

Article

Longitudinal Analysis of Contrasts in Gene Expression Data

Georg Hahn ^{1,*} , Tanya Novak ² , Jeremy C. Crawford ³ , Adrienne G. Randolph ² and Christoph Lange ¹¹ Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA² Critical Care Medicine, Department of Anesthesiology, Boston Children's Hospital, Boston, MA 02115, USA³ St. Jude Children's Research Hospital, Memphis, TN 38105, USA

* Correspondence: ghahn@hsph.harvard.edu

Abstract: We are interested in detecting a departure from the baseline in a longitudinal analysis in the context of multiple organ dysfunction syndrome (MODS). In particular, we are given gene expression reads at two time points for a fixed number of genes and individuals. The individuals can be subdivided into two groups, denoted as groups *A* and *B*. Using the two time points, we compute a contrast of gene expression reads per individual and gene. The age of each individual is known and it is used to compute, for each gene separately, a linear regression of the gene expression contrasts on the individual's age. Looking at the intercept of the linear regression to detect a departure from the baseline, we aim to reliably single out those genes for which there is a difference in the intercept among those individuals in group *A* and not in group *B*. In this work, we develop testing methodology for this setting based on two hypothesis tests—one under the null and one under an appropriately formulated alternative. We demonstrate the validity of our approach using a dataset created by bootstrapping from a real data application in the context of multiple organ dysfunction syndrome (MODS).

Keywords: contrasts; longitudinal data; gene expression; multiple organ dysfunction syndrome (MODS)



Citation: Hahn, G.; Novak, T.; Crawford, J.C.; Randolph, A.G.; Lange, C. Longitudinal Analysis of Contrasts in Gene Expression Data. *Genes* **2023**, *14*, 1134. <https://doi.org/10.3390/genes14061134>

Academic Editor: Stefano Lonardi

Received: 30 April 2023

Revised: 19 May 2023

Accepted: 21 May 2023

Published: 24 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this article, we present novel methodology whose development was motivated by an application in the context of multiple organ dysfunction syndrome (MODS). The data that motivated this research are structured as follows. We are given gene expression reads at two time points for $m \in \mathbb{N}$ genes and $n \in \mathbb{N}$ individuals. The gene expression reads at the two time points are translated into contrasts of gene expression per individual and gene, thus effectively reducing the input to a scalar value per individual and gene. Moreover, we possess the age of each individual. The individuals can be subdivided into two groups, denoted as groups *A* and *B*. For instance, the two groups can be defined as those individuals who recovered from MODS (group *A*) and those who suffer from a condition called “prolonged MODS” (group *B*).

For each gene separately, we want to perform a linear regression of the gene expression contrasts on the individual's age. As we are interested in a departure from the baseline, we look at the intercept of the linear regression. The aim of this work is to reliably single out those genes for which there is a difference in the intercept among those in group *A* and not in group *B*.

To approach this, we develop testing methodology based on two hypothesis tests. The two hypothesis tests are once under the null, and once under an appropriately formulated alternative. In an abstract setting, we are faced with two linear regressions— $L_i : y_i = \alpha_i a_i + \beta_i$, where $y_i \in \mathbb{R}^{n_i}$, $a_i \in \mathbb{R}^{n_i}$, and $\alpha_i, \beta_i \in \mathbb{R}$ for $i \in \{1, 2\}$, where n_1 and n_2 are the sizes of the groups *A* and *B*, respectively. For group *A*, we wanted to test the hypothesis $H : \beta_1 = 0$ versus its complement $H' : \beta_1 \neq 0$. For group *B*, given some level λ , we test under the alternative, meaning $\tilde{H} : \beta_2 > \lambda$ versus its complement $\tilde{H}' : \beta_2 \leq \lambda$.

The aforementioned tests are carried out on both groups for each gene contrast under consideration. Since there are $m \in \mathbb{N}$ genes, we have to use a multiple testing correction across the $2m$ p -values obtained by testing H_1 and H_2 for all m genes.

