The International Journal of Biostatistics

Volume 6, *Issue* 1 2010 *Article* 29

Statistical Methods for Comparative Phenomics Using High-Throughput Phenotype Microarrays

Joseph Sturino, *Texas A&M University*
Ivan Zorych, *Texas A&M University*
Bani Mallick, *Texas A&M University*
Karina Pokusaeva, *Texas A&M University*
Ying-Ying Chang, *Texas A&M University*
Raymond J. Carroll, *Texas A&M University*
Nikolay Bliznuyk, *Texas A&M University*

Recommended Citation:
Sturino, Joseph; Zorych, Ivan; Mallick, Bani; Pokusaeva, Karina; Chang, Ying-Ying; Carroll, Raymond J.; and Bliznuyk, Nikolay (2010) "Statistical Methods for Comparative Phenomics Using High-Throughput Phenotype Microarrays," *The International Journal of Biostatistics*: Vol. 6: Iss. 1, Article 29.
DOI: 10.2202/1557-4679.1227

Statistical Methods for Comparative Phenomics Using High-Throughput Phenotype Microarrays

Joseph Sturino, Ivan Zorych, Bani Mallick, Karina Pokusaeva, Ying-Ying Chang, Raymond J. Carroll, and Nikolay Bliznuyk

Abstract

We propose statistical methods for comparing phenomics data generated by the Biolog Phenotype Microarray (PM) platform for high-throughput phenotyping. Instead of the routinely used visual inspection of data with no sound inferential basis, we develop two approaches. The first approach is based on quantifying the distance between mean or median curves from two treatments and then applying a permutation test; we also consider a permutation test applied to areas under mean curves. The second approach employs functional principal component analysis. Properties of the proposed methods are investigated on both simulated data and data sets from the PM platform.

KEYWORDS: functional data analysis, principal components, permutation tests, phenotype microarrays, high-throughput phenotyping, phenomics, Biolog

Author Notes: Zorych and Bliznyuk were supported by the Texas A&M Postdoctoral Training Program of the National Cancer Institute (CA90301). The research of Carroll was supported by a grant from the National Cancer Institute (CA57030). Carroll and Mallick were also supported by Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST). The research of Sturino was supported by the United States Department of Agriculture, Cooperative State Research, Education and Extension Service, Hatch project TEX 09436. Acquisition of the Biolog Omnilog Phenotype Microaarray was supported by the State of Texas Permanent University Fund with matching funds from Texas AgriLife Research and Texas A&M University. In addition, the authors would like to thank the editor and anonymous reviewers for their valuable comments.

1 Introduction

Phenomics is the systematic study of global cellular phenotypes that arise as a function of genotype (or metagenotype) and its environmental context (Gowen and Fong, 2009). Differentiation of biological systems as a function of observable phenotype predates the discovery of their molecular components, which includes DNA (and its systems biology sub-discipline, genomics), epigenetic heritability (epigenomics), RNA (transcriptomics), proteins (proteomics), and metabolites (metabolomics). Nevertheless, due to the repetitive and labor intensive nature of phenotypic studies, cellular phenomics has struggled to become a vibrant functional discipline (Bochner, 2009; Joyce and Palsson, 2006).

Assay miniaturizations coupled with process automation have proven to be foundational advances that have enabled the rational design and implementation of high-throughput phenotyping (HTP) platform chemistries. Automated liquid handling systems can be used to accurately and precisely prepare low-volume microplate-based assays, while high-capacity and temperature-controlled automated microplate readers enable effectively parallel data collection (Gabrielson et al., 2002; Bochner, 2003). These technological advancements have paved the way for the development and commercialization of standardized platform chemistries for HTP; see Gowen and Fong (2009) for a review.

The Phenotype MicroArray (PM) platform for HTP was developed by Bochner et al. (2001) and commercialized by Biolog, Inc., whose web site is http://www.biolog.com/; for reviews, see Bochner et al. (2009) and Bochner (2003a). The complete Phenotype MicroArray for microbial cells is comprised of twenty pre-formulated 96-well microplates (PM1 to PM20). When used together, they enable researchers to simultaneously assay up to 1,920 different cellular phenotypes as a function of time (i.e., kinetic response). Individual PM microplates contain a large and heterogeneous collection of functionally-related chemical compounds or combinations thereof (e.g., up to 96 per PM microplate). These compounds may serve as a source of carbon (PM1-2), nitrogen (PM3, PM6-PM8), and phosphorus or sulfur (PM4). Other PM microplates are used to determine sensitivity to environmental stresses, such as ions or osmolytes stress (PM9), pH (PM10), and chemical agents (PM11-20) (Bochner, 2009).

Using this platform, microbial strains may be assayed individually or in parallel (e.g., a wild-type strain versus an isogenetic mutant) and subsequently compared for quantitative phenotypic differences as a function of time (i.e., kinetic activity). Many microplate-based HTP platforms monitor the accumulation of biomass directly as a function of time, which necessitates

cellular growth and division. In contrast, the PM platform employs a universal colorimetric reporter system to monitor cellular metabolism that is effective even in the absence of the accumulation of biomass. This colorimetric reporter indirectly measures cellular metabolism by directly measuring the irreversible reduction of tetrazolium-based dyes (colorless) to formazan (purple, Gabrielson et al., 2002; Bochner, 2003a). Conceptually, maintenance of a respiration-competent state requires fewer physiological systems than does the maintenance of a replication-competent state. As such, this platform may afford a distinct advantage over cultivation based platforms for the study of microorganisms that may be viable, but not cultivatable *ex vivo*, which includes most of the microorganisms present in the gastrointestinal tract of humans and animals (Savage, 1977; Xu and Gordon, 2003).

Standardized commercial chemistries for high-throughput phenotyping (HTP), such as the Biolog Phenotype Microarray (PM) platform, provide a universal platform to facilitate meta-analysis of phenomics data generated from the query of disparate biological systems.

These recent advancements have resulted in the generation of novel high-dimension phenomic data sets. The OmniLog PM system software is used to visualize phenotype curves and provides several basic analytical functions. For example, the PM Kinetic Analysis Module can be used to generate a mean kinetic phenotype curve and to (optionally) amend it (e.g., subtract background signal, crop/trim early or late time points, and several others.). The PM Parametric Analysis Module calculates summary values for each mean phenotype curve (e.g., area under the curve, min/maximum signal intensity, maximum slope, lag time, etc.), which enables two microorganisms to be compared. For example, the software allows users to select an ad hoc threshold (e.g., fixed cut-off k) that distinguishes wells that show "striking" phenotypic relative differences between microorganisms with respect to a given summary value (e.g., a k-fold change). The number of wells that differ under such an approach depends on the summary value and the stringency of the of selected threshold. Analysis of PM data using ad hoc summary values without statistical support has been reported extensively, for example (Bochner, et al., 2001; Bochner 2003 (2); Zhou, et al., 2003; Mukherjee, et al., 2006). Although this approach may be used judiciously to guide biological research, ad hoc experimenter-selected thresholds have much the same flavor of the 2-fold expression changes that used to occur in microarrays, and in that context the ad hoc threshold has given way to robust statistical analysis (Slonim, 2002).

