Phenotype on Genotype: Long-read sequencing based High-throughput Phenotype Analysis

Seong-Kun Bak[1, 2*], Wonjae Seong[1, 2*], A-young Park[1, 3], Eugene Rha[1], Hyewon Lee[1], Kilkoang Kwon[1], Haseong Kim[1, 2], and Seung-Goo Lee[1, 2]

[1] Synthetic Biology and Bioengineering Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 34141, Republic of Korea

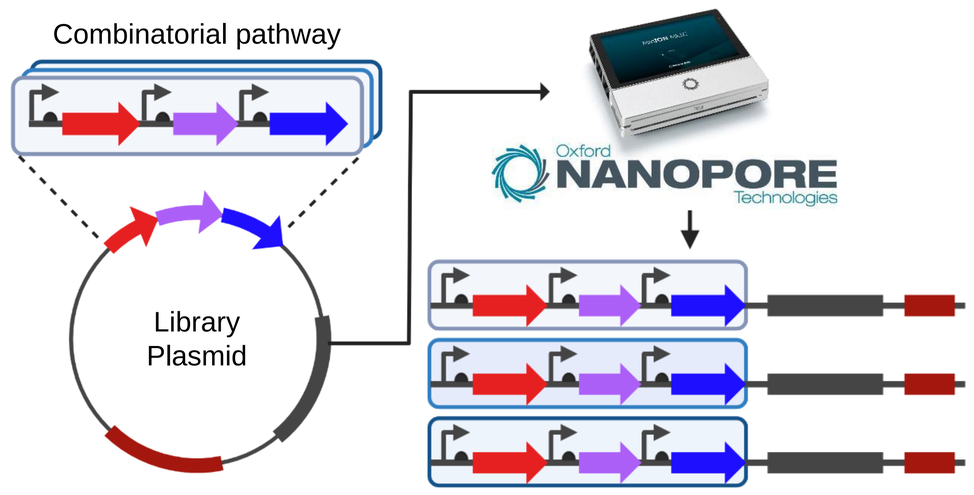
[2] Department of Biosystems and Bioengineering, KRIBB School of Biotechnology, University of Science and Technology, Daejeon 34113, Republic of Korea

[3] Department of Chemical Engineering and Applied Chemistry, Chungnam National University, Daejeon 34134, Republic of Korea

ABSTRACT

The essential elements to characterize, optimize and engineer genetic circuit are the multiple genotypes of genetic mutants and their respective phenotypes.
Currently, there are several high-throughput methods for measuring phenotype and genotype. However, they are measured independently using different methods and it requires substantial effort to connect and convert them to analytic data.
In our studies, we introduce a long-read sequencing based high-throughput method to measure phenotype and genotype simultaneously. **Long-read sequencing, can distinguish each genotype in DNA library and obtain all information of a single DNA in a concatenated state.** Using this characteristic, this study present two examples that obtain phenotype on genotype.
First, we tracked the DNA ratio of library in cultured cells. With this data, **we could calculate relative growth rate for individual genotype**. Second, we used T7 RNAP linked deaminase based DNA recorder system as a reporter of transcription factor-based biosensor to **represent gene expression rate by mutation rate in specific region**. Finally, we confirmed their feasibility by showing high correlation when compared to the existing methods such as growth rate.