We demonstrate the validity of our approach using a simulated dataset in the context of multiple organ dysfunction syndrome (MODS) [1]. The simulated dataset has the purpose of showcasing the new methodology, and it is generated by bootstrapping from real MODS data of the Pediatric Intensive Care Influenza (PICFLU) investigators group [2,3]. As the purpose of the present article is to introduce the specific testing scenario and one possible solution approach, the usage of simulated data is valid. An application to real MODS data and the subsequent biological interpretation of the results are deferred to a separate publication.

The article is structured as follows. We start with a brief literature review in Section 1.1. Section 2 gives details on the dataset under the investigation and introduces the methodology we employ to solve our testing problem. This methodology can be subdivided into a series of (mathematical) components; for instance, the calibration of the alternative, the precise calculation of p -values, and the hypothesis tests being carried out. A demonstration of the methodology on simulated data can be found in Section 3. The article concludes with a discussion in Section 4.

Throughout the entire article, we denote with $Y_{\cdot,j}$ the j 'th column of a matrix Y .

1.1. Literature Review

While, initially, studies often compared gene expression data between distinct groups at fixed time points, there is a growing literature which considers time dependent expression data, meaning studies which extract insights from mRNA (or similar) samples collected at successive time points.

The identification of differentially expressed genes in a time course study is an active area of research. Starting with [4], the authors consider a two-way analysis of variance (ANOVA) approach combined with a permutation test to obtain p -values, where both group and time are the main factors. Their aim is to quantify the group effects (via permutations over all levels of the group labels for a fixed time point) and the time effects separately. The proposed approach to handling this problem is a two-stage model which first removes the time effect and then looks at the group effect. The given model is linear and the effects are determined via hypothesis testing with multiplicity correction.

In another work [5], the authors propose to identify differentially expressed genes in a time course study using a parametric model for the expression values in connection with a false discovery rate approach. In particular, the authors aim to detect changes in either a single biological group or differences in expression over time between two or more groups. The proposed method, called EDGE, is designed for inputs with or without replicates, and for both single group and multiple group tests. The trajectories of expressed genes are modelled with cubic splines and their goodness-of-fit to the model under both the null and alternative is evaluated, where the distributions are approximated via bootstrapping.

Another contribution in the literature [6] considers the questions of how to combine simultaneous inferences across multiple time points, as well as how to best control for multiplicity while accounting for the strong dependence between measurements. The authors formulate a decision-theoretic framework in which a gene is significant if a certain combination of null hypotheses is rejected at a given level. The focus of the authors is specifically on the optimal combination of testing at multiple time points, suitable multiplicity correction, and dependence among hypotheses. The hidden Markov model of [7] is generalized to capture the temporal correlation in the gene expression data.

In a similar fashion, in [8], the authors study the question of how to identify genes associated with a biological process in data having multiple time points, though not necessarily coming from the same individuals. They propose an approach based on functional principal component analysis (FPCA) in connection with hypothesis testing that allows one to incorporate high dimensionality, a low number of replicates, missing values,

and measurement errors or time correlations. In their model, the parametric form of the null hypothesis is unknown, and thus has to be approximated via permutations.

In contrast to the methodology presented in our article, which aims to identify gene contrasts showing a departure from the baseline in one group and not another, the aforementioned publications differ from ours in that they either aim to identify different group and time effects, consider two null hypotheses, or are based on functional principal component analysis.

A second line of research available in the literature aims to process expression profiles with graphical methods as opposed to hypothesis testing. For instance, an algorithm to increase the temporal resolution of expression measurements and an application to skeletal muscle differentiation can be found in [9]. The algorithm is essentially a pipeline to process cell expression profiles, which includes dimensionality reduction via independent component analysis, the construction of a minimum spanning tree (MST), and the subsequent computation of the longest path through the MST, which corresponds to the longest sequence of transcriptionally similar cells.

In the exploratory study of [10], the authors combine libraries of single-cell RNA-Seq for primary mouse bone marrow-derived dendritic cells (DCs) and find substantial variation between identically-stimulated DCs.

