Electronic Supplementary Information (ESI)

Rapid detection of cancer DNA in human blood using cysteamine-capped AuNPs and a machine learning-enabled smartphone

Sireemas Koowattanasuchat[a], Sawinee Ngernpimai[a], Piyaporn Matulakul[a], Janpen Thonghlueng[a], Witthawat Phanchai[a], Apiwat Chompoosor[b], Uthumporn Panitanarak[c], Yupaporn Wanna[d], Thanapong Intharah[d], Kanokon Chootawiriyasakul[e], Pimjai Anata[f], Prajuab Chaimnee[f], Raynoo Thanan[g], Chadamas Sakonsinsiri[g], Theerapong Puangmali[a,*]

[a] *Department of Physics, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand*

[b] *Department of Chemistry, Faculty of Science, Ramkhamhaeng University, Bangkok 10240, Thailand*

[c] *Department of Biostatistics, Faculty of Public Health, Mahidol University, Bangkok 10400, Thailand*

[d] *Department of Statistics, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand*

[e] *Department of Medicine, Faculty of Madicine, Khon Kaen University, Khon Kaen 4002, Thailand*

[f] *Molecular Diagnosis Unit, Central Laboratory, Srinagarind Hospital, Khon Kaen University, Khon Kaen 4002, Thailand*

[g] *Department of Biochemistry, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand*

[*] *Corresponding author: E-mail: theerapong@kku.ac.th*

I. DATA ANALYSIS

A confusion matrix (Fig. 5(c)), also known as an error matrix in machine learning (ML), is a specific table structure that permits visualization of the performance of an algorithm, typically one that uses supervised learning. Each row of the matrix represents an actual class, whereas each column represents a forecast class. In the present work, cancer patients and normal cases are considered as positive and negative labels, respectively. The following definitions apply to terminology and derivations from a confusion matrix.

Condition positive (P)	represents the number of true positive cases in the data set.
Condition negative (N)	is the number of real negative cases in the data.
True positive (TP)	is a test result that correctly indicates the presence of a condition or characteristic.
True negative (TN)	is a test result that correctly indicates the absence of a condition or characteristic.
False positive (FP)	is a test result which wrongly indicates that a particular condition is present.
False negative (FN)	is a test result which wrongly indicates that a particular condition or attribute is absent.

Sensitivity

Sensitivity is defined as the fraction of true positive labels (TP) relative to the summation of true positive and false negative labels ($TP + FN$).

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{1}$$

Specificity

Specificity is defined as the fraction of true negative labels (TN) relative to the summation of true negative and false positive labels ($TN + FP$).

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2}$$

Positive predictive value (PPV)

Positive predictive value is defined as the fraction of the number of true positives (TP) to the number of all positive calls ($TP + FP$).

$$PPV = \frac{TP}{TP + FP} \tag{3}$$

Negative predictive value (NPV)

Negative predictive value is defined as the fraction of the number of true negatives (TN) to the number of all negative calls ($TN + FN$).

$$NPV = \frac{TN}{TN + FN} \tag{4}$$

Accuracy

Accuracy is the proportion of accurately predicted labels among all predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Fig. 5(c) is a summary of the results predicted by the machine learning analysis ($TP = 31$, $TN = 10$, $FP = 2$, and $FN = 0$). Calculating sensitivity, specificity, PPV, NPV, and accuracy is as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{31}{31 + 0} = 1.000 \tag{6}$$

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{10}{10 + 2} = 0.8383 \tag{7}$$

$$PPV = \frac{TP}{TP + FP} = \frac{31}{31 + 2} = 0.9394 \tag{8}$$

$$NPV = \frac{TN}{TN + FN} = \frac{10}{10 + 0} = 1.000 \tag{9}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{31 + 10}{31 + 10 + 2 + 0} = 0.9534 \tag{10}$$

Therefore, the sensitivity, specificity, PPV, NPV, and accuracy are 100.0%, 83.3%, 93.9%, 100.0%, and 95.3%, respectively.

