Novel Biodevice for Colorectal Cancer Screening using Escherichia coli Nissle 1917 with miRNAs as Biomarker

A. INTRODUCTION

Colorectal cancer is a disease characterized by the presence of abnormal cell growth in the colon and rectum. Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related deaths worldwide, with the Asian population ranking first in both incidence and mortality. Global Cancer Statistics 2020 state that the combination of colorectal and rectal cancers contributes to approximately 10% (1,880,725) and 9% (915,880) of all cancer occurrences and deaths worldwide. In Indonesia, colorectal cancer holds the fourth and fifth positions in terms of incidence and mortality rates. These figures contribute to around 8.4% (33,427) and 7.6% (17,786) of all cancer occurrences and deaths in Indonesia. Furthermore, data from the Jogja Cancer Registry reveals that CRC is the second most prevalent type of cancer in Yogyakarta. According to the 2018 Basic Health Research (Riskesdas) results from the Ministry of Health of the Republic of Indonesia, Yogyakarta Province has the highest number of cancer cases in Indonesia. The prevalence of cancer in Yogyakarta Province is three times higher than the overall cancer prevalence in Indonesia, at 4.9 per 1,000 population. This highlights the urgency of addressing cancer cases in Yogyakarta Province. According to the distribution map of cancer cases in Yogyakarta in 2022, the average age of cancer patients is 55.08 ± 15.46 years. The most commonly diagnosed types of cancer are breast cancer, cervical cancer, and colorectal cancer. Most cancer patients come from Sleman Regency and economically disadvantaged areas, with one of the factors influencing the high number of cancer cases being the low and limited cancer screening rates. Therefore, efforts to improve access to cancer screening and raise public awareness about the importance of early cancer detection are crucial in addressing this issue in Yogyakarta Province, as well as other parts of the world.

Early diagnosis plays a crucial role in determining the success of therapy. Consequently, routine CRC screening is highly recommended as it can detect CRC at an early stage, reducing the incidence and mortality rates of CRC. While there are various screening methods available, they have their limitations. Endoscopic procedures like colonoscopy are invasive, requiring bowel preparation and sedation, and carry a small risk of complications such as bleeding or perforation. Non-invasive stool tests like fecal occult blood test (FOBT) and fecal immunochemical test (FIT) do not directly detect cancer and can produce false positives and negatives. Stool DNA tests, such as Cologuard, have a higher false positive rate and are more expensive than other stool tests. Imaging procedures like virtual colonoscopy using CT scans still require bowel preparation, carry a risk of false positives, and involve radiation exposure. Additionally, we aim to develop an affordable and minimally invasive early screening test for colorectal cancer (CRC), especially in rural areas facing unique challenges in implementing and accessing CRC screening services. One crucial aspect is limited education or awareness, contributing to a lack of understanding about risk factors, preventive measures, and the availability of CRC screening services. Rural areas are characterized by dispersed populations and long distances to healthcare facilities, leading to reduced access to screening centers, making it difficult for residents to undergo regular CRC screenings. Furthermore, rural populations frequently experience lower socioeconomic status, impeding access to healthcare resources. Factors such as limited health insurance coverage, financial constraints, and high out-of-pocket costs for screening tests can prevent individuals from seeking preventive care services. Finally, rural regions often face a shortage of healthcare providers, including gastroenterologists and primary care physicians, resulting in longer wait times and delays in diagnosis for patients residing in rural areas. The challenges mentioned above indicate that, in addition to the need for non-invasive CRC screening methods with high sensitivity and specificity,

it is equally imperative to consider additional factors such as ease of use, cost-effectiveness, and the elimination of the requirement for healthcare providers' intervention for a better CRC screening method in rural areas.

For all these reasons, we carry a mission within our Human Practices vision to connect and engage with various stakeholders and potentially-affected parties to share our insights with them and relay questions that we are still uncertain about and need further evidence, confirmation, and information. Dialogues with these individuals will enable us to obtain and retain a better understanding of the puzzle we are trying to solve with our project. The stakeholder interviews allowed us to implement the human-centred design approach as we take in advice and recommendations from people affected by our project. The Family of Patient expressed how our project is a great solution, as it is not only affordable but also accessible for people in rural areas. The Sleman District Health Office suggested putting an emphasis on community involvement through conducting promotional initiatives, such as distributing educational colorectal cancer leaflets and holding screening awareness campaigns. The Ministry of Health of the Republic of Indonesia relayed that the major issue in Indonesia is despite the Ministry having CRC in the "Health Transformation" program, screening is still scarce as the majority of the population could not afford to spend time to screen. This further spotlight the urgent need of our project. We also got inspired to devise the scale-up plan for our project through designing a manufacturing strategy and cost-effective analysis after convening with Lecturers from the Faculty of Pharmacy and the Faculty of Agricultural Technology in Universitas Gadjah Mada. At the same time of collecting and incorporating inputs, we worked on our wet lab experiment and dry lab computation.