Finally, there is literature on software tutorials for gene expression analysis with different time points, which, however, does not present new methodology. A workflow for the statistics computer software R, for the purpose of analyzing data from a micro-array time-course experiment, is presented in [11]. The tutorial considers quality control and normalization, the identification of genes that are differentially expressed, the clustering of genes into distinct temporal patterns, and the biological interpretation of the clusters. Some of the examples considered in the tutorial cover the exposure of mice to three different strains of influenza and lung tissue data collected at 14 time-points after infection.

2. Methods

This section introduces the methodology pipeline we developed for the problem of longitudinal analysis of contrasts, which was briefly outlined in Section 1. We present our approach as a series of steps. We start with a mathematical abstraction of the problem in Section 2.1. A high level summary of our approach is given in Section 2.2 before giving details on the calibration of the alternative hypothesis (Section 2.3), the calculation of p -values (Section 2.4), and the testing of all generated hypotheses (Section 2.5). We conclude with a note on how to report findings (Section 2.6).

2.1. Problem under Investigation

Denote the sizes of the groups A and B as $n_1 \in \mathbb{N}$ and $n_2 \in \mathbb{N}$, respectively, and let $n = n_1 + n_2$. The first step is to reduce the input to a single contrast matrix $Y^{(i)} \in \mathbb{R}^{n_i \times m}$ for each group $i \in \{1, 2\}$, where $m \in \mathbb{N}$ is the number of genes. This effectively reduces the two endpoint problem we consider to a single endpoint. The matrices $Y^{(i)}$ for group $i \in \{1, 2\}$ contain the contrast for individual k and gene j at position (k, j) . This is visualized in Figure 1.

Formally, we are given two matrices $R^{(i)}, S^{(i)} \in \mathbb{R}^{n_i \times m}$ for $i \in \{1, 2\}$, where the matrices $R^{(i)}$ and $S^{(i)}$ contain expression data per individual (row) and gene (column). Using this data, we compute a contrast matrix $Y^{(i)} = R^{(i)} - S^{(i)} \in \mathbb{R}^{n_i \times m}$ per group, where $i \in \{1, 2\}$. The ages of the individuals in the two groups shall be given as a vector $a_i \in \mathbb{R}^{n_i}$ for $i \in \{1, 2\}$.

For each gene $j \in \{1, \dots, m\}$, we perform two linear regressions of the contrast on the age with intercept, that is, we perform $L_{i,j} : Y_{\cdot,j}^{(i)} = \alpha_{i,j} a_i + \beta_{i,j}$, where $\alpha_{i,j}, \beta_{i,j} \in \mathbb{R}$ for $i \in \{1, 2\}$. For group A (meaning $i = 1$), we want to test the hypothesis $H_j : \beta_{1,j} = 0$ versus its complement $H'_j : \beta_{1,j} \neq 0$ for each $j \in \{1, \dots, m\}$. For group B (meaning $i = 2$), we test under the alternative given some level λ_j for each gene j . To be precise, we test $\tilde{H}_j : \beta_{2,j} > \lambda_j$ versus its complement $\tilde{H}'_j : \beta_{2,j} \leq \lambda_j$.

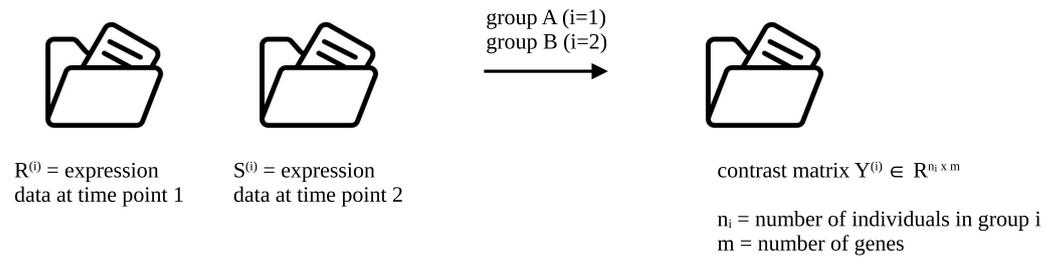


Figure 1. Preparation of the datasets $Y^{(i)}$ containing the contrast data for group A ($i = 1$) and group B ($i = 2$). The contrast data matrix $Y^{(i)} = R^{(i)} - S^{(i)}$ is obtained by computing the componentwise difference between the data at the two input timepoints $R^{(i)}$ and $S^{(i)}$.

