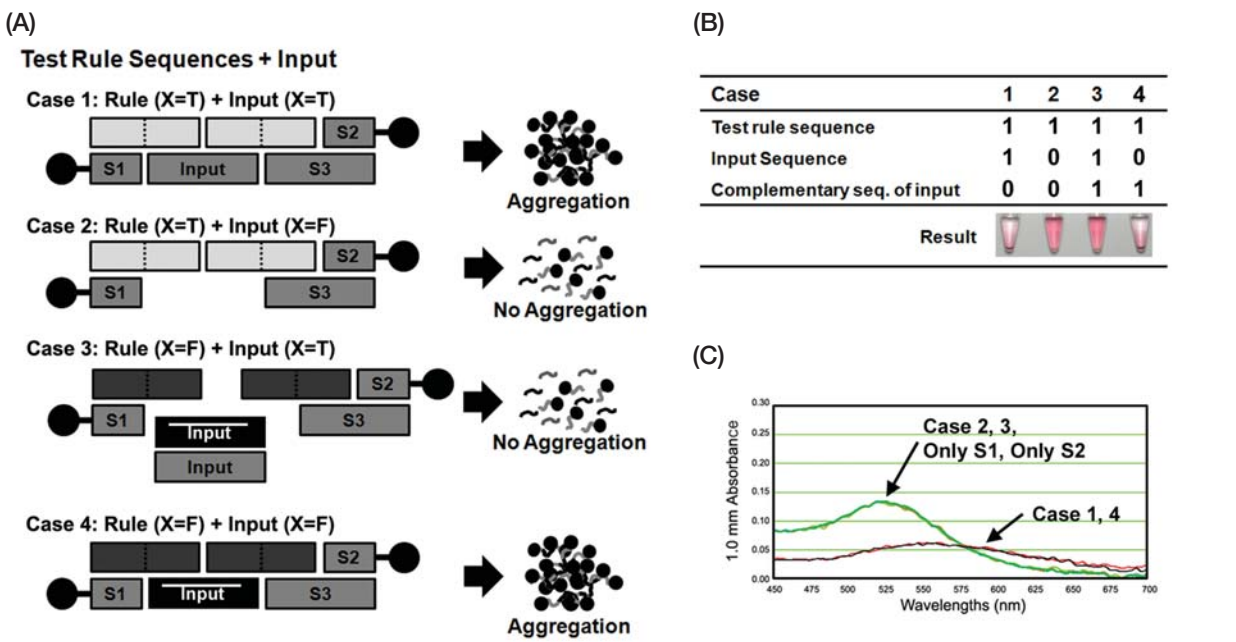

Figure 1. Verification of the proposed classifier. (A) Test cases. Case 1 occurs when input X=T (input is present) and the rule is X=T. Case 2 occurs when input X=F (input is absent) and the rule is X=T. Case 3 occurs when input X=T and the rule is X=F. Case 4 occurs when input X=F and the rule is X=F. (B) Test results. Cases 1 and 4 show the color changes as expected. (C) UV/Vis spectrometry results. The characteristic plasmon peak of cases 1 and 4 disappeared at approximately 520 nm, which indicates that the Au-NPs aggregated. Thus, the input of case 1 satisfies the rule of case 1. Also, the input of case 4 satisfied the rule of case 4.

the actual value for X was T (present). Because X=F, a complementary sequence of input ($\overline{\text{input}}$) is included in the rule sequences as explained in Material and Methods Section. Therefore, perfect double strands cannot be formed (no aggregation is observed), which complies with the fact that the input does not satisfy the rule. Case 4 occurs when the rule is X=F and the actual value for X was F. Since the input satisfies the rule of case 4, the rule sequences hybridize each other to from the perfect double strands for a color change. These results confirm that our approach can successfully function as designed. Figure 1C depicts the UV/Vis spectrometry results of Figure 1B. The characteristic plasmon peak at approximately 520 nm disappeared only when Au-NPs aggregated, implying that Figure 1B and 1C show essentially the same results. Therefore, the classification result (perfect double strands or not) is observable via Au-NP aggregation, which is easily observed with the naked eye without any additional devices[9].