B. MATERIALS AND METHODS

One of the most promising biomarkers for CRC is RNA in the form of miRNA. They have been found to be dysregulated in various types of cancer, making them potential biomarkers for cancer screening and diagnosis (Okugawa et al., 2014). Different miRNAs such as miR92a and miR21 exhibit upregulation during the early adenoma and advanced adenoma stages, respectively. Based on this, there is a LIRA, Loop-Initiated RNA Activator, which is a novel nucleic acid detection method based on RNA secondary structure. LIRA utilizes the concept of protein translation, wherein translation cannot commence if the ribosome fails to bind to the Ribosome Binding Site (RBS). By closing the RBS sequence in its secondary structure and opening it in the presence of specific biomarker RNA sequences, LIRA could be designed with various inputs and gates, including an OR gate, which facilitates the output or the expression of a reporter gene upon the presence of either of the two input sequences (Ma et al., 2022). Hence, in this study, we design LIRA OR Gate to detect at least one of two biomarkers that are upregulated in different CRC phases, those are miR21 (advanced adenoma) and miR92a (early adenoma) (Okugawa et al., 2014).

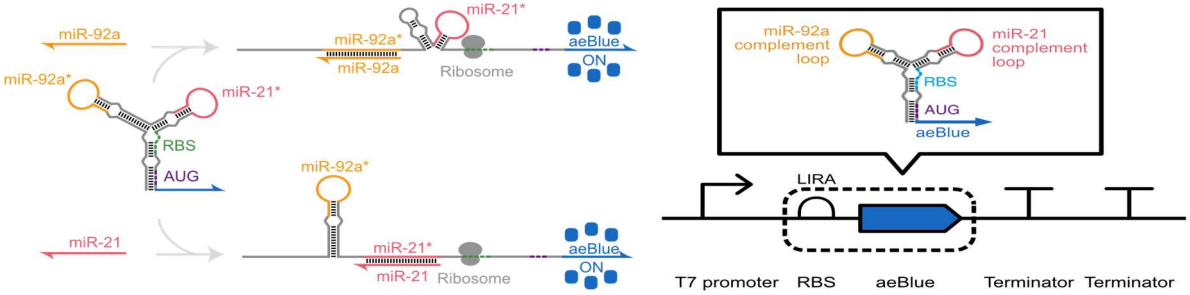


Figure 1. Working mechanism of Loop Initiated RNA Activator (LIRA) (left) and its SBOL circuit diagram (right)

However, in order to express LIRA, a safe and non-toxic biological host or chassis is required in the lumen. *Escherichia coli* Nissle 1917 or EcN is a bacterium that potentially could be used as chassis for human uses. It was developed by Dr. Alfred Nissle from the World War 1 era and since then has been marketed with the brand name Mutaflor® for Inflammatory Bowel Disease (IBD) treatment [14]. Today, EcN has been discovered to have anticancer activity by specifically colonizing and altering the signaling pathways of CRC cells (Chiang & Huang 2021; Alizadeh et al., 2020). In addition, the two endogenic plasmids of EcN could be engineered and maintained without the need for antibiotic selection (Kan et al., 2020). These properties of EcN make it a suitable chassis for use in CRC.

C. RESULTS

1. Metrics development for predicting ON OFF value

LIRA (Loop Initiated RNA Activator) is used as a design for EcN toehold RNA to get functional OR gate miRNA triggers. Since it's new and still under-researched, there is not yet a large dataset available of LIRA wet lab performance. Therefore, mathematical modeling is one of the alternatives to predict its performance and optimize the design process. Secondary structure of toehold has been one of the factors that are affected by the variation of sequences and would affect the toehold performance during the trigger and non-trigger to toehold reaction. Metrics development was already documented in the previous project https://2023.igem.wiki/ugm-indonesia/model/. Below is flowchart to get secondary structure metrics.