As such, there is an urgent need to design robust statistical methodologies to interrogate these data and to enable sound biological inferences to be made therefrom. In this paper, we propose several simple yet effective hy-

DOI: 10.2202/1557-4679.1227

pothesis testing frameworks to compare phenomic data that were generated using the PM platform for high-throughput phenotyping.

The paper is organized as follows. In Section 2, we describe the experiment undertaken. Section 3 gives the methods used. Section 4 gives the results of simulation studies, while Section 5 describes the analysis of the experimental data. Section 6 has concluding remarks.

2 Experimental Design

Sodium (Na+)/proton (H+) antiporter proteins are critical for maintaining intracellular pH, cell volume, and osmotic homeostasis (Padan et al., 2001). The *Escherichia coli* (*E. coli*) *nhaA* gene encodes the NhaA Na+/H+ antiporter, which protects the cell from sodium ion toxicity in high-sodium environment (Padan et al., 1989). The *Escherichia coli* K-12 BW25113 (relevant genotype: $nhaA^{+}$) and its isogenetic *nhaA*-deletion derivative *E. coli* ECK0020 (relevant genotype: $\Delta nhaA$) were obtained from the Keio Collection (Baba et al., 2006) curated by the Escherichia coli Genetic Stock Center at Yale University.

All PM procedures were performed as indicated by the manufacturer. In brief, strains were cultivated at 37°C on $R2A$ agar. A sterile cotton swab was used to transfer bacteria from the agar surface into IF-0a GN basal inoculation fluid until a cell density of 85% transmittance was reached using a turbidimeter. The resultant suspension was diluted 1 : 10 in IF-0a GN that incorporated tetrazolium dye mix D to a final concentration of 1% (vol/vol) and $20mM$ sodium succinate (Sigma).

The resultant BW25113 and ECK0020 suspensions were separately inoculated (100 μL/well) into the PM9 microplate, which challenges cellular responses to various osmotic and ionic stresses, including sodium chloride at concentrations ranging from 1% (w/v) (PM9 well A1) to 10% (w/v) (PM9 well A10). PM9 microplates were incubated at 37°C in an OmniLog PM microplate reader; an integral charge-coupled-device camera was used to record colorimetric signal intensity, represented as arbitrary OmniLog Units (OL), every 15 min. As such, the resultant assays provided both quantitative and kinetic information of the queried biological systems.

Strains BW25113 and ECK0020 were assayed in parallel in each of four independent biological replicates.

Unlike microarray data (Slonim, 2002), there has been very little discussion regarding the necessity of data normalization for for comparative phenomic analysis using PM. The only reference that we have found that argues explicitly for normalization in some cases is Borglin (2009). The notion here is

that Biolog readings for two organisms may be different either because of differences in metabolism or because of different growth rates. In the later case, one may try to normalize with respect to growth rate. The author concludes (page 16) that normalization is not necessary for PM plates. At the same time, the manufacturer stresses "high-reproducibility" of the results, and at least one paper confirms this for the substrates from the PM5 plate (Figure 1, Johnson, et al., 2009). See the data examples in Section 5 for more discussion.

There are other important issues related to normalization. Time 0 is not actually physical time zero. There is actually a 15-30 minute preprocessing stage where all the plates are inoculated and then loaded into the reader. Thus, Time 0 is the time at which reading commences, and one would not expect that the phenotype curves start at 0. Hence, some forms of normalization may hide initial growth rates. To get some idea of what the data look like, in Figure 1 we plot the results from 16 wells from the PM9 plate.

3 Statistical Approaches for Phenotype Comparison

The metabolic activity of both microorganisms was measured for each phenotype assay well over regular intervals of time. As noted above, time $t = 0$ means the time at which reading of the plate commences, and not the time when the plates become inoculated. Each observation of a well may be represented as a function $Y_{ij}(t)$, which reflects measurements of oxidation level of the $i = 1, ..., I = 2$ organisms on the $j = 1, ..., J$ biological replicate for time t: if technical replicates are also available, $Y_{ij}(t)$ refers to the mean of them. The population mean of the functions across the microplates in organism i is $E\{Y_{ij}(t)\} = f_i(t)$. Time t is measured every 15 minutes during the course of the experiment.

We propose two approaches to testing the null hypothesis that the two organisms have exchangeable phenotypes, which we abbreviate as stating that $H_0 : f_1(t) = f_2(t)$ for all t, with alternative hypothesis that $H_1 : f_1(t) \neq f_2(t)$ for all t. The first method we propose is a nonparametric permutation approach applied to the mean curves, median curves, and to the areas under the mean curves. The second approach employs functional data analysis methods.