To demonstrate the possibility that this method can be scaled up, we designed a decision tree with five nodes to perform *in vitro* classification. The test decision tree is shown in Figure 2A. This particular decision tree has two rules: The first rule is "if (A=T) AND (B=T) AND (C=F) then Abnormal". The second rule

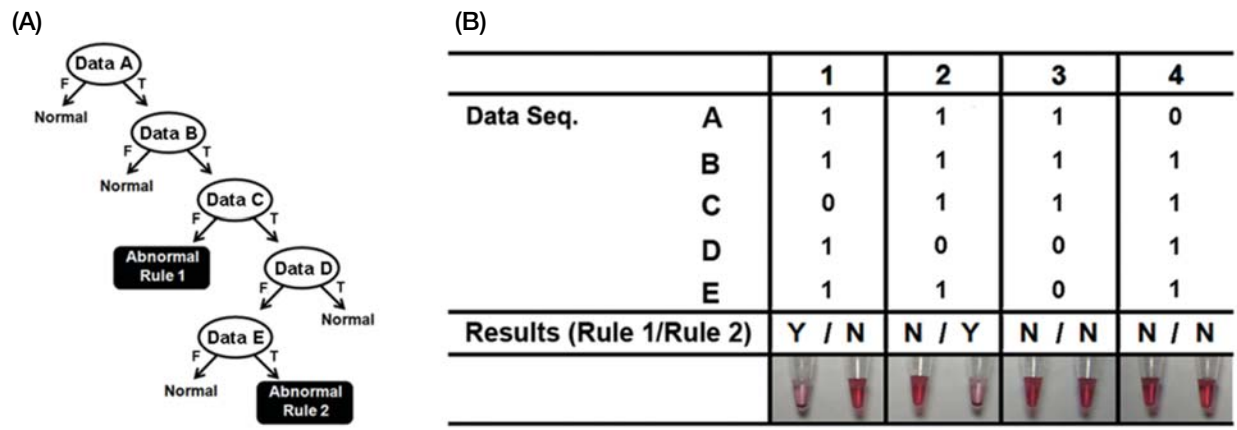

Figure 2. (A) Test scenario of the decision tree for scaling up the molecular classification protocol. (B) Data sequence for checking various cases. The test results were based on Au-NP aggregation in each case. Y or N in the "Results" row indicates "aggregation" or "no aggregation", respectively. Case 1 represents an abnormal case for Rule 1. Case 2 is an abnormal case for Rule 2. Cases 3 and 4 represent normal samples.

Table 2. Sequence sets for scaling up the molecular classification protocol. The same S sequences are used (as explained in Materials and Methods).

Rule detecting sequences (5′→3′)			
Rule 1		Rule 2	
Rule1 L1	ATGGCAACTTACAGC	Rule2 L1	ATGGCAACTTACAGC
Rule1 L2	GTTGGTACCGGTAGT	Rule2 L2	GTTGGTACCGGTAGT
Rule1 L3	CGTGATGACAACACA	Rule2 L3	CGTGATGACGTATGC
Rule1 L4	GGCATACGAGTCGTTC	Rule2 L4	CTGTGTTGTTCAGACT
S1	GTTGCCAT-thiol (C3)	Rule2 L5	GCCGTACACGTCTG
S2	GTTCGTTT-thiol (C3)	Rule2 L6	ATTCTGCCAGTCGTTC
S3	AAACGAACGAACGACT	S1	GTTGCCAT-thiol (C3)
$\overline{\text{Input C}}$	CGTATGCCTGTGTTG	S2	GTTCGTTT-thiol (C3)
		S3	AAACGAACGAACGACT
		$\overline{\text{Input D}}$	GTACGGCAGTCTGAA
Input data sequences (5′→3′)			
Input A	GTACCAACGCTGTAA		
Input B	TCATCACGACTACCG		
Input C	CAACACAGGCATACG		
Input D	TTCAGACTGCCGTAC		
Input E	GGCAGAATCAGACGT		

is "if (A=T) AND (B=T) AND (C=T) AND (D=F) AND (E=T) then Abnormal". Each rule has a DNA sequence to classify the result DNA sequences is provided in Table 2. The two rules were tested in different tubes for parallel testing. Figure 2B shows the classification results of the test samples. Case 1 represents an abnormal sample satisfying Rule 1. Therefore, the Rule 1 tube (left tube in the second column of Figure 2B) shows Au-NP aggregation whereas the Rule 2 tube (right tube in Figure 2B) does not. Similarly, case 2, which is the abnormal case that satisfies Rule 2, shows aggregation results only in the Rule 2 tube. Cases 3 and 4 are normal cases. Case 3 has only inputs A, B, and C, while D and E are missing. This case does not satisfy either of the rules because of the existence of input C (violates Rule 1) and the absence of input E (violates Rule 2). Likewise, case 4 does not satisfy either of the rules because input A has value 'F' (absent). In both cases, no Au-NP aggregation is shown in either the Rule 1 or Rule 2 tubes.