II. ADDITIONAL EXPERIMENTAL METHOD AND RESULTS

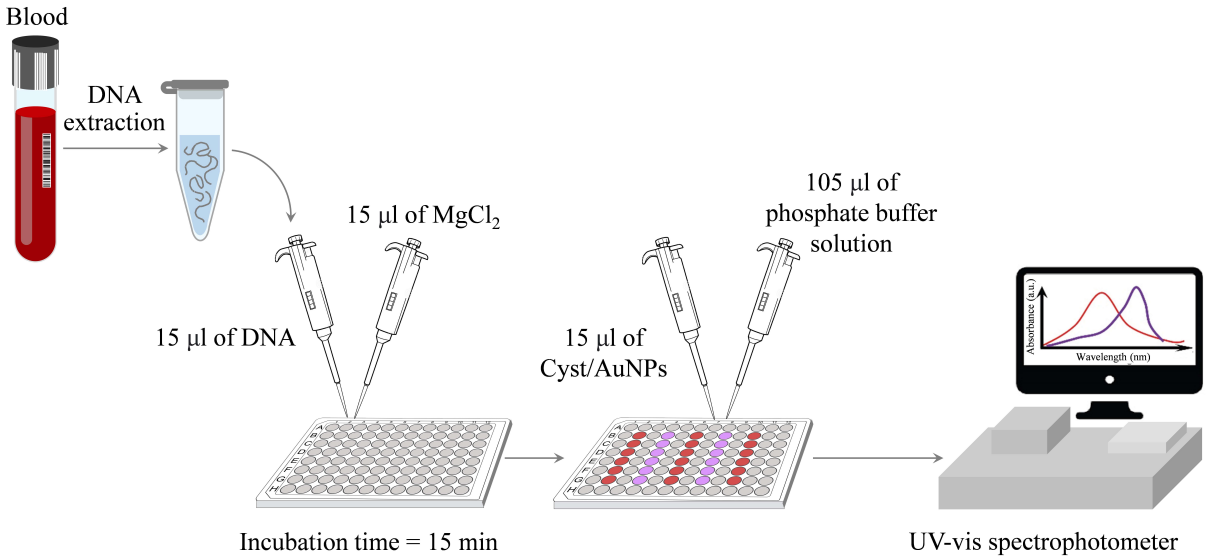


Fig. S1: Schematic illustration of the procedure for Methylscape detection.

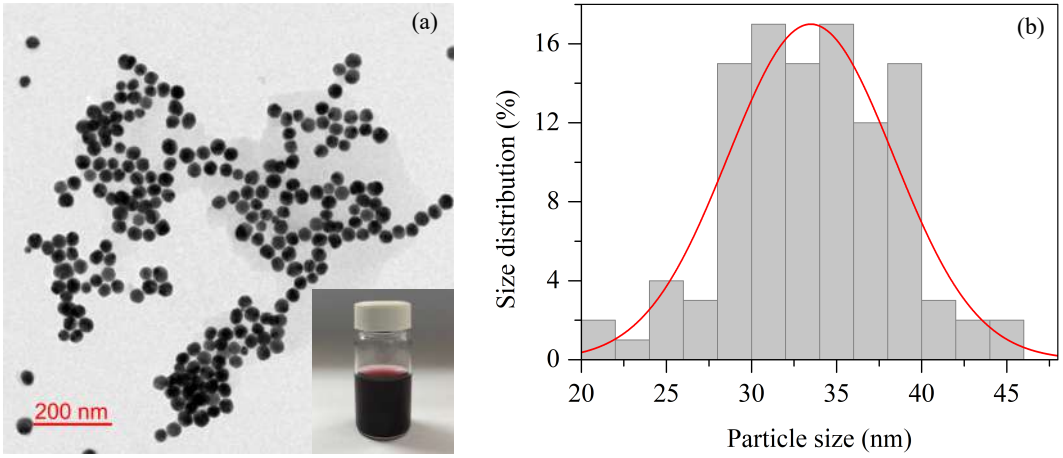


Fig. S2: (a) Transmission electron microscopy image of cysteamine-capped gold nanoparticles (Cyst/AuNPs). (b) Size distribution of Cyst/AuNPs.

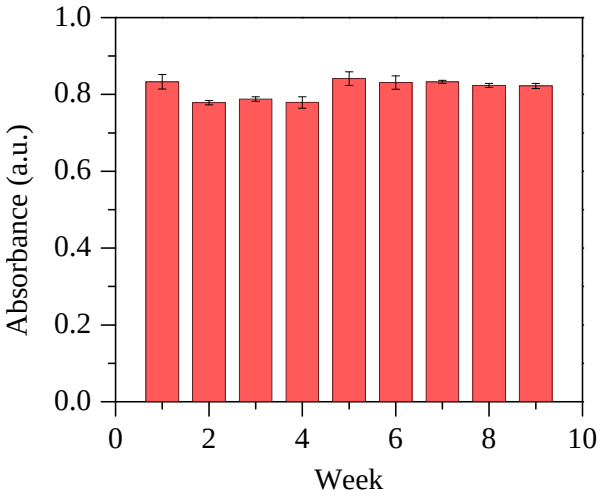


Fig. S3: Stability of cysteamine-decorated gold nanoparticles (Cyst/AuNPs). The absorbance (at 525 nm) of the nanoparticle solution had been measured every week for nine weeks. The solution was stored at a temperature of 4 C°.