DOI: 10.2202/1557-4679.1227

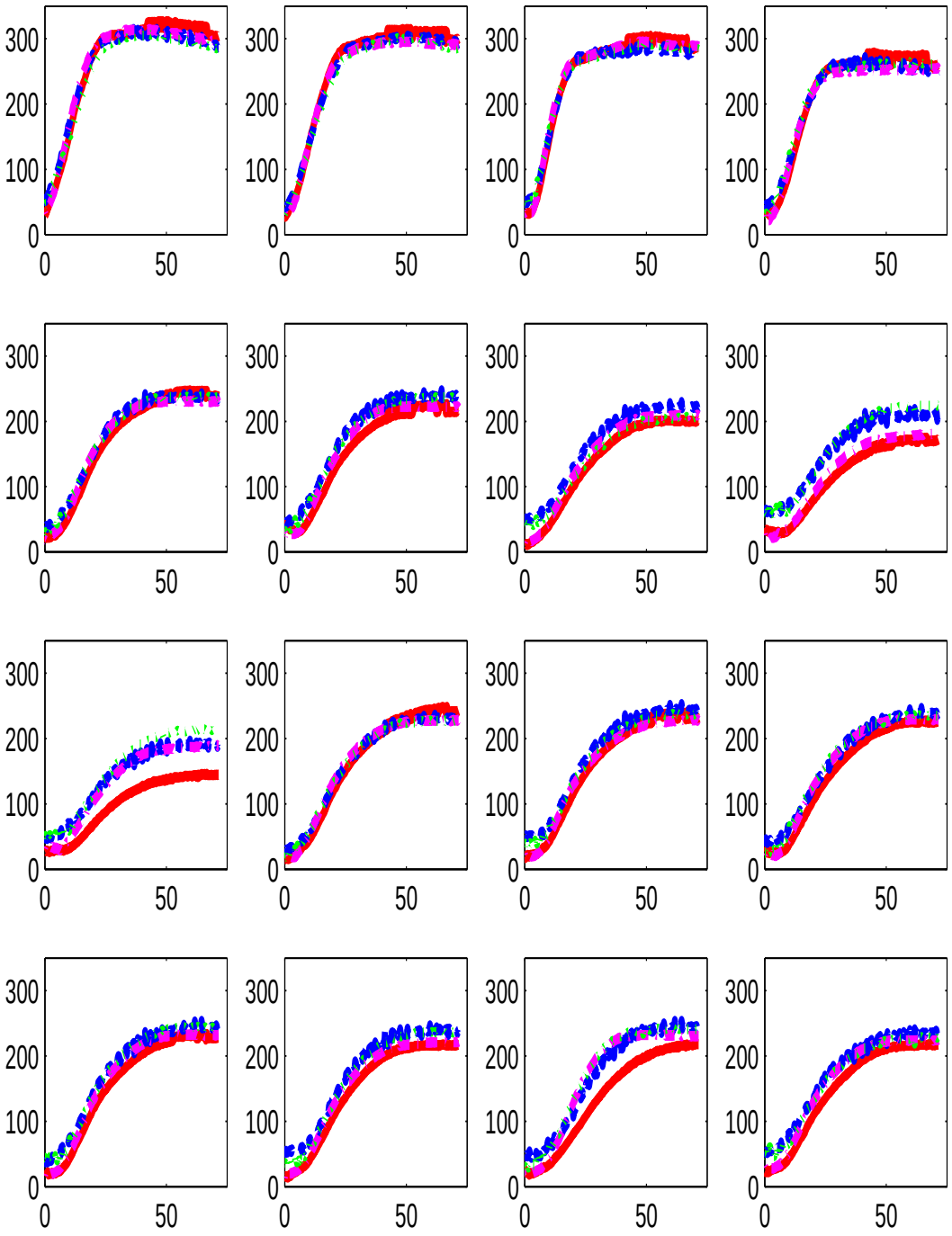

Figure 1: Sixteen wells from the PM9 plate. Plotted are the four biological replicates in the mutant strain. The horizontal axis is the time since the wells begin to be read. The vertical axis is the mean optical density.

3.1 Permutation Approach

The permutation-based method for comparing two groups of curves works as follows. We define $Y_{ij}(t)$ to be an observed phenotype curve for organism $i = 1, ..., I = 2$ and replicate $j = 1, ..., J$, and let the sample mean curve in organism i be

$$\widehat{f}_{i\cdot}(t) = J^{-1} \sum_{j=1}^{J} Y_{ij}(t). \tag{1}$$

Our first proposed test statistic is based upon the overall squared difference between the two mean curves, i.e.,

$$S = \int \{\widehat{f}_{1\cdot}(t) - \widehat{f}_{2\cdot}(t)\}^2 dt.$$

In practice, the curves are only observed at fixed times, in which case the intergral in (1) is replaced by the average over the observation times.

One can test the null hypothesis that the two organisms have exchangeable phenotypes using permutation testing, because under the null hypothesis, all permutations of the observed biological replicate data have the same distribution. To implement permutation testing, we use the standard device of taking all permutations of the indices (i, j) in such a way that there are exactly J functional observations for each of $i = 1, 2$. Let the number of unique permutations be B. Then for each of the $b = 1, ..., B$ unique permutations, recompute the test statistic (1) and record it as $\mathcal{S}_b$. Then the p-value for the test is

$$p = B^{-1} \sum_{b=1}^{B} I(\mathcal{S}_b > \mathcal{S}).$$

If there are $J \geq 4$ biological replicates, there are a sufficient number of unique permutations for this test to make sense, exactly 70 if $J = 4$.

Besides the permutation test that is based on statistic (1), we also consider two other tests. Instead of mean curves (1), one may employ median curves defined as $\widetilde{f}_{i\cdot}(t) = \text{median}_{j \in J}\{Y_{ij}(t)\}$. This approach may be of interest when functional observations, $Y_{ij}(t)$, exhibit high levels of skewness. Using mean absolute error, the test statistic then changes to $S_M = K^{-1} \sum_{k=1}^{K} |\widetilde{f}_{1\cdot}(t_k) - \widetilde{f}_{2\cdot}(t_k)\}|$, and the test is carried out in the same way as the mean curves-based test.

Another approach uses the area under the mean curve of each group. Let $\overline{f}_{i\cdot}$ be the area under the mean curve of group i, defined as $\overline{f}_{i\cdot} = \int \widehat{f}_{i\cdot}(t) dt$. Then test statistic S_A is defined as the absolute difference between areas under the mean curves, i.e., $S_A = \mathsf{I}\overline{f}_{1\cdot} - \overline{f}_{2\cdot}\mathsf{I}$.

DOI: 10.2202/1557-4679.1227

3.2 Functional Data Analysis Approach

Another method of testing employs a functional data analysis (FDA) approach (Ramsay and Silverman, 2006).

- Based on a parsimonious basis function approximation, functional principal component analysis allows us to reduce two functions to a set of a few coefficients.
- Under the null hypothesis that the two organisms have exchangeable phenotypes, all coefficients have a common mean. Under the alternative hypothesis, the assumption of a common mean will not hold.

As before, we have a set of functional observations $Y_{ij}(t)$. In the first step, we center all the functional observations across the organisms by subtracting the mean curve $\overline{Y}(t) = (IJ)^{-1} \sum_{i=1}^{I} \sum_{i=1}^{J} Y_{ij}(t)$, thus forming centered data $\widetilde{Y}_{ij}(t) = Y_{ij}(t) - \overline{Y}_{ij}(t)$.

The covariance function of the process, here denoted by $\sigma(s,t)$, is then estimated by the usual method of moments as

$$\widehat{\sigma}(s,t) = (IJ-1)^{-1} \sum_{i=1}^{I} \sum_{j=1}^{J} \widetilde{Y}_{ij}(s) \widetilde{Y}_{ij}(t).$$

Let $\sigma(s,t)$ be the theoretical covariance function. In theory, there are orthonormal functions $\psi_1(t), \psi_2(t), ...$ which are solutions to the eigenequation

$$\int \sigma(s,t)\psi(t)dt = \lambda\psi(s), \tag{2}$$

with corresponding eigenvalues $\lambda_1 > \lambda_2 > ... > 0$. Then, according to the Karhunen-Loeve expansion, every centered functional observation $\widetilde{Y}_{ij}(t)$ can be approximated as

$$\widetilde{Y}_{ij}(t) \approx \sum_{k=1}^{K} \gamma_{ijk}\psi_k(t),$$

where the γ_{ijk} are the corresponding principal component scores with $\gamma_{ijk} = \int \widetilde{Y}_{ij}(t)\psi_k(t)dt$, and are zero-mean random variables with $\text{cov}(\gamma_{ijk}, \gamma_{ij\ell}) = I(l=k)\lambda_k$, where $I(\cdot)$ is the indicator function.

Of course in practice, the calculations are done using the estimated covariance function $\widehat{\sigma}(\cdot)$. Using standard calculations (Jolliffe and Silverman,

2002; Ramsey and Silverman, 2006), this produces estimates $\widehat{\gamma}_{ijk}$, with the approximate property that $\widehat{\gamma}_{ijk} \approx \text{Normal}(\mu_{ik}, \sigma^2_{ik})$, which are approximately independent across $i = 1, ..., I$ and $k = 1, ..., K$. We use the first two principal components, $k = 1, 2$, and test the null hypothesis that the mean of the principal component scores are the same, i.e., $H_0 : \mu_{11} = \mu_{21}$ and $\mu_{12} = \mu_{22}$.

