Introduction:

In the modeling part, we decide to prove that using NASBA to amplify the circulating tumor DNA is necessary and to estimate the optimal concentration of the circulating tumor DNA for detection.

The influence of cell free DNA on the effectiveness of the detection device:

In real physiological environment of the plasma, there are many factors influencing the effectiveness of dcas9 protein binding with the targeted DNA sequence (EML4-ALK), the most prominent one of which is cell free DNA. Cell free DNAs come from apoptosis and necrosis of normal cells and may include gene EML4 and gene ALK, which are potential ligands bind with dcas9 protein and lead to a decrease of the effective concentration of NT7-dcas9 and CT7-dcas9 protein. The effective concentration is referred as the concentration of NT7-dcas9 and CT7-dcas9 which bind to the target sequence (EML4-ALK fusion gene), leading to one DNA molecule bears both NT7-dcas9 and CT7-dcas9 to enable the transcription of report gene. In order to determine the magnitude of the influence from the cell free DNAs on the effectiveness of split-T7-dcas9 in binding with EML4-ALK, the following model is made. Assumption that NT7-dcas9 binds with sequence of EML4 and CT7-dcas9 binds with sequence of ALK is made in the formula.

$$\frac{\mathrm{d[NT7\ eff]}}{\mathrm{d[EML4]}} = [\mathrm{NT7\ total}] \times \theta_N \times \frac{-[\mathrm{EML4-ALK}]\times\Delta[\mathrm{EML4}]}{([\mathrm{EML4}]+[\mathrm{EML4-ALK}])\times([\mathrm{EML4}]+\Delta[\mathrm{EML4}]+[\mathrm{EML4-ALK}])}$$

$$\frac{\mathrm{d[CT7\ eff]}}{\mathrm{d[ALK]}} = [\mathrm{CT7\ total}] \times \theta_C \times \frac{-[\mathrm{EML4-ALK}]\times\Delta[\mathrm{ALK}]}{([\mathrm{ALK}]+[\mathrm{EML4-ALK}])\times([\mathrm{ALK}]+\Delta[\mathrm{ALK}]+[\mathrm{EML4-ALK}])}$$

Where θ_N and θ_C are the fraction of NT7-dcas9 and CT7-dcas9 which bind with potential binding sites, calculated by hill equation as $\theta_N = \frac{[\mathrm{NT7}]^n}{K_d+[\mathrm{NT7}]^n}$ and $\theta_C = \frac{[\mathrm{CT7}]^n}{K_d+[\mathrm{CT7}]^n}$.

Symbols are shown in the table below

[NT7 eff]	The concentration of NT7-dcas9 binding with target sequence (EML4-ALK)
[CT7 eff]	The concentration of CT7-dcas9 binding with target sequence (EML4-ALK)
[NT7 total]	The total concentration of NT7-dcas9 in the system
[CT7 total]	The total concentration of CT7-dcas9 in the system
[EML4-ALK]	The concentration of circulating tumor DNA EML4-ALK
[EML4]	The concentration of cell free DNA EML4
[ALK]	The concentration of cell free DNA ALK
[NT7]	The concentration of NT7-dcas9
[CT7]	The concentration of CT7-dcas9
K_d	Apparent dissociation constant
n	Hill coefficient

As shown in the formula above, given the total concentration of NT7-dcas9 and CT7-dcas9, the effective concentration of NT7-dcas9 and CT7-dcas9 decreases as the concentration of cell free DNA increase. In real physiological environment of plasma, the concentration of cell free DNAs is seven orders of magnitude larger than the

concentration of circulating tumor DNAs. Therefore, it can be assumed that d[EML4] and d[ALK] approach to infinity. Hence, in this scenario, the change in effective concentration of NT7-dcas9 and CT7-dcas9 is equal to negative infinity, revealed as no signal emitted even if EML4-ALK fusion gene exists.

Based on the modeling, in order to gain valid result of the presence and concentration of EML4-ALK fusion gene, amplifying the concentration of EML4-ALK fusion gene is necessary. By employing the Nucleic Acid Sequence-Based Amplification (NASBA) **(Figure 1)**, the concentration of EML4-ALK fusion gene increases at an exponential rate **(Figure 2)**. Within eight rounds, the concentration of EML4-ALK approaches to a detectable value. Therefore, it is proved that NASBA is a necessary and effective signal amplification approach before detection.