Classification of miRNA-related cancer types

To demonstrate the possibility of disease diagnosis, this rule-based pattern classification method was used for the classification of miRNA-related cancer types. The miRNA expression profile data for human cancer consist of 89 samples from human cancer patients and non-patients, and each sample contains the expression levels of 151 miRNAs[18]. To simplify the learning process, we transformed the expression levels into binary values (0 or 1) based on median expression levels. Among 151 miRNAs, 22 miRNAs were selected by information gain, and 12 miRNAs were further removed due to self-dimer formation. The final classification rules were constructed by repeatedly training a decision tree with different subsets of miRNAs, and the chosen decision tree showed an accuracy of 80%. Although we were capable of building more complicated and accurate decision trees, we chose the above tree to demonstrate our approach (Figure 3). The classification

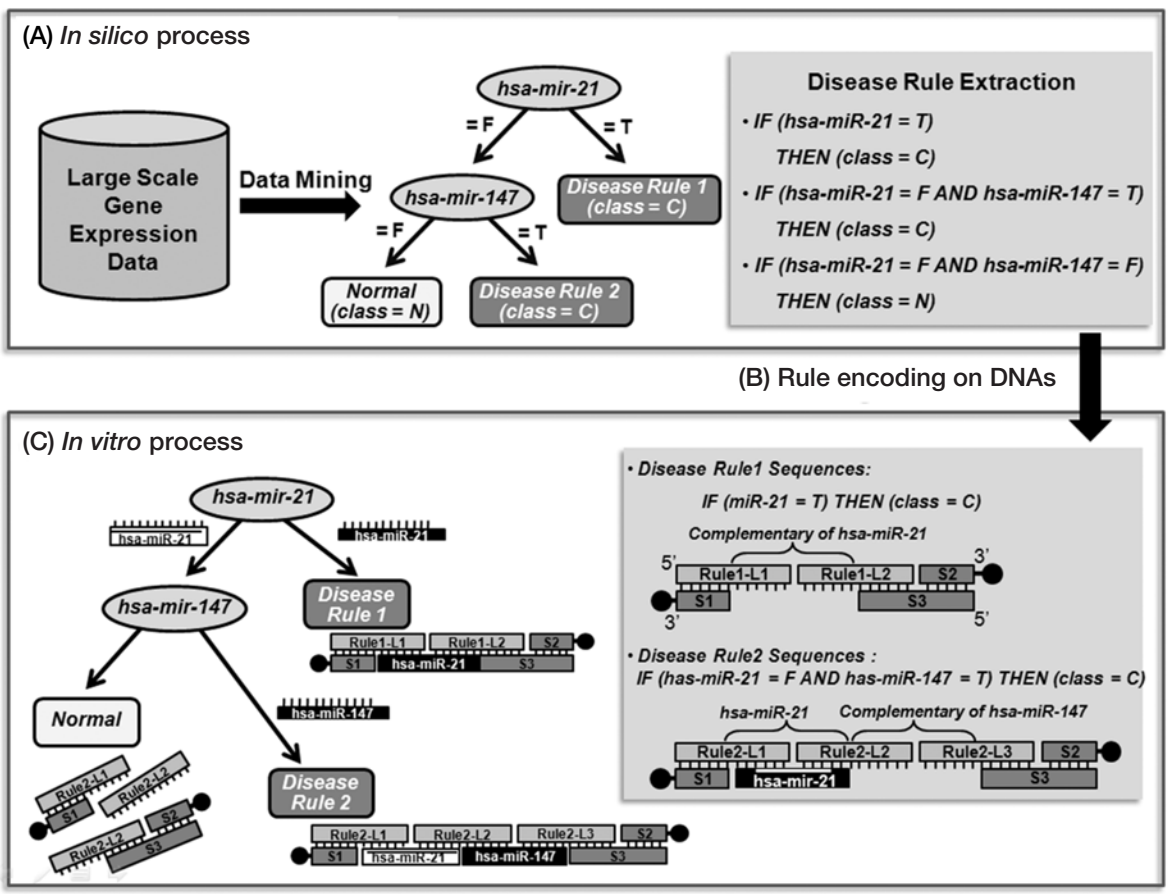

Figure 3. Rule-based molecular classification of miRNA-related disease types. Class 'C' indicates cancer and class 'N' indicates normal. (A) Three rules are extracted using the decision tree classifier. (B) Two rules for cancer detection are encoded in DNA molecules. (C) Demonstration of the *in vitro* procedure.