In our implementation, we do not assume that the principal component scores are exactly independent across $k = 1, ..., K$, and instead "spend α" by weighting. Having conducted $k = 1, 2$ t-tests that produce $k = 1, 2$ p-values p_k, we reject H_0 at level α if either $p_1 < \alpha\lambda_1/(\lambda_1 + \lambda_2)$, or $p_2 < \alpha\lambda_2/(\lambda_1 + \lambda_2)$, where λ_1 and λ_2 are corresponding eigenvalues. In work not shown here, we have found that these weighted critical values have much greater power than simple rejection of the null hypothesis if $p_1 < \alpha/2$ or $p_2 < \alpha/2$.

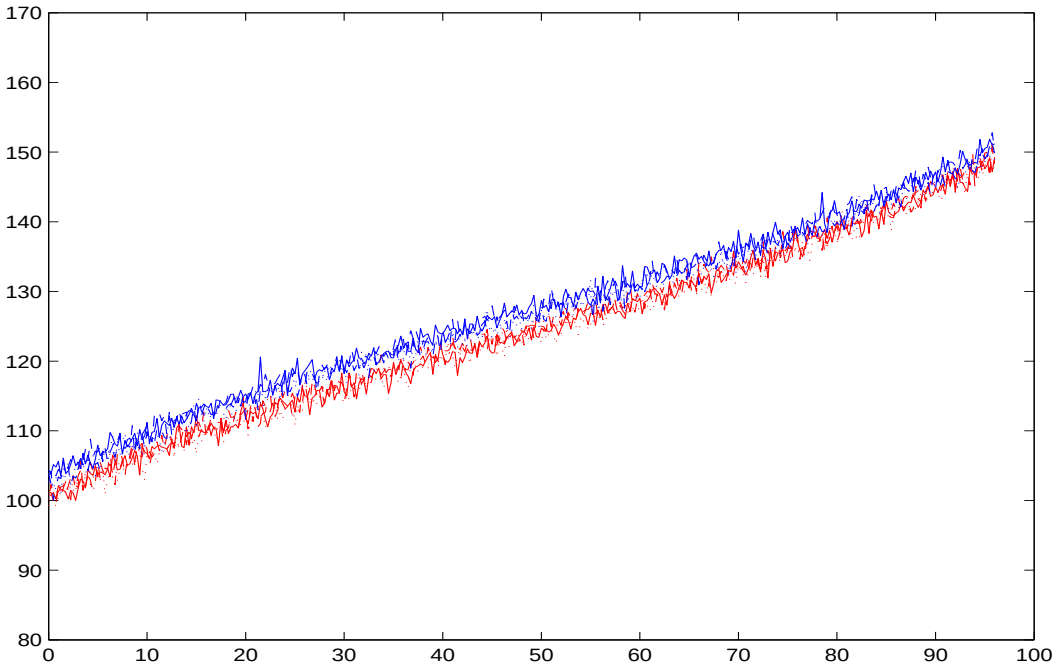

Figure 2: Simulated data in Section 4.1 with $\Delta = 3.0$ in (3), showing level shifts. There are four simulated data sets for each of $i = 1$ (red lines) and $i = 2$ (blue lines).

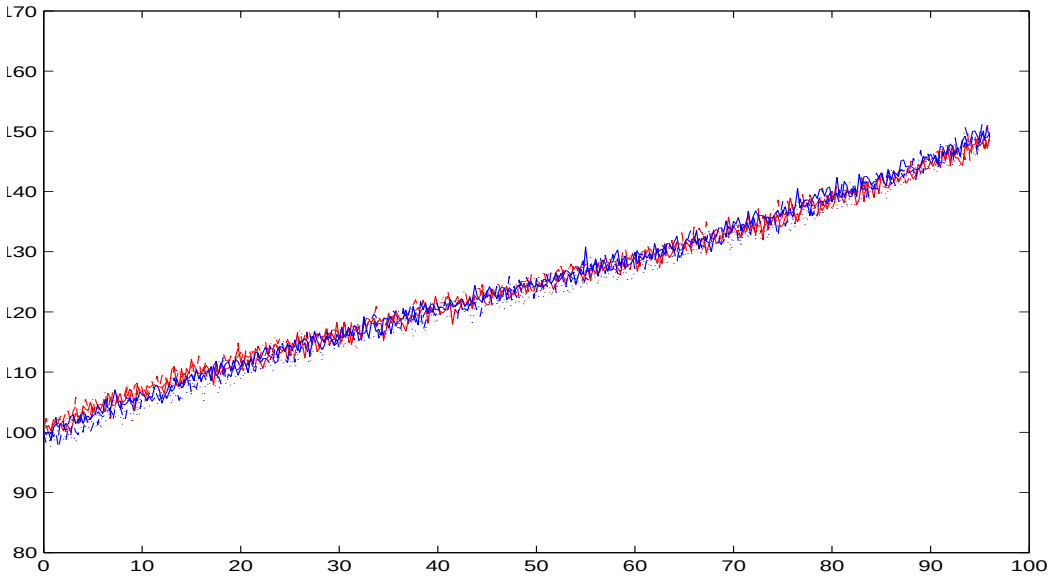

Figure 3: Simulated data in Section 4.1 with $\gamma = 0.02$ in (4), showing crossing effects. There are four simulated data sets for each of $i = 1$ (red lines) and $i = 2$ (blue lines).

4 Simulations

4.1 Simulations in the Linear Part of the Range

The first set of simulations were meant to mimic a case where only the linear part of the curve is analyzed. We performed two types of simulations. In the first, the curves were level shifts of one another, and in the second, the curves crossed at a random location. In both cases, the number of groups was $I = 2$, and there were $J = 4$ biological replicates. The curves were sampled at 385 time points t which were equally spaced on the interval $[0, 96]$. The first model we used was that

$$Y_{ij}(t) = \Delta I(i = 2) + \sum_{k=0}^{3} \beta_k t^k + \sum_{k=0}^{3} \theta_{kj} t_k + \epsilon_{ij}(t), \tag{3}$$

where $I(\cdot)$ is the indicator function, $\epsilon_{ij}(t) = \text{Normal}(0, 1)$ are independent, $(\beta_0, ...\beta_3) = (100, 0.6925. - 0.065, -0.01)$, $\theta_{0j} = \text{Normal}(0, 1)$ are independent, and for $k = 1, ..., 3$, $\theta_{kj} = \text{Normal}(0, .001)$. We varied Δ from 0.0 to 6.0, with $\Delta = 0$ being the null model. The results in terms of test level and power from

Δ	Perm Perm	Perm Area	Perm Median	FDA
0.0	0.036	0.045	0.037	0.027
0.5	0.082	0.093	0.083	0.064
1.0	0.173	0.193	0.173	0.125
1.5	0.316	0.345	0.309	0.243
2.0	0.509	0.532	0.498	0.412
2.5	0.682	0.696	0.656	0.570
3.0	0.812	0.829	0.791	0.736
3.5	0.910	0.908	0.887	0.853
4.0	0.969	0.971	0.955	0.926
4.5	0.987	0.988	0.979	0.979
5.0	0.997	0.997	0.992	0.992
5.5	0.999	0.999	0.995	0.999
6.0	1.000	1.000	0.999	0.998

Table 1: Results of the simulation study in Section 4.1 when the two population mean functions differ by the amount Δ. When $\Delta = 0$, the values are the levels of a nominal 5% level test. When $\Delta > 0$, the results are the powers. Here "Perm" is the permutation test based on the mean, "Perm Area" is the permutation test based on the area under the mean curve, "Perm Median" is the permutation test based on the differences of the medians, and "FDA" is the 2-principal component test.

