Phenotype MicroArray Analysis of *Escherichia coli* K-12: An *in silico* Approach

Hiroaki Gobara\(^1\)  
Tomoya Baba\(^1\)  
Md. Altaf-Ul-Amin\(^1\)  
Masanori Arita\(^2,3\)  
Ken Kurokawa\(^1\)  
Shigehiko Kanaya\(^1\)  
Hirotada Mori\(^1,2\)

\(^{1}\) Nara Institute of Science and Technology (NAIST), 8916-5 Takayama, Ikoma, Nara 630-0101, Japan  
\(^{2}\) Inst. Adv. Bioscience, Keio University, Baba-machi, Tsuruoka, Yamagata 997-0035, Japan  
\(^{3}\) Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo and PRESTO, JST, Kashiwanoha, 5-1-5 CB05, Kashiwa 277-8561, Japan

**Keywords**: Phenotype MicroArray, KEIO collection, t-value, p-value

1 **Introduction**

Phenotype MicroArray (PM) is a screening technology that was developed by Biolog Inc. It is an integrated system of cellular assays, instrumentation, and bioinformatic software for high-throughput screening of cells, and is available for bacterial and fungal cells [1]. The technology and the testing process are shown in Figure 1. Biolog reconfigures a wide range of phenotypic tests into sets of arrays. Each well in the array is designed to test a different phenotype. Wells are prepared for a total of 1,920 conditions. A standardized cell suspension is inoculated into the wells of the MicroArray, thereby testing thousands of phenotypes at once. The MicroArray is then incubated, typically for 24 hours. PM uses Biolog's patented redox chemistry, employing cell respiration as a universal reporter.

If the phenotype is strongly "positive" in a well, the cells respire actively, reducing a tetrazolium dye and forming a strong color (Figure 1). If it is weakly positive or negative, respiration is slowed or stopped, and less color or no color is formed. The redox assay provides for both amplification and precise quantitation of phenotypes. Incubation and recording of phenotypic data is performed by the patented OmniLog® instrument which captures a digital image of the MicroArray several times each hour and stores the quantitative color changes into computer files. The computer files can be displayed in the form of kinetic graphs. Thousands of phenotypes are monitored simultaneously by the OmniLog® and up to 450,000 data points can be generated in one 24-hour run. PM can monitor, either directly or indirectly, most known aspects of cell function [2, 3]. Purpose of the present study is to quantify the effect of mutants to phenotype. To handle this problem, we combine PM with gene deletion mutants of *Escherichia coli*.

2 **Material and Method**

The PM data for the 120 strains were obtained from Biolog Inc. the 120 strains of Escherichia coli were selected from the KEIO collection (Baba T. et al, in preparation): the systematic construction of single gene knock out mutants of all gene/ORFs including putative ones previously performed by our group using the Wanner’s method (Datsewnko and Wanner, 1997). Experiments were conducted using each strain for a total of 1,920 conditions to generate the PM data. One PM data consist of consecutive 94 observation points for about 24 hours. First, the obtained data underwent the correction process. Original data were smoothed by

![Figure 1: Phenotype MicroArray [3]](image-url)
taking an average of consecutive 5 points. Subsequently, the largest gradient for 13 consecutive points was computed as the culture’s estimated growth slope. For each mutant strain, the maximum gradients for 1,920 different cultures were computed as a vector of the same length. From the estimated gradient vectors, all-against-all correlations were computed using the following formula (Pearson correlation; x, y: any two vectors).

\[
    r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} = \frac{s_{xy}}{\sqrt{s_{xx} \cdot s_{yy}}}
\]

Next, the scatter plot was calculated for wild type and mutant slope data. Here we used only those mutant slope data that have high correlation with calculated wild type slope data. From the scatter plot, outliers were calculated by the t-test with the confidence limit of 95% and with infinity as the degree of freedom. That is, when a calculated t-value is higher than 1.960 (df = \infty), the corresponding slope data were considered an outlier. Subsequently, the p-value was computed as follows.

\[
y = Y \pm t \cdot s \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}} \\
s = \sqrt{\frac{S_{yy} - bS_{xy}}{n - 2}} \\
p = \int_{t}^{\infty} f(u)du
\]

3  Result and Discussion

Phenotype MicroArray is useful for screening technology. In the present study, we performed quantitative analysis of time-series data obtained from Phenotype MicroArray to quantify differences of growth in 1,920 conditions between targeted mutant and wild type, and then compute their correlations. We could obtain some expression profiles and growth curves that had significant similarities. From KEIO collection, we selected 120 mutants including 77 deletion mutants of metabolic genes. Among them, 26 mutants showed > 0.8 correlations with the wild type. Such mutants, however, also exhibited data that had locally significant differences among the 1,920 experimental conditions (p-value < 0.05). Some differences may be attributed to experimental noises, but most of local difference in growth conditions suggests that the genes deleted in this study regulated signal cascade in particular growth phase. The present results contribute to investigate gene function in *Escherichia coli* in its growth phases. Furthermore, comprehensive analysis of Phenotype MicroArray Data obtained in the present study together with genome information such as GenoBase (http://ecoli.aist-nara.ac.jp/) could establish relationship between genes and metabolic pathways.

4  References