process can be further simplified by using only two rules because the input pattern is classified as "Normal" if it satisfies none of the two rules for class "Cancer". The two rules for class "Cancer" are as follows: "If (hsa-miR-21=T) then Cancer" (Rule 1) and "If (hsa-miR-21=F) AND (hsa-miR-147=T) then Cancer" (Rule 2). The sequence set is shown in Table 3. Note that the complementary sequence of hsa-miR-21 ($\overline{\text{hsa-miR-21}}$) is included as a rule sequence for Rule 2 because the antecedent of Rule 2 contains "hsa-miR-21=F".

Table 3. The sequence sets for miRNA-related cancer classification.

Rule detecting sequences (5′→3′)		
S	S1	GTTGCCAT-thiol (C3)
	S2	GTTCGTTT-thiol (C3)
	S3	AAACGAACGAACGACT
Rule 1	Rule1 L1	ATGGCAACTCAACATCAGT
	Rule1 L2	CTGATAAGCTAAGTCGTTC
Rule 2	Rule2 L1	ATGGCAACTAGCTTATCAG
	Rule2 L2	ACTGATGTTGAGCAGAAGCAT
	Rule2 L3	TTCCACACACAGTCGTTC
	$\overline{\text{hsa-miR-21}}$	TCAACATCAGTCTGATAAGCTA
Marker sequences (5′→3′)		
Marker	hsa-miR-21	UAGCUUAUCAGACUGAUGUUGA
	hsa-miR-147	GUGUGUGGAAAUGCUUCUGC

These two rules were tested in separate tubes for parallel operation. In each tube, the input miRNAs were hybridized with rule sequences for each rule (Figure 3C). Figure 4 summarizes the classification results using Rules 1 and 2 with various settings. Figure 4A shows the observed colors, and 4B represents the UHR-SEM image. Tubes (1) through (3) in Figure 4 correspond to Rule 1 with different inputs, and tubes (4) through (6) correspond to Rule 2 with various inputs. Significantly, tubes (3) and (6) represent cases in which the rules are satisfied by their corresponding markers; the input belongs to the class "Cancer".

Therefore, we can observe a significant color change accompanied by aggregation in Figure 4A. Additionally, we can observe an extensively cross-linked network of Au-NPs in the microscopic images (Figure 4B) of tubes (3) and (6), which satisfies the rules. Tubes (1) and (4) represent cases in which the rule sequence set is not complete. Tube (1) contains only S1 and S2 sequences, which can be used as a negative control. Therefore, no color change is observed. In tube (4), the complementary sequence of hsa-miR-21 is missing. Thus, no color change is observed even if the input