1,000 simulated data sets are displayed in Table 1, and show little difference among the methods.

Another set of simulations (Table 2) compares two approaches when the second group of curves, instead of being 'shifted' by Δ units, 'crosses' the curves from the first group. The model was

$$Y_{ij}(t) = I(i = 2)(t - t_0)\gamma + \sum_{k=0}^{3} \beta_k t^k + \sum_{k=0}^{3} \theta_{kj} t_k + \epsilon_{ij}(t), \tag{4}$$

where we varied γ from 0.0 to 0.04, and t_0 was randomly generated across the range of time values. The results are shown in Table 2.

To illustrate the generated functions, in Figures 2 and 3, we generate a single simulated data set when $I = 2$ and $J = 4$. In the shift case (Figure 2), we set $\Delta = 3$, while in the crossing case (Figure 3) we set $t_0 = 48$ and $\gamma = 0.02$.

In the shift case of Figure 2 and (3), all methods have roughly comparable power, although the FDA approach was slightly less powerful. In the

$\tan(\gamma)$	Perm Perm	Perm Area	Perm Median	FDA
0.005	0.117	0.096	0.118	0.088
0.010	0.356	0.211	0.341	0.308
0.015	0.700	0.316	0.673	0.626
0.020	0.901	0.470	0.860	0.810
0.025	0.973	0.573	0.956	0.936
0.030	0.992	0.655	0.986	0.982
0.035	0.999	0.710	0.997	0.997
0.040	1.000	0.723	1.000	0.999

Table 2: Results of the simulation study in Section 4.1 when the two population mean functions rotated by γ. For the different levels of γ, the results are the powers. Here "Perm" is the permutation test based on the means, "Perm Area" is the permutation test based on the area under the mean curve, "Perm Median" is the permutation test based on the differences of the medians, and "FDA" is the 2-principal component test.

crossing case of Figure 3 and (4), the method based on the area under the curve is distinctly less powerful than the others, as expected, since this is the case the the areas under the curve will be similar even though the curves are different.

4.2 Simulations of Four-Parameter Logistic Functions

We simulated curves following a 4-parameter logistic model, with the j^{th} curve for organism $i = 1, 2$ at time t_k being 385 equally spaced values on the interval $[0, T]$ given as

$$\begin{aligned} Y_{1jk} &= \beta_{1j} + (\beta_{2j} - \beta_{1j})/\{1 + (x_k/\beta_{3j})^{-2-\beta_{4j}}\}; \\ Y_{2jk} &= \theta_{1j} + (\theta_{2j} - \theta_{1j})/\{1 + (x_k/\theta_{3j})^{-2-\theta_{4j}}\}. \end{aligned}$$

Our first two simulations involved cases where the curves separated as shifts. In the first, we had $T = 48$, $\beta_{1j} = \theta_{1j} = 50$, $\beta_{2j} = \theta_{2j} = 300$, $\beta_{3j} = \text{Uniform}(11, 13)$, $\theta_{3j} = \text{Uniform}(11, 13) + \Delta$, with $c = 0.0, 0.5, ..., 3.0$. In addition, $\beta_{4j} = \text{Uniform}[4, 6]$ and $\theta_{4j} = \text{Uniform}[4, 6]$. The data for the first strain, Y_{1jk}, closely mimic an earlier experiment for actual data from 2% (w/v) NaCl (PM9 well A2) described in Section 5. In Figure 4, we evaluate the mean of the functions for various levels in Δ. In this first simulation, all the tests were roughly equivalent in terms of power.

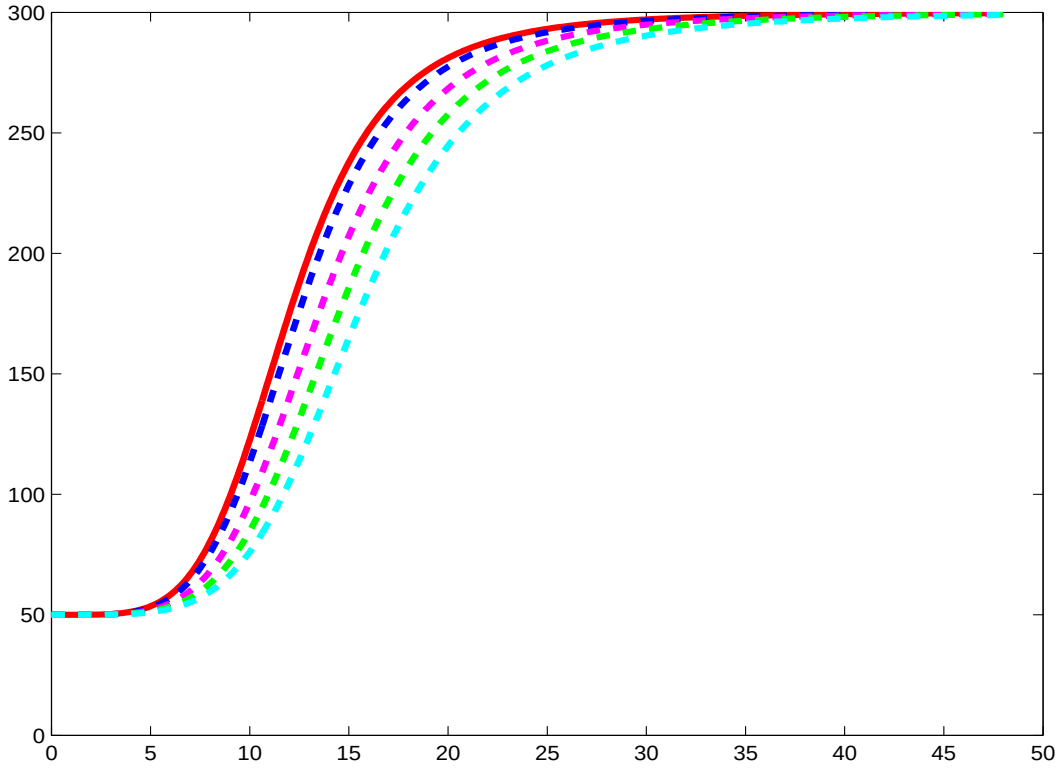

Figure 4: Mean curves for the first simulation in Section 4.2. The solid line is the null model mean curve, while the dashed lines illustrate some of the mean curves for the alternatives.

Our second simulation was similar, but with $\theta_{2j} = \beta_2 - \kappa$, where we varied the levels of $(\Delta, \kappa) = (0.0, 0.0), ..., (1.5, 6.0)$, see Figure 5 for the mean curves. This situation is meant to be a case where the permutation test based on the area under the mean curve will do the best, because of the level shifting inherent in κ. This indeed transpired: for example, with $\Delta = 1.0$ and $\kappa = 4.0$, the power of the test based on the area under the curve was 70%, while that for the comparison of the means was 50%.