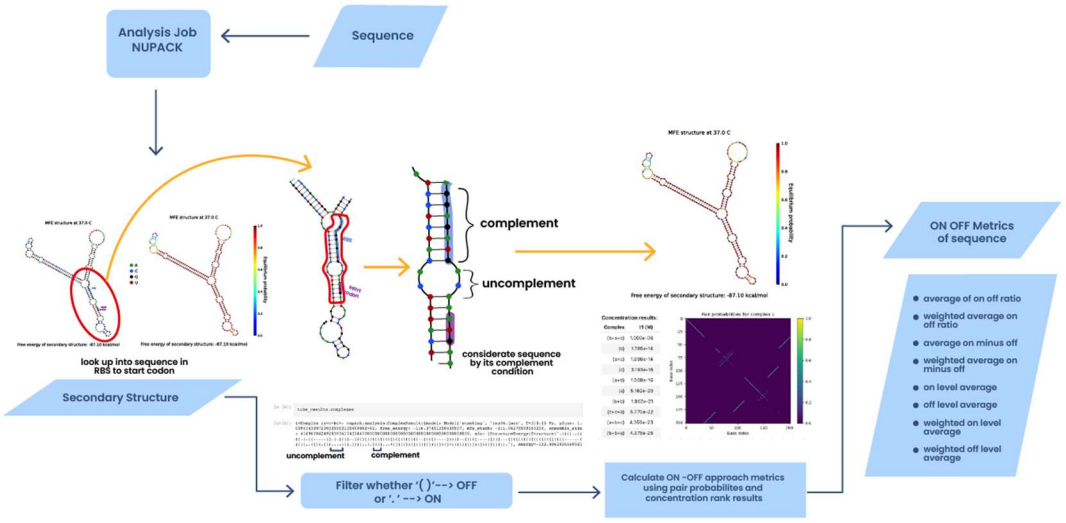


Figure 2. Flowchart of methods in obtaining secondary structure metrics

The metrics development needs to be validated into wet-lab data. Metrics that built upon LIRA context were then converted into RNA Toehold Sequence (RTS) context to be validated on Angenent-Mari et al (2020) dataset. The dataset flow process consists of pre-processing, NUPACK analysis job as RTS reaction simulation, nucleotide secondary structure approached metrics calculation with addition of epsilon metric (**Appendix A**), then we calculate the performance with R-squared, RMSE, and MSE metrics. The dataset processed only contains 1000 random sample rows due to limitation of computation and time to process all data. The epsilon metric was inspired to handle NaN value in ratio-ish approached metrics. Below the flowchart process of metrics validation.

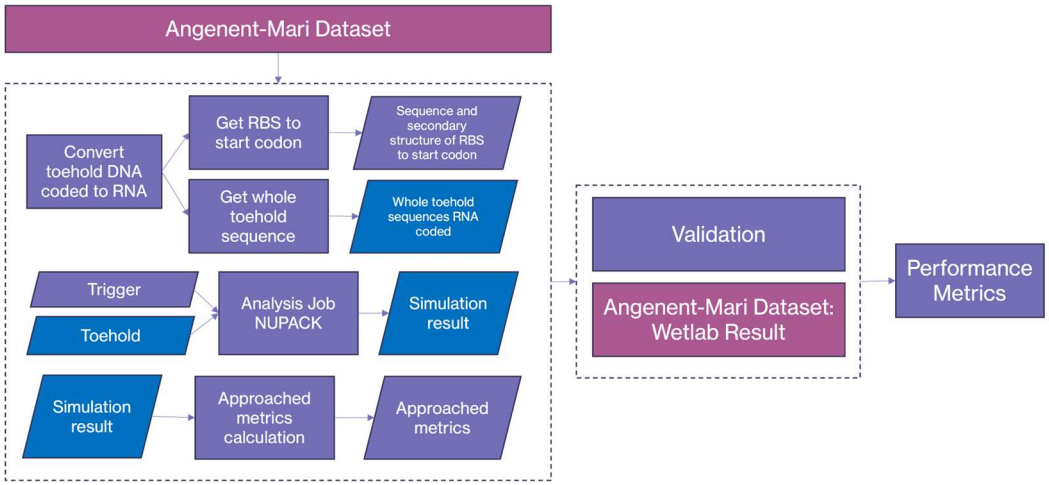


Figure 3. Validation Metrics Process, also documented here: github.com/iGEM-UGM/gogec-2024

Table. 1 Validation Score of Secondary Structure Metrics Over Wetlab Data

Secondary structure metrics	RMSE	MSE	R-squared	Secondary structure metrics	RMSE	MSE	R-squared
	On off ratio epsilon wet lab				Off level wetlab		
Epsilon avg	0.39943	0.15954	-1.45813	**a+b**	0.32630	0.10647	-1.18799
Epsilon a+b	0.36294	0.13172	-1.02950	**weighted**	0.32630	0.10647	-1.18799
	On off minus wetlab				On level wetlab		
Minus avg	0.52491	0.27553	-2.44555	**On level**	0.43360	0.18801	-0.77217
weighted	0.52491	0.27553	-2.44555	**weighted**	0.43360	0.18801	-0.77217
a+b	0.43845	0.19224	-1.40394	**a+b**	0.41621	0.17323	-0.63287

From the validation process, all metrics have an error under 0.5 (based on RMSE). We also got that the lowest error (RMSE) in ratio-ish and minus metrics are in epsilon a+b (0.36). Meanwhile, from overall metrics, the lowest error is on predicted off-level toehold (0.32).