2.2. Summary of the Approach

We aim to identify those individuals that are in both group A and *not* in group B , and whose gene expression reads are explained by a linear model that includes the age without intercept (see Section 1). Since the selection criterion is different for the two groups, the formulation of two simple hypothesis tests is chosen as it provides a valid framework to draw statistical inference.

The complete approach we use to identify such genes is summarized in Figure 2. Specifically, we start with the two contrast matrices $Y^{(1)}$ and $Y^{(2)}$ for groups A and B , respectively. For both groups, the linear model of Section 2.1 is fitted separately to each gene $j \in \{1, \dots, m\}$. For group A , the one for which we want to test for an intercept of zero, we test the null hypothesis $H_j : \beta_{1,j} = 0$ versus the complement $H'_j : \beta_{1,j} \neq 0$ (see Section 2.1) for each gene $j \in \{1, \dots, m\}$. For group B , we test under the alternative $\tilde{H}_j : \beta_{2,j} > \lambda_j$ versus its complement $\tilde{H}'_j : \beta_{2,j} \leq \lambda_j$ for each gene $j \in \{1, \dots, m\}$. Details on the calibration of the alternative are deferred to Section 2.3. In total, we thus observe $2m$ p -values (computed as described in Section 2.4), one per group and per gene $j \in \{1, \dots, m\}$. As a final step, we evaluate the resulting $2m$ p -values to single out only those genes $j \in \{1, \dots, m\}$ for which both H_j (having p -values p_j) and \tilde{H}_j (having p -values \tilde{p}_j) are rejected. Details on this multiple testing problem are given in Section 2.5.

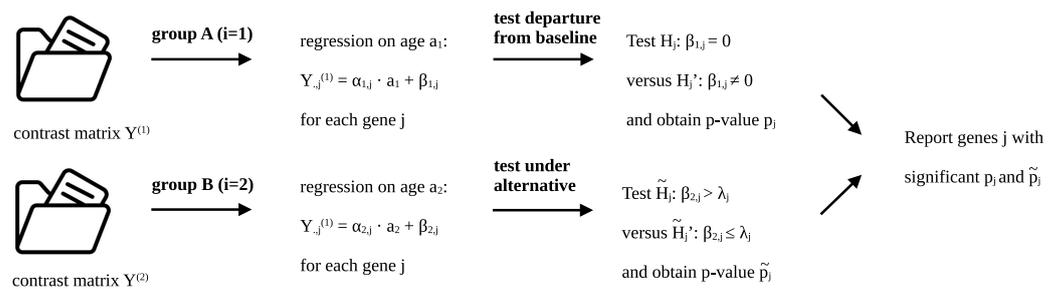


Figure 2. Summary of the testing pipeline. For each of the two groups A ($i = 1$) and B ($i = 2$), the same regression is carried out for each gene $j \in \{1, \dots, m\}$, using the data across the population of individuals (n_1 for group A , and n_2 for group B). However, the intercepts β_1 and β_2 are tested in two different ways. For group A , the departure from the baseline is tested, meaning $H_j : \beta_{1,j} = 0$ for each gene j . For group B , the hypothesis $\tilde{H}_j : \beta_{2,j} > \lambda_j$ is tested under the alternative for an appropriately selected λ_j and for all $j \in \{1, \dots, m\}$. The reported genes j must have significant p -values p_j and \tilde{p}_j .

2.3. Calibration of the Alternative

To have a well-defined alternative \tilde{H}_j , it is necessary to specify a level $\lambda_j \in \mathbb{R}$ for each gene $j \in \{1, \dots, m\}$. This means we consider an individual level for each test under the alternative. The calculation of the level under the alternative is performed as follows.