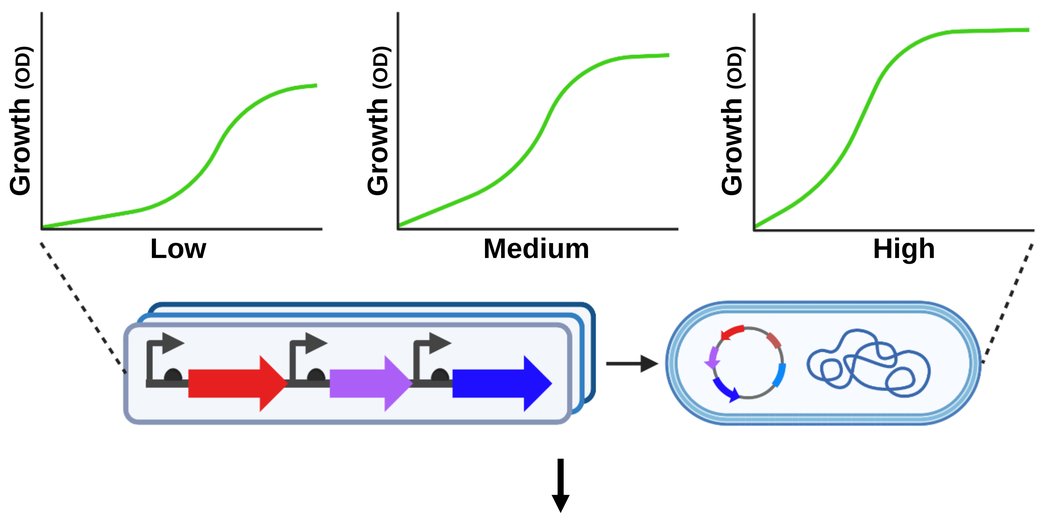
Why Long-read sequencing ?

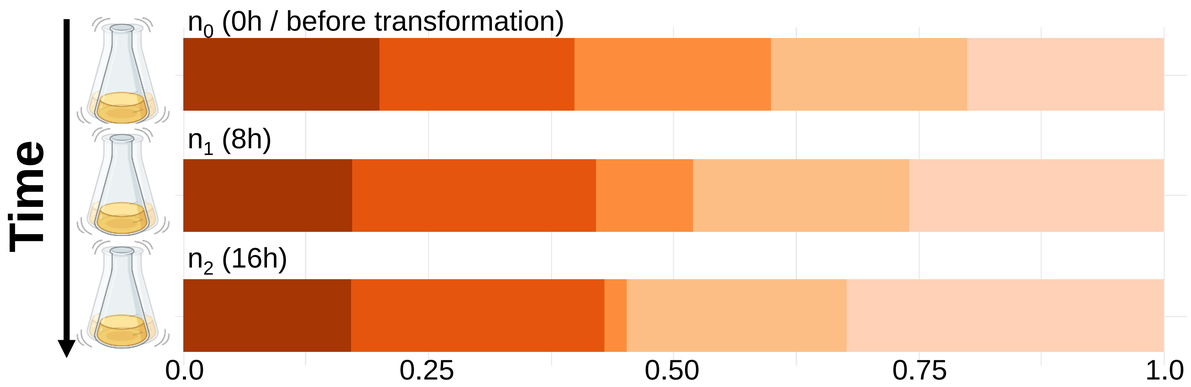


- Long-read sequencing can distinguish and analysis each genotype simultaneously in DNA Library.

GROWTH RATE

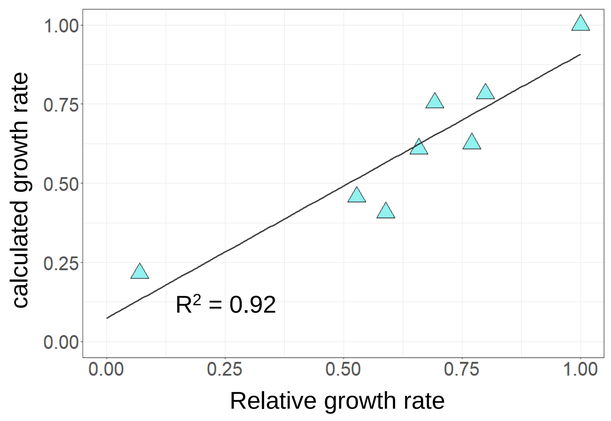
(A) Combinatorial pathway Library with Different growth rate





- Tracking DNA ratio of library while culturing.