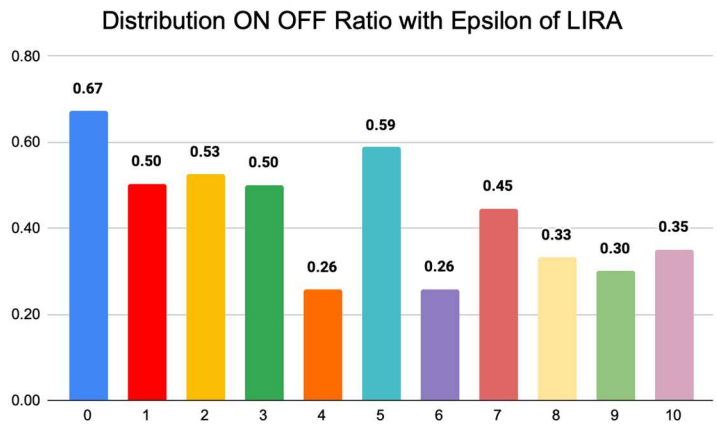


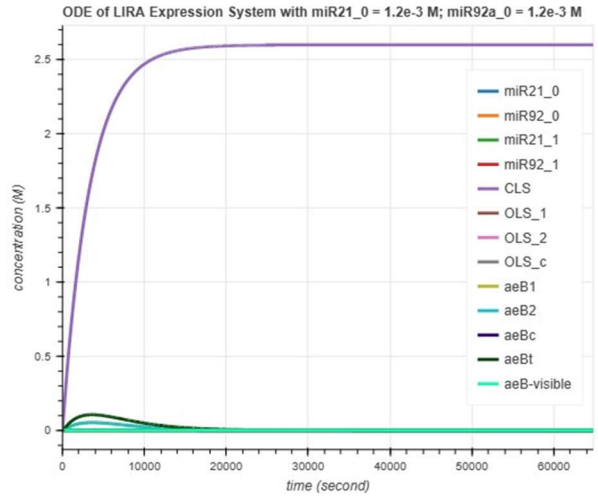
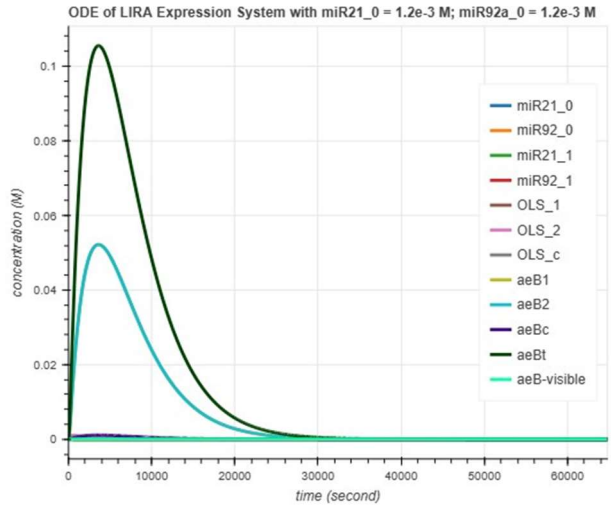
Figure 4. ON OFF Ratio distribution with epsilon of LIRA. vertical axis shows epsilon metric, horizontal axis shows sequence number