In our third simulation, we generated data in which the curves cross. Here $T = 100$, $\beta_{1j} = \theta_{1j} = 100$, $\beta_{2j} = \theta_{2j} = 200$, $\beta_{3j} = \text{Uniform}(47, 53)$, $\theta_{3j} = \text{Uniform}(47, 53)$, $\beta_{4j} = \text{Uniform}(-3.5, -0.5)$ and $\beta_{4j} = \Delta\text{Uniform}(-3.5, -0.5)$, where $\Delta = 0$ representing the null hypothesis. The expectations of these two functions are given in Figure 6 when $\Delta = 1.5$. As expected, the permutation test using the area under the curve performs poorly here, see Table 3.

DOI: 10.2202/1557-4679.1227

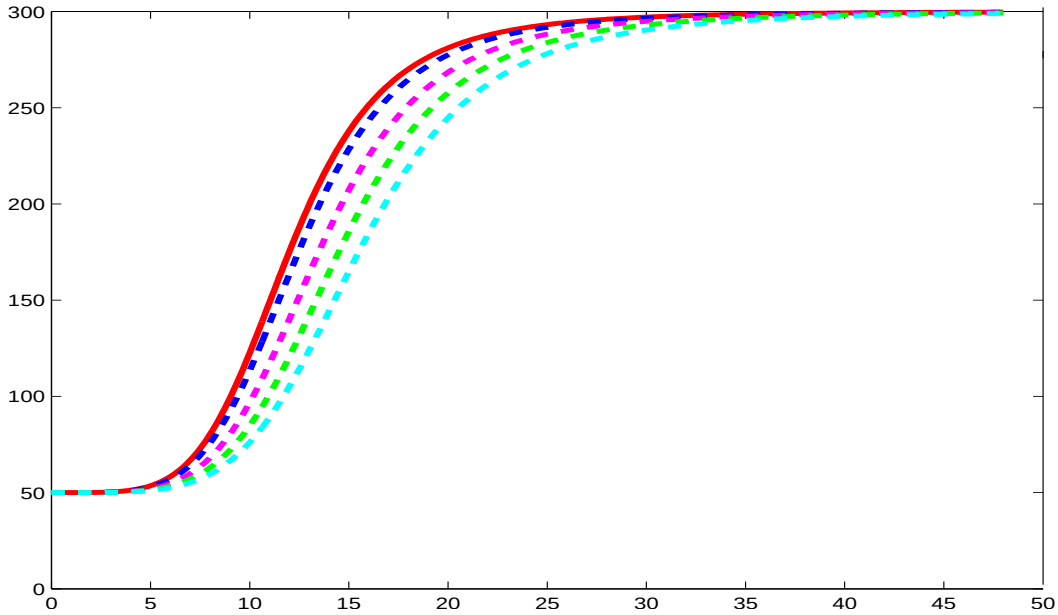

Figure 5: Mean curves for the second simulation in Section 4.2. The solid line is the null model mean curve, while the dashed lines illustrate some of the mean curves for the alternatives.

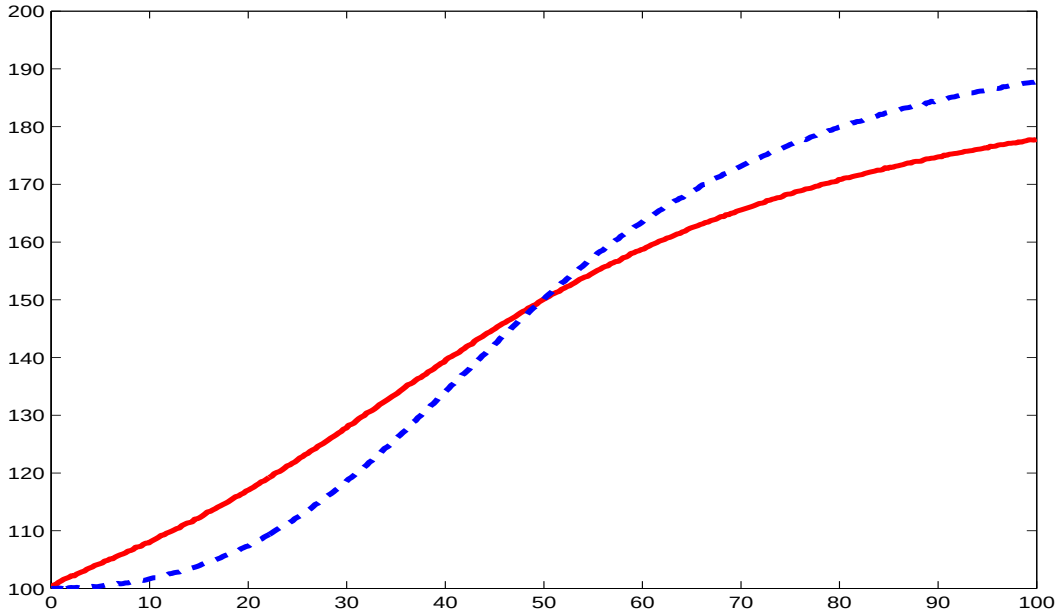

Figure 6: Mean curves for the Mean curves for the third simulation in Section 4.2. The solid line is for $\Delta = 0$, the null case, and the dashed for $\Delta = 1.5$.

Δ	Perm Perm	Perm Area	Perm Median	FDA
0.00	0.05	0.05	0.06	0.04
0.37	0.08	0.06	0.08	0.06
0.75	0.17	0.06	0.17	0.12
1.12	0.28	0.06	0.29	0.20
1.50	0.50	0.09	0.49	0.36
1.87	0.66	0.14	0.65	0.51
2.25	0.85	0.20	0.85	0.65
2.62	0.97	0.23	0.97	0.79
3.00	0.99	0.36	0.99	0.89

Table 3: Results of the second simulation study in Section 4.2, when there are crossing lines, see Figure 6. Here "Perm" is the permutation test based on the means, "Perm Area" is the permutation test based on the area under the mean curve, "Perm Median" is the permutation test based on the differences of the medians, and "FDA" is the 2-principal component test.

5 Comparative Analysis of Phenomics Data

While the *nhaA* gene is critical for the survival of E. coli in high-sodium environments, *nhaA* mutant strains are far less sensitive to unrelated ions (e.g., potassium) at equivalent concentrations (Padan et al., 1989; Trchounian and Kobayashi, 1999). As such, deletion of *nhaA* ($\Delta nhaA$) should result in decreased kinetic activity in the presence of sodium ion but not in the presence of unrelated ions or osmolytes. To test this, the osmotic and ionic sensitivity profiles were determined for *E. coli* K-12 BW25113 $nhaA^+$ and its isogenetic *nhaA*-deletion derivative ECK0020 $\Delta nhaA$ (Baba et al., 2006) using Biolog phenotype microarrays.

The PM9 microplates were used to compare the kinetic activity for these model bacteria, with four biological replicates. We were able to detect significant difference in this activity (Figure 7) between the two strains when they were cultivated in the presence of 2% (w/v) NaCl (PM9 well A2). The p-values for the permutation tests based on differences of the means, the differences of the medians, and the area under the curve were all equal to 0. The functional data analysis approach also rejected null-hypothesis about the equality of kinetic profiles at the level $\alpha = 0.05$.