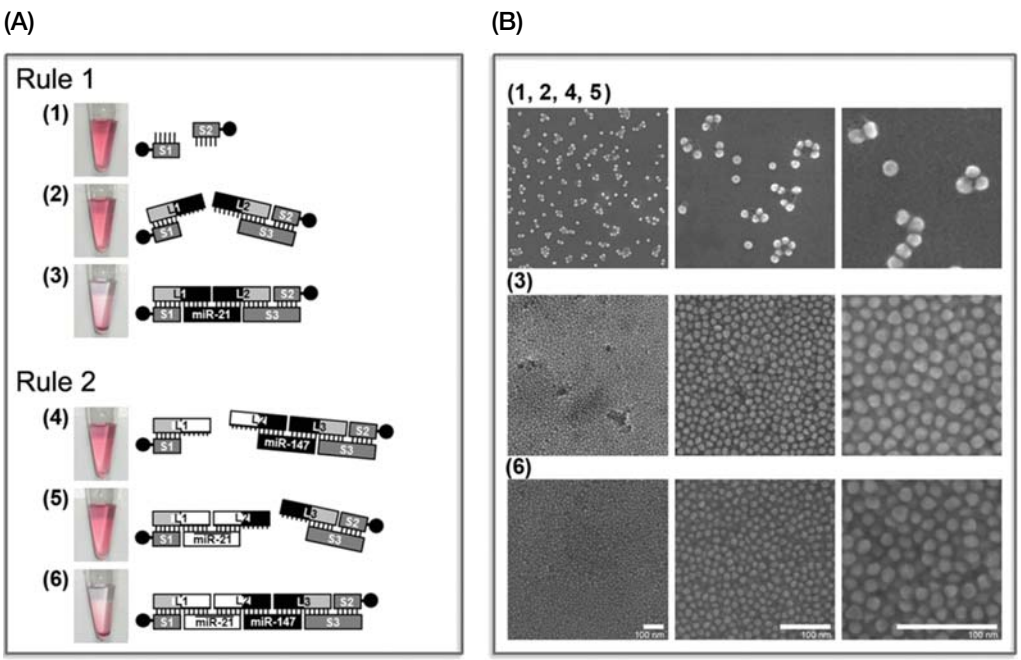

Figure 4. Demonstration of a molecular diagnosis. (A) Color responses of the given cases. (1) Only S1 and S2 sequences exist in the tube. (2) The sequence for Rule 1 (hsa-miR-21) is missing. (3) All necessary sequences for rule 1 exist. (4) The rule sequence set is incomplete because ($\overline{\text{hsa-miR-21}}$) is missing. (5) The sequence for Rule 2 (hsa-miR-147) is absent. (6) All necessary sequences for Rule 2 exist. (B) UHR-SEM analysis. The first, second and third columns show the UHR-SEM images under magnifications of 100 K, 250 K, and 500 K, respectively. Bars indicate 100 nm. A massive aggregation can be observed only in cases (3) and (6). The nanoparticles are distributed evenly within the aggregation, which may have been caused by the double-stranded DNA structures.

(or marker) pattern (hsa-miR-21 is absent and hsa-miR-147 is present) satisfies Rule 2. Tubes (2) and (5) are the cases in which the input pattern represents the sample without "Cancer". Tube (2) represents the case in which hsa-miR-21 is absent, and tube (5) corresponds to the case in which hsa-miR-147 is absent. In these cases, the input sequences do not satisfy the encoded rules; thus, no color change is observed in Figure 4A and no aggregation of the Au-NP network is observed in Figure 4B. These results demonstrate that the proposed method can successfully diagnose disease *in vitro* using pre-defined diagnostic rules.

Conclusions

We developed a rule-based *in vitro* molecular classification method and successfully applied this method to the classification of miRNA-related cancer types. This demonstrates the possibility of an intelligent classification system using *in vitro* information directly. Our method highlights several significant aspects of the diagnostic applications of DNA computing. First, we introduced a new rule-based diagnosis method that handles information *in vitro*. Depending on the composition of rule sequences, many variables (markers) can be efficiently analyzed at the same time without individual assays, which is an important advantage of our method. Second, the simple binary output using Au-NP aggregation allows for easy detection of the target class and improves practical availability. This approach could reduce the classification steps carried out by current laboratory systems. Finally, if this method can be combined with lab-on-a-chip technologies[19], a new portable disease diagnosis device can be implemented. However, the proposed method will require further detailed analysis before it can be used in practical applications because real disease samples might be more complicated than those used in this study. For example, the minimum detection limit of Au-NP aggregation is important because the amount of sequences might vary in real samples. The maxi-