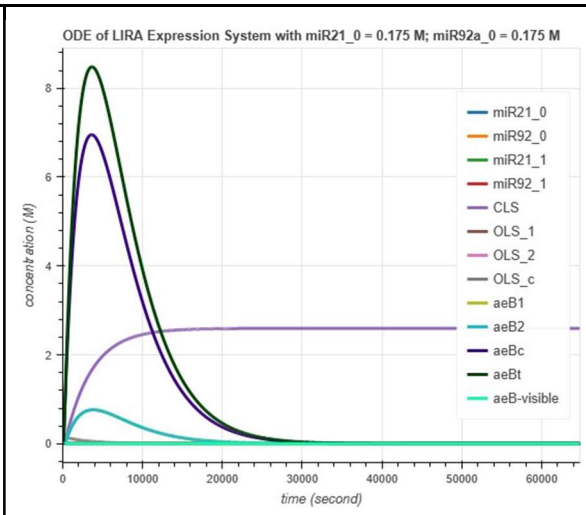
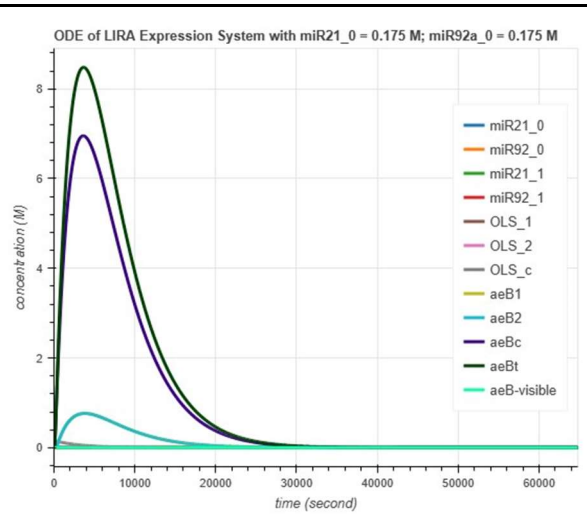
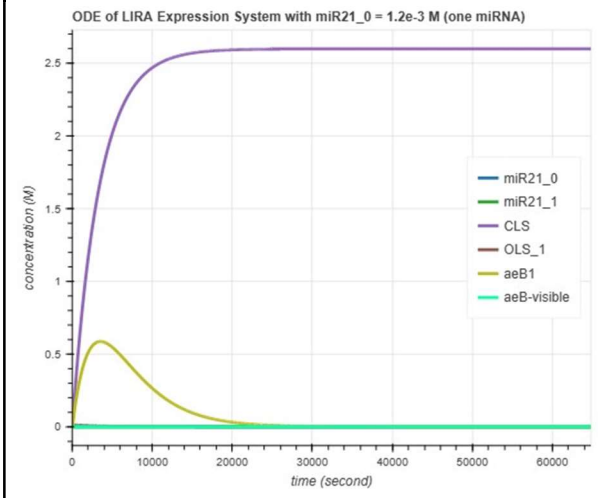
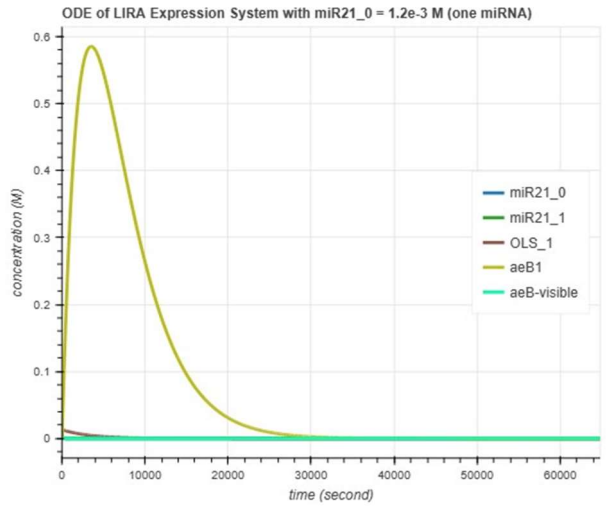
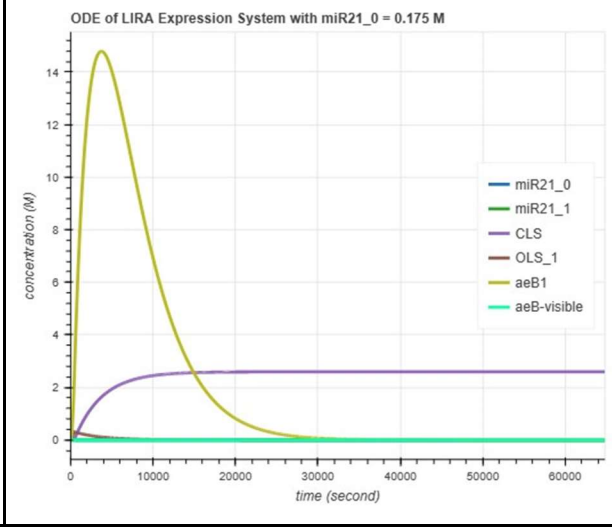
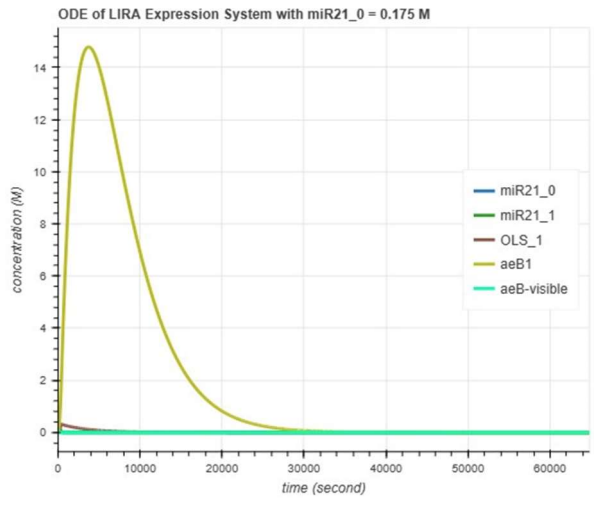
The on off ratio with epsilon average then applied to 11 sequences of LIRA that were constructed before using NUPACK to determine which sequence will be used. From the epsilon metric, the best sequence is 0th, 5th, 2nd, and so on.

2. Kinetic modeling of LIRA expression systems

We aim to simulate LIRA to understand how it responds to increased aeBlue concentration and extended aeBlue visibility by varying miRNA within a certain range. Since we lack miRNA concentration data in CRC patients at different stages, we simulate LIRA with initial miRNA concentrations spanning from the lowest normal value to the approximate short-term nonlethal limit on the human extracellular fluid ($1.2x10^{-3}$ M to $175x10^{-3}$ M, Hall and Hall 2020). Find the kinetic modeling steps and abbreviations in the attached document [Kinetic Modeling Equation for LIRA (1).pdf] and here are the results:

Table. 2 Kinetic Modelling Equation Results

No.	All Substance fluctuations	Zoom in (without CLS fluctuation)
1.		