We now turn to the presence of unrelated ions or osmolytes, where we do not expect to see an effect. Figure 8 shows results for PM9 well D2 (4% KCl). For this well, we conducted permutation and functional data analysis

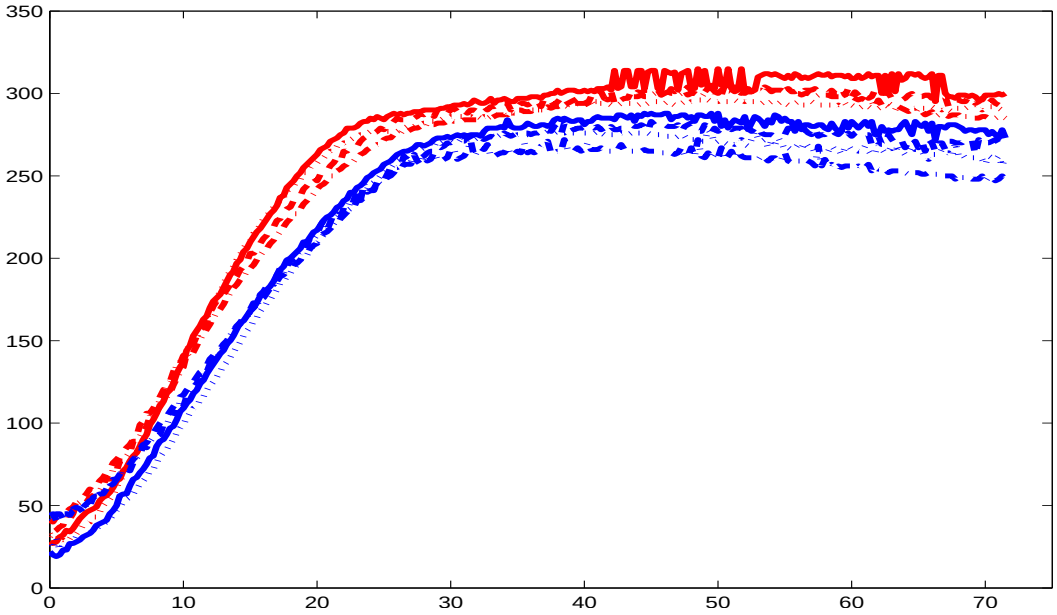

Figure 7: Kinetic activity of wild-type strain BW2511 (red lines, n = 4) and the *nhaA* mutant ECK0020 (blue lines, n = 4) in 2% (w/v) NaCl (PM9 well A2). These data are not ratio normalized. The p-values for the permutation tests based on differences of the means, the differences of the medians, the area under the curve were all 0.00. The FDA approach also rejects H_0, at the level $\alpha = 0.05$.

tests that produced the following results: the p-values for the permutation tests based on differences of the means, the differences of the medians, the area under the curve were 0.11, 0.09, 0.14, respectively, so that we found no significant difference between the two organisms, although the results are suggestive. The FDA approach also had a p-value > 0.05. These results are in agreement with cultivation-based studies, which indicate *nhaA* mutants are capable of growth in the presence of high concentrations (i.e., 0.15 M) of KCl, but not NaCl (Trchounian and Kobayashi, 1999).

Earlier, we discussed normalization of the data. One such normalization method has been suggested in the literature (Hackett and Griffiths, 1997; Weber, et al., 1007), namely at each time point and for each plate, we divided the raw signal for a well by the average across all the wells. This is similar to what is done with microarrays. Thackett and Griffiths (1997) are cautious about this methodology, as are we because these are not microarrays, and they suggest that "we prefer to investigate the effect of this transformation on each data set".

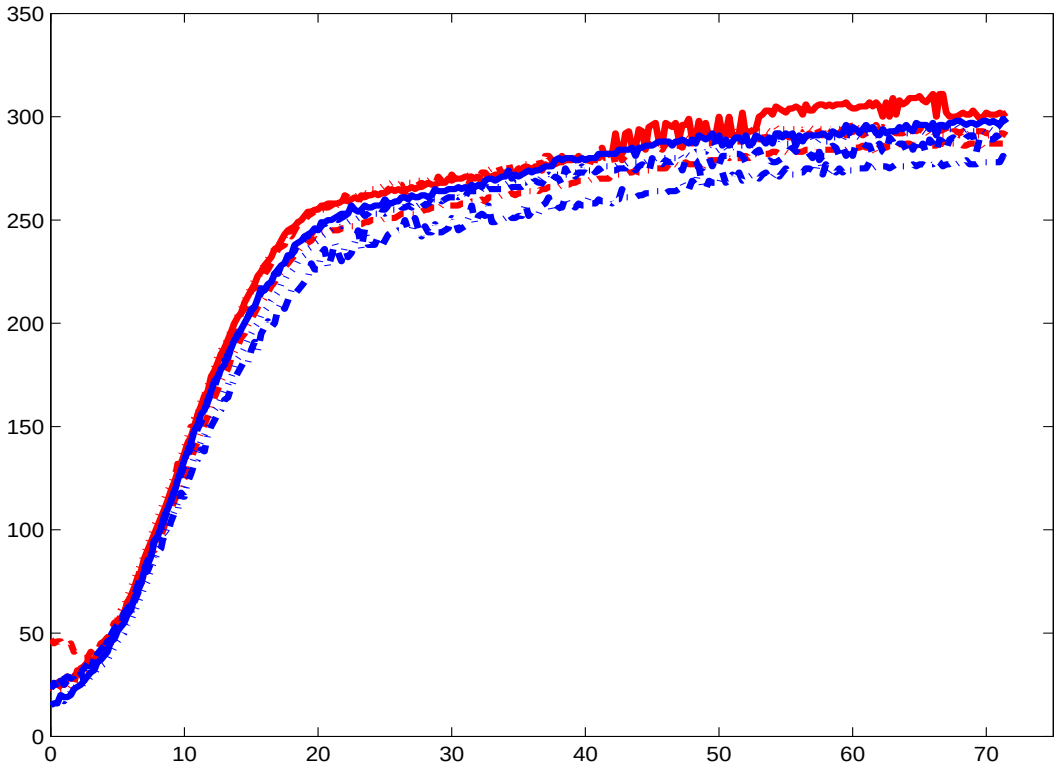

Figure 8: Kinetic activity of wild-type strain BW2511 (red lines, n = 4) and the *nhaA* mutant ECK0020 (blue lines, n = 4) in 4% (w/v) KCl (PM9 well D2). The p-values for the permutation tests based on differences of the means, the differences of the medians, the area under the curve were 0.11, 0.09, 0.14. The FDA approach also does not reject the null hypothesis.

We applied this type of normalization to the D2 well in the PM9 plate, where as described previously, we would not expect to have a statistically significant difference between the organisms. See Figure 9 below, for the vast and surprising differences between raw and normalized data. When we used the raw data, all the p-values exceeded 0.05. However, the normalized data show vast differences, with crossing lines such as seen in some of our simulations, and the p-values all = 0.00. We do not believe on biological grounds that there should be such a highly statistically significant effect, and conjecture that it is an artifact caused by the attempt to normalize. Further investigation of normalization is clearly warranted. In addition, we believe that phenotype microarrays and other standardized commercial platforms for high-throughput phenotyping also provide an intriguing platform for retrospective meta-analysis of phenomic data.