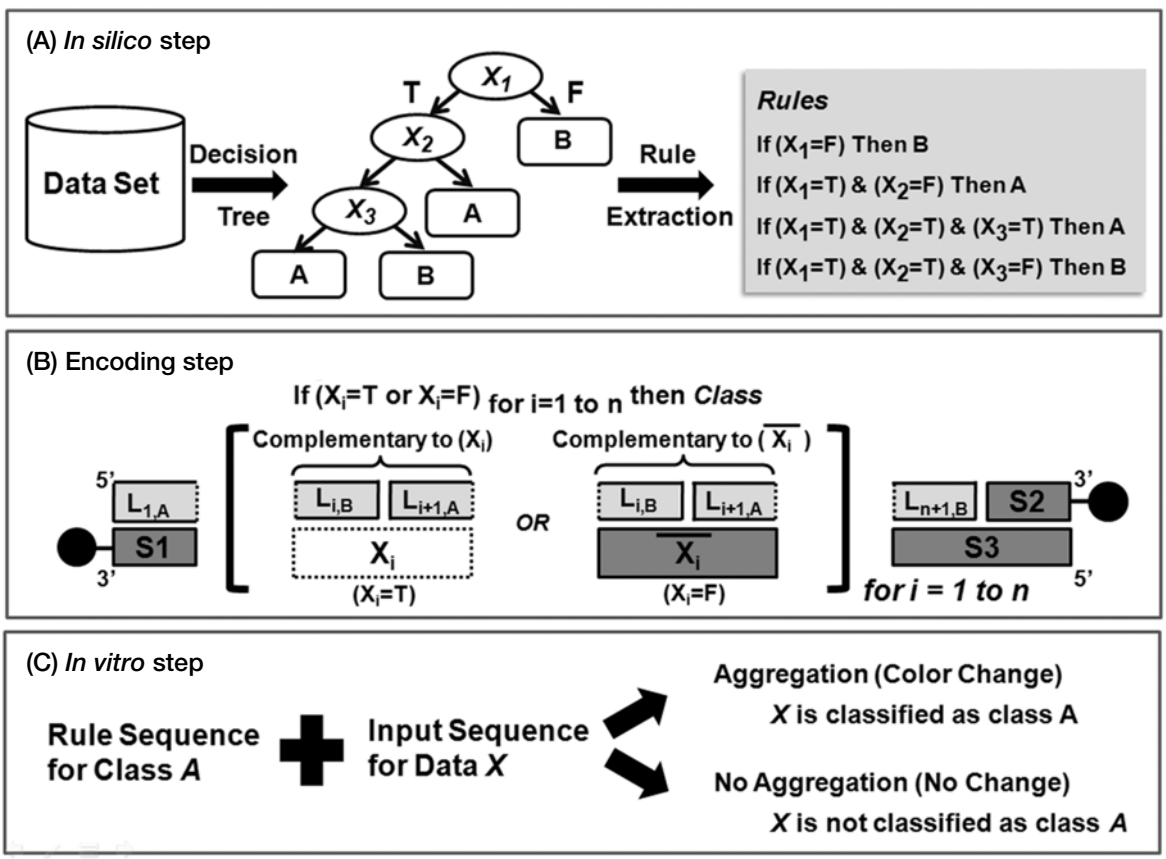

Figure 5. The overall procedure of the rule-based *in vitro* molecular pattern classification method. (A) *In silico* Step: Classification rules are formulated using pattern classification algorithms. (B) Encoding Step: A sequence design step that encodes the extracted rules in DNA sequences. (C) *In vitro* Step: When the rule sequences are hybridized with input molecules, the aggregation results indicate how to classify the input.

mum number of markers that can be handled using this approach should also be studied. Additionally, this method should be extended for the quantitative analysis of markers in addition to the presence/absence of markers.

Materials and Methods

Overall *in vitro* molecular classification procedure

The overall procedure of the proposed method is illustrated in Figure 5. The preliminary step is the rule extraction step (Figure 5A), and the classification rules are extracted from the given data *in silico*. Any pattern classification algorithm that can build conjunctive classification rules can be used in this step. In the next step (Figure 5B), the extracted rules are encoded in DNA sequences as rule sequences. In the final step (Figure 5C), the DNA computing algorithm interprets the input molecules and classifies them using the given rules.

In silico step for rule-based molecular pattern classification

To construct a set of classification rules from the data, a decision tree algorithm was applied, which is one of the most widely and successfully used classification methods among various machine learning techniques. A trained decision tree can be interpreted as a set of simple if-then classification rules[20]. Each rule can be generated by each path from the root to a leaf node. The antecedent of a rule is a conjunction for every node on the path from the root, and the consequent of the rule is the class assigned by the leaf node. Figure 5A shows an example of a decision tree with 3 variables (X_1, X_2, and X_3) and 2 classes (A and B). Let us assume that each variable has a binary value (T or F), and we take the left or right branch when the variable in the node has the value T or F. Because there are 4 leaf nodes in the tree (marked as A or B in Figure 5A), we can extract four classification rules from the tree as shown in Figure 5A. Additionally, the rule extraction step is independent of the molecular classification step and can be performed beforehand *in silico*. We used WEKA 3.4.8[21] for training a decision tree.

DNA sequence design for *in vitro* molecular pattern classification

Each classification rule is encoded in DNA molecules (called 'rule sequences') so that the rule sequences and input molecules form perfect double strands only when the corresponding rule is satisfied (Figure 5B). In this way, each rule can be tested in separate tubes in parallel. We assumed binary values (T or F) for each variable (X_i), and the values (y_i) are represented as the existence (T) or the absence (F) of the corresponding molecules. Therefore, each classification rule will have the form: "If (X_1=y_1) AND ⋯ AND (X_n=y_n) then Class A", where y_i=T or F for i=1 to n.