2.		
3.		
4.		

The literature suggests that the transit time for food from the beginning of the colon to feces is approximately 12-24 hours (Sensoy, 2021). Some studies show varied transit times: ascending (9.5 ± 2.3

hours) and descending colon (5.5 ± 4.1 hours) are shorter than sigmoid-rectum (12.7 ± 2.1 hours) and transverse colon (4.2 ± 2.1 hours) (Tomita et al. 2011). Additionally, most colorectal cancer (CRC) cases occur in the sigmoid-rectum (59% in men, 46% in women). Our biodevice shows promise in detecting CRC. At the lowest input miRNA concentration (1.2x10-3 M), the visible aeBlue color lasts about 12.15 hours. When increased 10x, it extends to 14.53 hours, falling within the 12-24 hour range (Sensoy, 2021) and sigmoid-rectum transit time (12.7 ± 2.1 hours) (Tomita et al. 2011), the region with the highest cancer distribution. Predictably, our biodevice may detect CRC in feces, especially with elevated miRNA levels. Additionally, factoring in EcN cell population dynamics in future simulations may further extend visibility and allow adjustments based on EcN probiotic dosage.

3. Plasmid Design

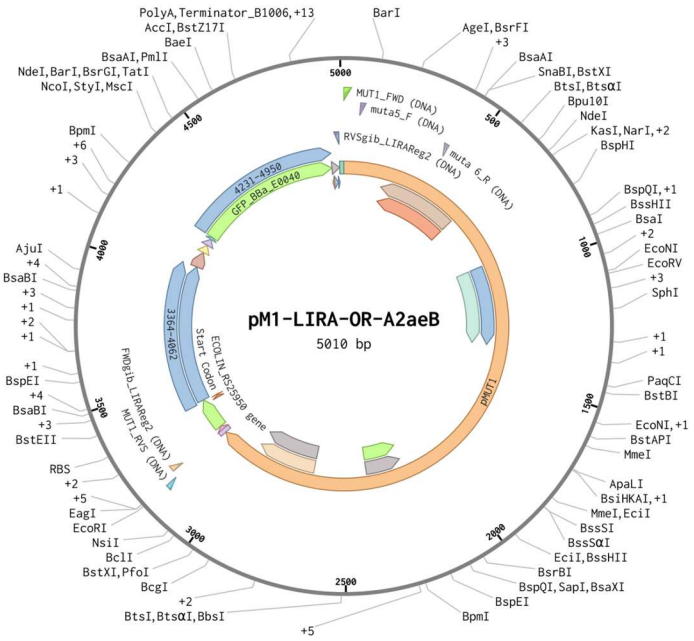


Figure 5. LIRA Plasmid expression vector design, based on the endogenous plasmid pMUT1 backbone.

In implementing the LIRA biodevice, we designed a plasmid as a vector of the LIRA System. We use an endogenous plasmid of EcN called pMUT1, which allows for excellent plasmid maintenance. In addition, our design allowed for plasmid selection to be done without using antibiotics, which reduces the possibility of unwanted antimicrobial resistance when implemented. The positive colonies were selected by their appearance, colored blue and then confirmed using PCR using a specifically designed primer.

Regarding the Biocontainment Plan Guideline 2018 by Center for Disease Control and Prevention, biocontainment development is associated with the laboratory procedures explained in training and equipment, such as a Biological Safety Cabinet, decontamination equipment, and gloves. For instance, an isolation for the member who handles the specific contaminant should be provided. Moreover, a killing switch design will be an improvement for naturally reducing the risk of contamination in the environment.

4. Kill Switch Design for Biocontainment

One of the most prevalent biocontainment strategies to prevent adverse environmental risks is kill switches. Therefore, we plan to implement 2 kill switch systems designed to (1) ensure bacterial death in

the outside environment and (2) ensure bacterial death when not localized to the human colon. To achieve those goals we designed one temperature-based kill switch and one chemical-based kill switch.

The first kill switch system, the temperature kill switch is based on a design developed and optimized by Striling et al. (2017). The switch consists of an antitoxin (ccdA) regulated by a weak constitutive promotor and a toxin (ccdB) regulated by a temperature-sensitive promoter (P_{cspA}) that is only activated when the environmental temperature drops below 37C. Therefore, when EcN, along with the feces, is defecated outside the human body, the low ambient temperature induces toxin production which overwhelms the weak antitoxin production thus leading to cell death. This system was proven to be highly robust with a 10^-5 survival ratio, better compared to similar temperature-based kill switches (Piraner et al., 2017). Apart from that, a modified version of the system was found to be evolutionarily stable for 140 generations (Stirling et al., 2017).