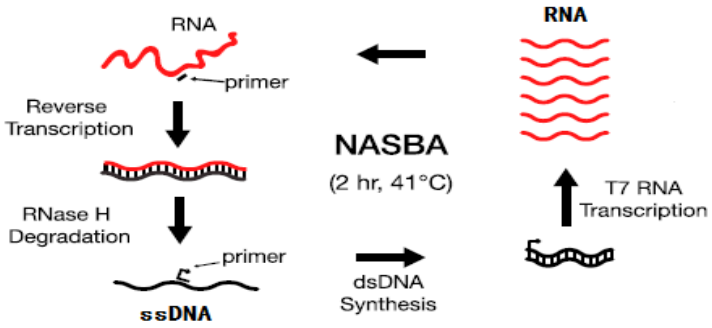

Figure 1. The condition and process of NASBA

NASBA is performed in 41 degree Celsius. Starting from single strand DNA (lower left corner), double strand DNA is synthesized. Then RNAs are transcribed from the double strand DNA. After reverse transcription and RNase H degradation, new single strand DNAs are synthesized, entering the next round.

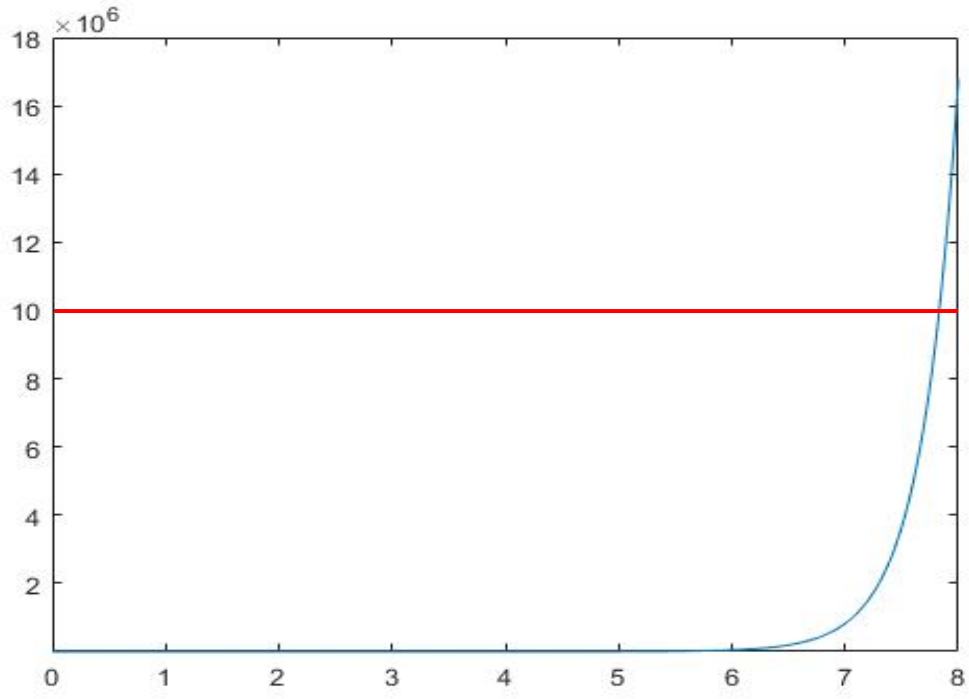

Figure 2. The concentration of EML4-ALK versus the rounds of NASBA

Based on the process of NASBA, the number of DNA molecules (x) after n rounds of NASBA can be represented as $x=n^n-n$. By sketching the curve $x=n^n-n$ (blue line) and $x=10^7$ (red line), the intersection point is revealed prior to n=8.

Determining the upper limit of target DNA concentration:

We decide to determine whether there is an upper limit of the target DNA concentration for split-T7-dcas9 system, since it is hypothesized that the effective concentration of NT7-dcas9 and CT7-dcas9 will decrease after the concentration of the target DNA reaches an upper limit; since the increase in potential binding sites increases the probability of one DNA molecule bears only one protein **(Figure 3)**. The upper limit is the optimal point for the highest activity of the detection device.