The detailed procedures of rule encoding are illustrated in Figure 5B and are as follows:

1. X_i=T is represented by the existence of the sequence for X_i.
2. X_i=F is represented by the absence of the sequence for X_i.
3. For a rule with n terms of (X_i=y_i), the rule sequence consists of (n+1) linkers, 3 additional sequences (S1, S2, S3), and the necessary complementary sequences $\overline{X_i}$, s, which represent absence, to make fully double-stranded DNAs.
4. The linker sequences L_i (for i=1, ⋯ , n+1) are concatenations of two subsequences $L_{i+1,A}$ and $L_{i,B}$.
5. If y_i=T, then $L_{i,B}$ and $L_{i+1,A}$ are complementary to the half of X_i from the 3′-end and the 5′-end, respectively.
6. If y_i=F, $L_{i,B}$ and $L_{i+1,A}$ are complementary to the half of $\overline{X_i}$ (complementary sequence of X_i) from the 3′-end and the 5′-end, respectively.
7. $L_{1,A}$ and $L_{n+1,B}$ are complementary to S1 and the half of S3 from the 3′-end, respectively. S2 is complementary to the half of S3 from the 5′-end.

S1 and S2 are attached to Au-NPs for detection purposes. S3 caps the remaining part of sequences to make fully double-stranded DNAs. Pre-designed S sequences (S1, S2, and S3) are required for convenience in handling different input sequences. If S1 and S2 are designed to hybridize to input sequences directly, they should be changed whenever the input sequences are changed. However, it is expensive and time consuming to prepare different S sequences and attach them to Au-NPs each time.

Results classification using Au-NP aggregation

To monitor the classification results, we used Au-NP aggregates. Au-NPs provide a convenient way to distinguish the solutions containing perfect double-strands via a color change[22]. As the Au-NP complexes grow and reach a critical size, they precipitate out of solution, and the characteristic plasmon peak is decreased causing a color change. This color change can be observed with the naked eye as well as a UV/Vis spectrometer[22]. The ease of visualization by Au-NP aggregation also eliminates the need for laborious detection methods and thus reduces the time required for the total procedure and enhances the practical availability.

Preparation of sequences

If mismatches occur during hybridization, the final results could be incorrect. To prevent mismatches, we optimized the sequences when encoding the rules by minimizing the possibility of cross-hybridization. NACST/Seq[23] and OligoAnalyzer 3.1[24] were used for designing the sequences and checking for cross-hybridization. Two thiol-modified sequences (S1 and S2) that are attached to Au-NPs were modified based on Mirkin group's work[22]. These S sequences are also shown in each table. To classify the miRNA-related cancer type, the miRNA sequences were searched (miRNA database, http://www.mirbase.org/), and the rule sequences were designed based on the miRNA sequences. All DNA sequences used in this experiment were purchased from Bioneer (Daejeon, Korea). Each pellet was resuspended to a stock concentration of 100 μM in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and stored at −20°C until use.

Au-NP preparation and concentration

To disrupt disulfide dimers of thiol-modified oligonucleotides, the modified sequences were treated with 100 mM dithiothreitol (DTT) (Sigma, MO, USA) for 10 min, respectively. DTT was removed by extracting 3 times with 4 volumes of ethyl acetate (Sigma, USA). The mixture was vortexed for 5 min and centrifuged for 5 min at 12,000 rpm (Eppendorf, Germany). After discarding the upper layer, the solutions were dried for 1 hr at 37°C to remove the ethyl acetate. The water that evaporated during the ethyl acetate removal steps was replenished to maintain a final concentration of 100 μM. Subsequently, we immediately added thiol-tagged oligomers to Au-NP to avoid disulfide dimerization of free sulfhydryl groups after removal of DTT. The final concentration of the thiol-modified sequences was 1 μM in the colloidal Au-NP solution. The colloidal gold nanoparticles (13 ± 2 nm; 1.4×10^{12} particles/mL) were purchased from BBInternational (London, UK). The mixtures were then incubated for 24 hr at room temperature with gentle agitation. We carried out a 40-hr incubation after adding aging buffer containing 0.1 M NaCl and 0.1 M sodium phosphate. Subsequently, we removed the unreacted sequences by centrifugation (13,200 rpm, 30 min, 2 times), dissolved the DNA-Au-NP complexes in the same buffer and stored the samples at 4°C until use. The final concentration of the samples was approximately 400 strands per particle (6×10^{14} DNA strands/1.4×10^{12} particles).