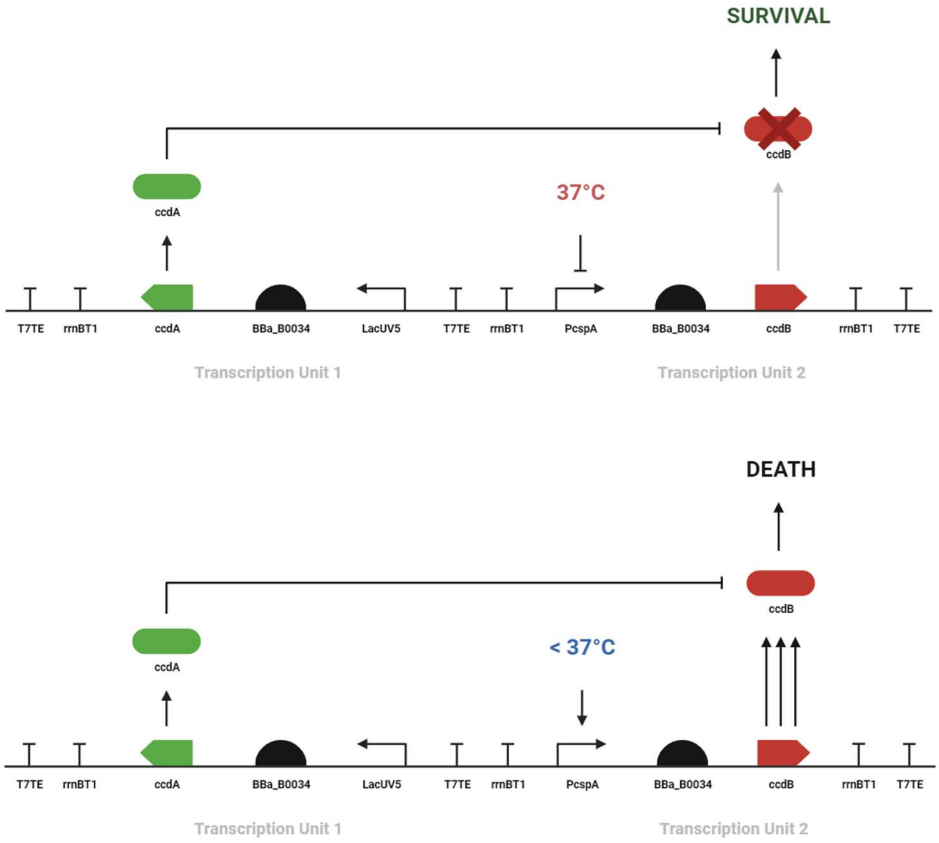


Figure 6. Temperature-based kill switch design using ccdA-ccdB antitoxin-toxin pairs and thermosensitive promoter P_{cspA} in permissive (top) and non-permissive (bottom) condition. Adapted from Stirling et al., (2017).

The second kill switch system uses a similar toxin-antitoxin system to the first kill switch system. The system consists of a toxin (parE) regulated by a weak constitutive promotor and an antitoxin (parD) regulated by a short chain fatty acid-sensitive promoter (P_{fliC}). The promotor fliC is developed by Kraus (2019) and only activated when short chain fatty acid is present in the environment. Short chain fatty acids (SCFA) are volatile fatty acids produced by the gut microbiota in the colon as fermentation products from food components that are unabsorbed/undigested in the small intestine Ríos-Covián et al. (2016). Therefore,

when EcN colonizes organs other than the colon, the absence of SCFA will stop the production of antitoxin thus leading to cell death.