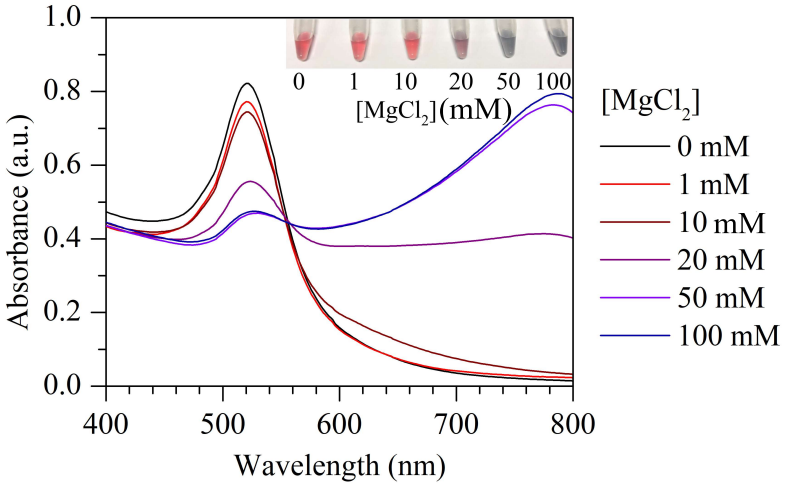
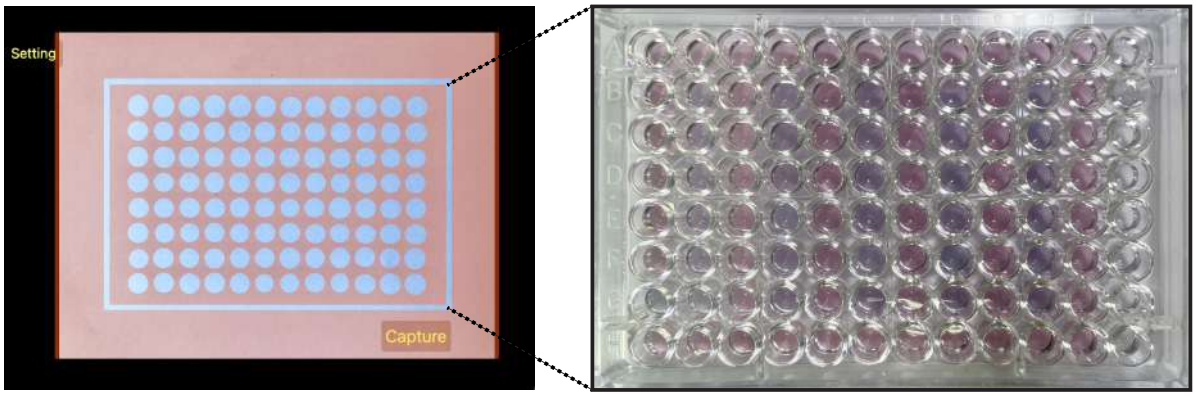


Fig. S4: The absorbance of AuNPs solution in the presence of $MgCl_2$ with different concentrations.



(a) (b)

Fig. S5: (a) The screen of the mobile phone application used for taking photograph. (b) An example of the image taken with our developed application for the colorimetric cancer screening.

Table S1: Clinical information of representative DNA samples extracted from the blood.

Sample	Gender	Age (Year)	Absorbance($\Delta A_{650/525}$)	Type	Diagnostic
C1	Male	44	0.599	Leukemia	Cancer
C2	Female	54	0.452	Leukemia	Cancer
C3	Female	63	0.387	Leukemia	Cancer
C4	Female	24	0.655	Leukemia	Cancer
C5	Male	60	0.703	Leukemia	Cancer
C6	Female	59	0.600	Leukemia	Cancer
C7	Female	54	0.398	Leukemia	Cancer
C8	Female	74	0.421	Leukemia	Cancer
C9	Female	61	0.845	Leukemia	Cancer
C10	Male	61	0.705	Leukemia	Cancer
C11	Female	79	0.532	Leukemia	Cancer
C12	Female	37	0.700	Leukemia	Cancer
C13	Male	66	0.155	Leukemia	Cancer
C14	Female	52	0.226	Leukemia	Cancer
C15	Female	63	0.373	Leukemia	Cancer
C16	Male	77	0.554	Leukemia	Cancer
C17	Male	52	0.601	Leukemia	Cancer
C18	Male	29	0.503	Leukemia	Cancer
C19	Male	35	0.468	Leukemia	Cancer
C20	Male	41	0.380	Leukemia	Cancer
C21	Male	18	0.604	Leukemia	Cancer
C22	Female	48	0.385	Leukemia	Cancer
C23	Female	71	0.525	Leukemia	Cancer
C24	Male	51	0.661	Leukemia	Cancer
C25	Male	53	0.307	Leukemia	Cancer
C26	Male	45	0.149	Leukemia	Cancer
C27	Male	60	0.277	Leukemia	Cancer
C28	Male	65	0.157	Leukemia	Cancer
C29	Female	27	0.407	Leukemia	Cancer
C30	Female	59	0.613	Leukemia	Cancer
C31	Male	74	0.666	Leukemia	Cancer
N1	N/A	N/A	0.960	Normal	Normal
N2	N/A	N/A	0.891	Normal	Normal
N3	N/A	N/A	0.996	Normal	Normal
N4	Male	26	0.789	Normal	Normal
N5	Male	25	0.785	Normal	Normal
N6	Male	22	0.894	Normal	Normal
N7	Male	32	0.737	Normal	Normal
N8	Male	22	0.792	Normal	Normal
N9	Female	22	0.869	Normal	Normal
N10	Male	25	0.924	Normal	Normal
N11	Male	26	0.973	Normal	Normal
N12	Female	28	0.838	Normal	Normal