Hybridization and visualization

Each of the rule and input sequences was added to 100 μL of DNA-Au-NP complexes in 200 μL PCR tubes, and additional NaCl was added. The final concentration of the rule sequences was 1 μM, and the final NaCl concentration was 0.2 M (4 μL of 5 M NaCl per 100 μL reaction solution). The reaction mixture was incubated at 80°C for 10 min, and the temperature was steadily lowered to 25°C at 1°C per 10 sec using a thermal cycler (Bio-Rad, USA). The mixture was then incubated at room temperature for approximately 3 hr. All reactions were performed in triplicate with independent sample preparation, and the reliability of solutions was tested by repeating the heating and cooling cycles. The results of molecular classification were also characterized by ultraviolet-visible (UV/Vis) spectrometry (Nanodrop, DE, USA) and ultra high resolution scanning electron microscopy (S-5500, Hitachi, Japan).

Acknowledgements This work was supported by NRF (2012-0005643, 2012-0005801) and BK21-IT. KAY was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2008-532-C00031). The ICT at Seoul National University provided the research facilities for this study.

References

1. Adleman, L.M. Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021-1024 (1994).
2. Reif, J.H. The emergence of the discipline of biomolecular computation in the US. *New Generat. Comput.* **20**, 217-236 (2002).
3. Amos, M., Paun, G., Rozenberg, G. & Salomaa, A. Topics in the theory of DNA computing. *Nat. Comp.* **287**, 3-38 (2002).
4. Chen, X. & Ellington, A.D. Shaping up nucleic acid computation. *Curr. Opin. Biotech.* **21**, 392-400 (2010).
5. Lee, J.Y., Shin, S.-Y., Park, T.H. & Zhang, B.-T. Solving traveling salesman problems with DNA molecules encoding numerical values. *Biosystems* **78**, 39-47 (2004).
6. 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).
7. Martinez-Perez, I.M., Zhang, G., Ignatova, Z. & Zimmermann, K.-H. Computational genes: a tool for molecu-lar diagnosis and therapy of aberrant mutational phenotype. *BMC Bioinformatics* **8**, 365 (2007).
8. Rinaudo, K. *et al.* A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat. Biotechnol.* **25**, 795-801 (2007).
9. Lee, I.-H. *et al.* The use of gold nanoparticle aggregation for DNA computing and logic-based biomolecular detection. *Nanotechnology* **19**, 395103 (2008).

10. Qian, L. & Winfree, E. Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **332**, 1196-1201 (2011).
11. Qian, L., Winfree, E. & Bruck, J. Neural network computation with DNA strand displacement cascades. *Nature* **475**, 368-372 (2011).
12. Stojanovic, M.N. & Stefanovic, D. A deoxyribozyme-based molecular automaton. *Nat. Biotechnol.* **21**, 1069-1074 (2003).
13. Macdonald, J. *et al*. Medium scale integration of molecular logic gates in an automaton. *Nano Lett.* **6**, 2598-2603 (2006).
14. 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).
15. Lund, K. *et al*. Molecular robots guided by prescriptive landscapes. *Nature* **465**, 206-209 (2010).
16. Gu, H., Chao, J., Xiao, S.-J. & Seeman, N.C. A proximity-based programmable DNA nanoscale assembly line. *Nature* **465**, 202-205 (2010).
17. Stojanovic, M. Some experiments and directions in molecular computing and robotics. *Israel J. Chem.* **51**, 99-105 (2011).
18. Lu, J. *et al*. MicroRNA expression profiles classify human cancers. *Nature* **435**, 834-838 (2005).
19. Lee, S.H. *et al*. Biomolecular theorem proving on a chip: A novel microfluidic solution to a classical logic problem. *Lab. Chip* **12**, 1841-1848 (2012).
20. Mitchell, T. Machine Learing. McGraw-Hill (1997).
21. Witten, I.H. & Frank, E. Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition. Morgan Kaufmann (2005).
22. Mirkin, C.A., Letsinger, R.L., Mucic, R.C. & Storhoff, J.J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607-609 (1996).
23. Shin, S.-Y., Lee, I.-H., Kim, D. & Zhang, B.-T. Multiobjective evolutionary optimization of DNA sequences for reliable DNA computing. *IEEE T. Evolut. Comput.* **9**, 143-158 (2005).
24. Owczarzy, R. *et al*. IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Res.* **36**, W163-W169 (2008).