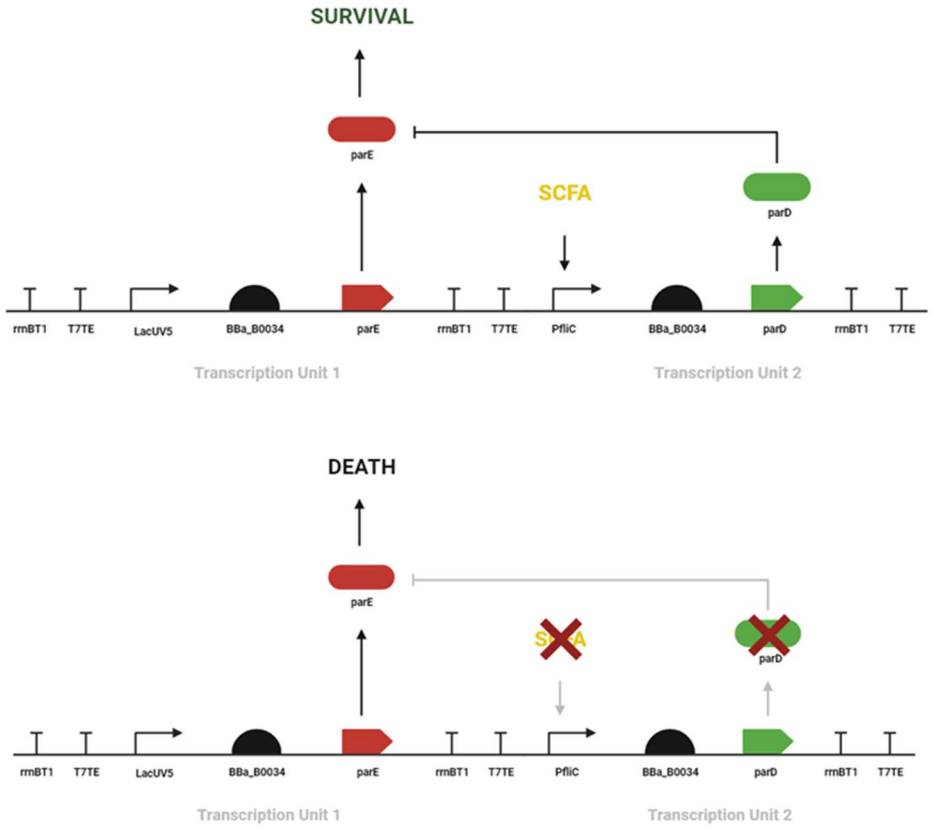


Figure 7. Chemical-based kill switch design using parD-parE antitoxin-toxin pairs and SCFA-sensitive promoter (P_{fliC}) in permissive (top) and non-permissive (bottom) condition.

D. CONCLUSIONS AND FUTURE WORK

1. Conclusions

- The epsilon metric with validation score 0.36 (RMSE) potential to predict ON OFF performance of LIRA as design consideration. Shows the 0th sequence is superior (0.67).
- Our biodevice shows promise in detecting CRC, visible aeBlue color lasts about 12.15 hours at the lowest input miRNA concentration (1.2x10-3 M).
- A well-maintained safe expression vector was also designed using the endogenous EcN pMUT1 plasmid backbone.
- A double temperature-based and chemical-based kill-switch system was also design to ensure robust biocontainment of engineered EcN inside the colon

2. Future Work

- Registering product with national authorization to ensure the legality of product marketing.
- Facilitating the availability of product for purchase in pharmacies or through government distribution.
- Scaling up to an industrial scale.

- Conducting a cost-effectiveness analysis of an affordable CRC screening method. According to the incremental cost-effectiveness ratio (ICER) analysis, ColDBlu has been proven to be cost-effective with a per-product cost of Rp14,500.

References

Alizadeh S, Esmaeili A, Omidi Y. Anti-cancer properties of Escherichia coli Nissle 1917 against HT-29 colon cancer cells through regulation of Bax/Bcl-xL and AKT/PTEN signaling pathways. Iranian Journal of Basic Medical Sciences. 2020 Jul;23(7):886.

Angenent-Mari, N.M., Garruss, A.S., Soenksen, L.R. *et al.* A deep learning approach to programmable RNA switches. *Nat Commun* 11, 5057 (2020).

Centers for Disease Control and Prevention. 2018. Select Agents and Toxins Biosafety/ Biocontainment Plan Guidance. Division of Select Agents and Toxins.

Chiang CJ, Huang PH. Metabolic engineering of probiotic Escherichia coli for cytolytic therapy of tumors. Scientific reports. 2021 Mar 12;11(1):5853.

Hall, J.E. and Hall, M.E., 2020. Guyton and Hall textbook of medical physiology e-Book. Elsevier Health Sciences.

Kan A, Gelfat I, Emani S, Praveschotinunt P, Joshi NS. Plasmid vectors for in vivo selection-free use with the probiotic E. coli Nissle 1917. ACS synthetic biology. 2020 Dec 10;10(1):94-106.

Ma D, Li Y, Wu K, Yan Z, Tang AA, Chaudhary S, Ticktin ZM, Alcantar-Fernandez J, Moreno-Camacho JL, Campos-Romero A, Green AA. Multi-arm RNA junctions encoding molecular logic unconstrained by input sequence for versatile cell-free diagnostics. Nature biomedical engineering. 2022 Mar;6(3):298-309.

Okugawa Y, Toiyama Y, Goel A. An update on microRNAs as colorectal cancer biomarkers: where are we and what's next?. Expert review of molecular diagnostics. 2014 Nov 1;14(8):999-1021.

Sensoy, I., 2021. A review on the food digestion in the digestive tract and the used in vitro models. Current research in food science, 4, pp.308-319.

Stirling, F., Bitzan, L., O'Keefe, S., Redfield, E., John, Way, J.C. and Silver, P.A. (2017). Rational Design of Evolutionarily Stable Microbial Kill Switches. Molecular Cell, [online] 68(4), pp.686-697.e3. doi:https://doi.org/10.1016/j.molcel.2017.10.033.

Tomita, R., Igarashi, S., Ikeda, T., Sugito, K., Sakurai, K., Fujisaki, S., Koshinaga, T. and Shibata, M., 2011. Study of segmental colonic transit time in healthy men. Hepato-gastroenterology, 58(110-111), pp.1519-1522.