DOI: 10.2202/1557-4679.1227

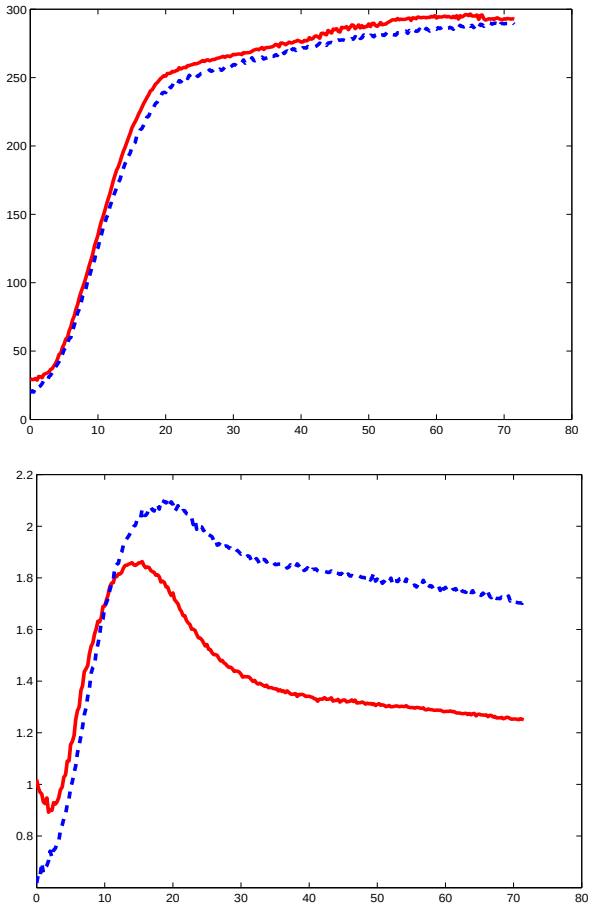

Figure 9: This is a plot of the mean functions across 4 biological replicates for the wild-type strain BW2511 (red line) and the *nhaA* mutant ECK0020 (blue line) in 4% (w/v) KCl (PM9 well D2). The left panel uses the raw data, while the right panel normalizes the data at each time point by dividing the raw data by the mean of the raw data across all the wells in the plate.

6 Concluding Remarks

The BIOLOG Phenotype Microarray platform offers a wealth of possible analyses, and this paper is by no means a final analysis. For example, while we have focused on the case of comparing phenotypes between two organisms, increasing many organisms are to be compared simultaneously. Our methodology is easily extended to allow for a simultaneous test of whether any of the organisms are statistically significantly different from others. For example, (1) can be extended by taking the sum of squares about the overall mean function.

References

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular System Biology*, 2, 2006.0008, doi: 10.1038/msb4100050.

Bochner, B. R., Gadzinski, P. and Panomitros, E. (2001). Phenotype microArrays for high-throughput phenotypic testing and assays of gene function. *Genome Research*, 11, 1246-1255.

Bochner, B. R. (2003a). New technologies to assess genotype-phenotype relationships. *Nature Review Genetics*, vol. 4, pp. 309-14.

Bochner, B. R. (2003b). Phenotype microArrays: their use in antibiotic discovery, Chapter 9 in *Microbial Genomics and Drug Discovery*, T. J. Dougherty and S. J. Projan, editors. Marcel Dekker, Inc.

Bochner, B. R. (2009). Global phenotypic characterization of bacteria. *FEMS Microbiology Reviews*, 33, 191-205.

Borglin, S. E. (2009). Overcoming the anaerobic hurdle in phenotypic microarrays: generation and visualization of growth curve data for *Desulfovibrio vulgaris* Hildenborough. Lawrence Berkeley National Laboratory: Lawrence Berkeley National Laboratory. LBNL Paper LBNL-1822E. Retrieved from: http://escholarship.org/uc/item/20q2b1g9.

Gabrielson, J., Hart, M., Jarelov, A., Kuhn, I., Mckenzie, D. and Mollby, R. (2002). Evaluation of redox indicators and the use of digital scanners and spectrophotometers for quantification of microbial growth in microplates. *Journal of Microbiological Methods*, 50, 63-73.

Gowen, C. M. and Fong, S. (2009). Phenome analysis of microorganisms. In *Bioinformatics Tools and Applications*, D. Edwards, J. Stajich and D. Hansen, editors. Springer, New York.

Hackett, C. A. and Griffiths, B. S. (1997). Statistical analysis of the time-course of Biolog substrate utilization. *Journal of Microbial Methods*, 30, 63-69.

Johnson, D. A., Tetu, S. G., Phillppy, K., Chen, J., Ren, Q. and Paulsen, I. T. (2009). High-throughput phenotypic characterization of *Pseudomonas aeruginosa* membrane transport genes. *PLoS Genetics*, 4, issue 10, e1000211.

Joyce, A. R. and Palsson, B. (2006). The model organism as a system: integrating 'omics' data sets. *Nature Reviews Molecular Cell Biology*, 7, 198-210.

Jolliffe, I. T. and Silverman, B. W. (2002). Principal Component Analysis. Springer, New York.

Kaufmann, K. W. (1981). Fitting and using growth curves. *Oecologia* 49, 293-299.

Mukherjee, A., Jackson, S. A., LeClerc, J. E. and Cebula, T. A. (2006). Exploring genotypic and phenotypic diversity of microbes using microarray approaches. *Toxicology Mechanisms and Methods*, 16, 121-128.

Padan, E., Maisler, N., Taglicht, D., Karpel, R. and Schuldiner, S. (1989). Deletion of ant in *Escherichia coli* reveals its function in adaptation to high salinity and an alternative Na+/H+ antiporter system(s). *Journal of Biological Chemistry*, 264, 20297-20302.

Padan, E., Venturi, M., Gerchman, Y. and Dover, N. (2001). Na+/H+ antiporters. *Biochimica et Biophysica Acta - Bioenergetics*, vol. 1505, issue 1, pp. 144-57.

Ramsay, J. O. and Silverman, B. W. (2006). *Functional Data Analysis*. Springer, New York.

Savage, D. C. (1977). Microbial ecology of the gastrointestinal tract. *Annual Review of Mirobiology*, 31, pp. 107-133.

Slonim, D. K. (2002). From patterns to pathways: gene expression data analysis comes of age. *Nature Genetics*, 32, 502-508.

Trchounian, A. and Kobayashi, H. (1999). Fermenting *Escherichia coli* is able to grow in media of high osmolarity, but is sensitive to the presence of sodium ion. *Current Microbiology*, 39, 109-114.

Weber, K. P., Grove, J. A., Gehder, M., Anderson, W. A. and Legge, R. L. (2007). Data transformation in the analysis of community-level substrate utilization data from microplates. *Journal of Microbial Methods*, 69, 461-469.

Xu, J. and Gordon, J. I. (2003). Inaugural article: Honor thy symbionts. *Proceedings of the National Academy of Sciences USA*, 100, 10452-10459.

Zhou, L., Lei, X. H., Bochner, B. R. and Wanner, B. L. (2003). Phenotype microarray analysis of *Escherichia coli* K-12 mutants with deletions of all two-component systems. *Journal of Bacteriology*, 185, 4956-4972.