(C) Comparing with existing method



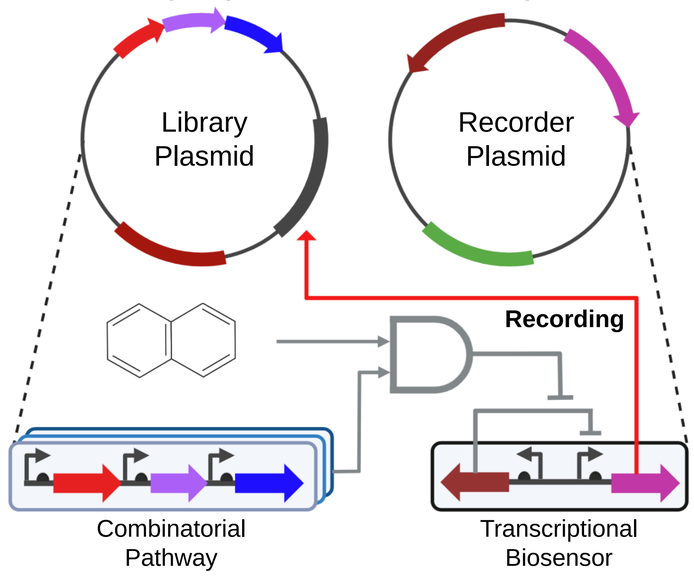
Using library's OD and ratio (in n_0, n_2) calculated each growth rate.

It showed a high correlation compared with conventional growth rate.

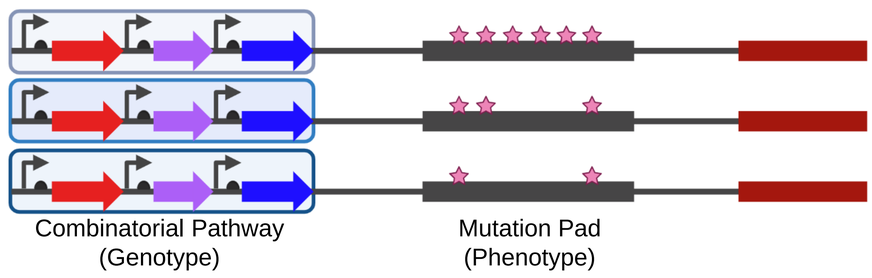
Fig 1. Growth rate estimation based on long-read sequencing.
While culturing combinatorial pathway library containing cells, the DNA ratio for individual genotype was confirmed through long-read sequencing. And we estimated individual growth rates using the ratio and final OD of library cells.
(A) The scheme of combinatorial library with different growth rate.
(B) Calculated DNA ratio by counting sequencing read in three time points (0h, 8h, 16h).
(C) The result of comparing the estimated growth rate by sequencing to conventional method.
* Among the sequencing reads, only 90% or more querycover reads were selected and calculated.

EXPRESSION RATE

(A) DNA recording reporter with transcriptional biosensor.



(B) Read genotype and phenotype on DNA simultaneously



- Convert pathway's expression rate to number of mutations on target sequence (mutation pad).
- Then, can read genotype(pathway) and phenotype(mutation pad) simultaneously on unbroken read.

Fig 2. Gene expression rate estimation based on long-read sequencing.
Using DNA recorder as a reporter in transcriptional biosensor, record expression rate on specific region of plasmid as mutation. With long-read sequencing can read mutation rate (phenotype) linked to each combinatorial pathway (genotype).
(A) The plasmid construction to calculate expression rate by sequencing.
(B) Simple diagram of read sequence, the genotype representing region and phenotype representing region are linked.

CONCLUSIONS & DISCUSSION

- By analyzing the results of long-read sequencing, it was confirmed that can distinguish each genotype in library and analysis was possible even there is sequencing error.
- With only sequencing, can estimate high accuracy growth.
- Although could not achieve a sufficient mutation rate for analysis using T7 RNAP based recorder, it was confirmed that long-read sequencing based method can distinguish a certain number of mutation or more in single read units.
- Using the method introduced here, multiple genotypes and phenotypes can be measured with one method.

ACKNOWLEDGEMENTS

This work was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science & ICT (2018M3A9H3024746) and the KRIBB Research Initiative Program(KGM5402221).

REFERENCES

1. Vladimir Potapov, Gregory J S Lohman. (2018). *Nucleic Acid Res*. July 13;46:e79.
2. Park H, Kim S. (2021). *Nucleic Acids Res.* Apr 6;49(6):e32.
3. Choi SL, Lee SG. (2014). *ACS Synth. Biol*. 3,3, 163-171.
4. Created with BioRender.com