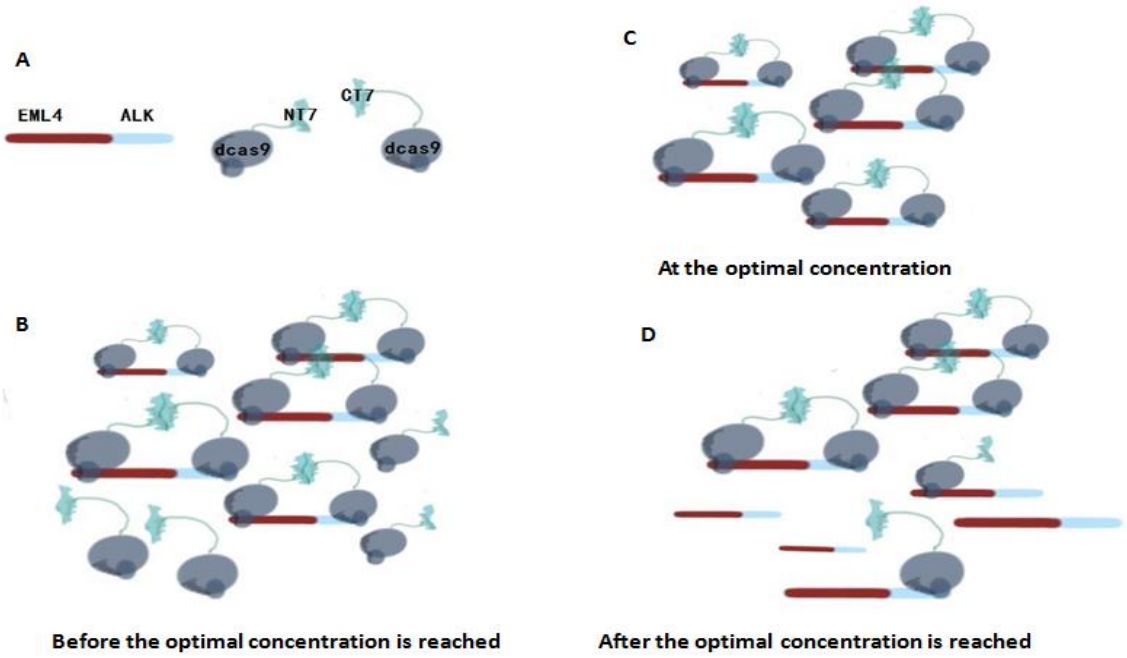

Figure 3. The binding of NT7-dcas9 and CT7-dcas9 with target DNA (EML4-ALK) before, at, and after the optimal DNA concentration

A. The EML4-ALK target DNA, NT7-dcas9, and CT7-dcas9 are represented by the geometric figures shown above.

B. The binding condition before the optimal concentration is reached. The number of NT7-dcas9 and CT7-dcas9 binding to DNA is far greater than the number of target DNA; therefore, all target DNA molecules bear both NT7-dcas9 and CT7-dcas9.

C. The binding condition at the optimal concentration. The number of NT7-dcas9 and CT7-dcas9 binding to DNA is exactly equal to the number of target DNA; therefore, all target DNA molecules bear both NT7-dcas9 and CT7-dcas9.

D. The binding condition after the optimal concentration is reached. The number of DNA molecules is greater than the number of NT7-dcas9 and CT7-dcas9; therefore, there is probability that one DNA molecule bears only one protein, either NT7-dcas9 or CT7-dcas9, decreasing the effective concentration of recombined T7 polymerase

To determine the optimal point for detection, the concentration of NT7-dcas9, the concentration of CT7-dcas9, and their fraction of binding with DNA are considered. The formula for the optimal point is established as:

$$2 \times \left[\mathrm{DNA}_{\mathrm{optimal}}\right] = [\mathrm{NT7\ total}] \times \theta_{\mathrm{N}} + [\mathrm{CT7\ total}] \times \theta_{\mathrm{C}},$$

where $[\mathrm{DNA}_{\mathrm{optimal}}]$ represents the optimal concentration of DNA for detection system.

When the concentration of DNA is lower than the optimal concentration, the concentration of NT7-dcas9 and CT7-dcas9 binding with the DNA increases linearly,

as the concentration of target DNA increases, despite there is the dissociation between target DNA and dcas9 (represented by θ). The change in effective concentration of NT7-dcas9 and CT7-dcas9 are represented in the formula below:

$$\frac{\mathrm{d[NT7\ eff]}}{\mathrm{d[DNA]}} = \mathrm{K_N} \times \theta_\mathrm{N}$$

$$\frac{\mathrm{d[CT7\ eff]}}{\mathrm{d[DNA]}} = \mathrm{K_C} \times \theta_\mathrm{C}$$

$$\frac{\mathrm{dRFU}}{\mathrm{d[DNA]}} = \mathrm{K_{T7}} \times \mathrm{MIN}(\frac{\mathrm{d[NT7\ eff]}}{\mathrm{d[DNA]}}, \frac{\mathrm{d[CT7\ eff]}}{\mathrm{d[DNA]}})$$

Symbols are shown in the table below

[DNA]	The concentration of target DNA
K_N	The increase in the concentration of effective NT7-dcas9 responding to one unit's increase in the concentration of target DNA
K_C	The increase in the concentration of effective CT7-dcas9 responding to one unit's increase in the concentration of target DNA
RFU	The relative fluorescent emit unit by the GFP (report gene) transcribed by the recombined T7 polymerase
K_{T7}	The increase in the relative fluorescent caused by one unit's increase in recombined T7 polymerase

When the concentration of target DNA is low, the dissociation between the DNA and the proteins almost does not occur, therefore, the dissociation constant is negligible and the relationship between effective concentration of NT7-dcas9 and CT7-dcas9 and the concentration of target DNA can be considered as pure linearity. However, after the concentration of DNA reaches a tipping point, the dissociation becomes prominent and the increase in effective concentration of NT7-dcas9 and CT7-dcas9 decreases as the concentration of DNA increases.

After the concentration reaches the optimal concentration for detection, the probability of one DNA molecule bearing both NT7-dcas9 and CT7-dcas9 and enabling transcription of report gene decreases as the concentration continually increases. The probability of one DNA molecule bearing both NT7-dcas9 and CT7-dcas9 is represented as the formula below:

$$\mathrm{P(NT7\ effective\ and\ CT7\ effective)} = \frac{\mathrm{[NT7]\times[CT7]}}{\mathrm{[DNA]^2}}.$$

Therefore, the change in the concentration of effective T7 polymerase responding to the change in the concentration of target DNA is represented as:

$$\frac{\mathrm{d[T7]}}{\mathrm{d[DNA]}} = \frac{-\Delta\mathrm{[DNA]\times[NT7]\times[CT7]}}{\mathrm{[DNA]\times([DNA]+\Delta[DNA])}} \times \theta_\mathrm{N} \times \theta_\mathrm{C},$$

where [T7] represents the concentration of recombined T7 polymerase.
Change in relative fluorescent unit is represented by:

$$\frac{\mathrm{dRFU}}{\mathrm{d[DNA]}} = \frac{-\Delta\mathrm{[DNA]\times[NT7]\times[CT7]}}{\mathrm{[DNA]\times([DNA]+\Delta[DNA])}} \times \theta_\mathrm{N} \times \theta_\mathrm{C} \times \mathrm{K_{T7}}.$$

Employing the formula above and the data from Peking 2015, the best-fitted curve for the relationship between relative fluorescent unit and target DNA concentration has been made in **Figure 4**. Revealed in the curve, before the optimal concentration is reaches, in the interval of low concentration (), the line rise linearly; whereas after the concentration reaches a certain value (), the slope of the curve decreases as a factor of θ. However, after the optimal concentration is reaches, the slope becomes negative and decreases as a factor of $\frac{[\mathrm{NT7}]\times[\mathrm{CT7}]}{[\mathrm{DNA}]^2}$. Therefore, the best-fitted curve proves that our hypothesis is correct.

Conclusion:

From the two models made above, we conclude that using NASBA to amplify the target DNA (EML4-ALK) is necessary and efficient for the detection device in order to avoid the interference of cell free DNAs from normal cells. Also, there is an upper limit of target DNA concentration for the detection device to create the most intense signal. The upper limit of the concentration of DNA is determined by the concentration of NT7-dcas9, CT7-dcas9, and their dissociation constant with target DNA.