First, we compute the linear regression fit $L_{i,j} : Y_{i,j}^{(i)} = \alpha_{i,j}a_i + \beta_{i,j}$ for $i = 2$ and all $j \in \{1, \dots, m\}$. We only consider the case $i = 2$ here, since the calibration of the alternative applies to the hypotheses \tilde{H}_j only. Second, for the fitted coefficients $\alpha_{2,j}$ and $\beta_{2,j}$, we

compute the R^2 measure of the goodness-of-fit, and then scale the intercept $\beta_{2,j}$ (in either direction) until it explains only a proportion $0 < \pi < 1$ of the original R^2 . Computationally, since the R^2 measure decreases monotonically for misspecified coefficients, a binary search can be used across an interval of values for $\beta_{2,j}$ until the value explaining the fraction π of the initial R^2 is found. The value of $\beta_{2,j}$ explaining the proportion π of the initial R^2 is recorded as the level λ_j for each $j \in \{1, \dots, m\}$. In our simulations in Section 3, we use $\pi = 0.35$. This choice is arbitrary but motivated by practical applications [3].

2.4. *p*-Value Calculation

After having conducted all linear regressions of the null H_j (see Section 2.1), we compute the *p*-values of the contrasts as described in [12,13]. Let $q_d(\delta)$ denote the lower δ quantile of the *t*-distribution with *d* degrees of freedom.

In particular, let $t_{i,j}$ be the *t*-value statistic calculated for each linear regression $L_{i,j}$. For H_j , that is, in the scenario of group *A* (meaning $i = 1$), when the level is $\lambda_j = 0$, the distribution of $t_{1,j}$ is a *t*-statistic with $n_j - 2$ degrees of freedom. Therefore, the (two-sided) *p*-value of the contrast is given as $p_j = 2q_{n_j-2}(t_{1,j})$.

For the alternative \tilde{H}_j , that is, in the scenario of group *B* (meaning $i = 2$), we test $\tilde{H}_j : \beta_{2,j} > \lambda_j$. This is equivalent to testing $\tilde{H}_j : \beta_{2,j} - \lambda_j > 0$. Therefore, the (one-sided, upper tail) *p*-value of the contrast is given as $\tilde{p}_j = 1 - q_{n_j-2}(t_{2,j} - \lambda_j)$.

2.5. Multiple Hypothesis Testing

The *p*-value calculations of Section 2.4 result in $2m$ *p*-values, a vector (p_1, \dots, p_m) for (H_1, \dots, H_m) , and a vector $(\tilde{p}_1, \dots, \tilde{p}_m)$ for $(\tilde{H}_1, \dots, \tilde{H}_m)$. We are interested in evaluating those *p*-values in such a way as to find those indices $j \in \{1, \dots, m\}$ for which both H_j and \tilde{H}_j are rejected.

As we evaluate several hypotheses at the same time, a multiple testing correction is necessary. We consider two classical options in the remainder of the article, the Bonferroni correction [14] to control the Familywise Error Rate (FWER), and the Benjamini–Hochberg procedure to control the False Discovery Rate (FDR) [15]. Either procedure can be used, so long as one discloses that the reported significances are with respect to FWER or FDR control, respectively.

To stay conservative, we fix a testing threshold $\alpha \in (0, 1)$ which is prespecified by the user. We evaluate the hypotheses of both groups *A* and *B* separately, using a testing threshold of $\alpha/2$ for each to keep the overall type I error under control. We denote with \mathcal{R}_H^B and \mathcal{R}_H^{BH} the sets of rejected hypotheses among $H = (H_1, \dots, H_m)$ based on the *p*-values (p_1, \dots, p_m) for the Bonferroni correction and Benjamini–Hochberg procedure, respectively. Similarly, we denote with $\mathcal{R}_{\tilde{H}}^B$ and $\mathcal{R}_{\tilde{H}}^{BH}$ the sets of rejected hypotheses among $\tilde{H} = (\tilde{H}_1, \dots, \tilde{H}_m)$ based on the *p*-values $(\tilde{p}_1, \dots, \tilde{p}_m)$ for the Bonferroni correction and the Benjamini–Hochberg procedure, respectively.

2.6. Reporting the Findings

After evaluating all $2m$ hypotheses, we determine whether there exist one or more indices $j \in \{1, \dots, m\}$ such that both H_j and \tilde{H}_j are rejected. In this case, we report these indices as findings. If no such index exists, we report an empty set.

To be precise, we report the hypotheses in the set $\mathcal{R}_H^B \cap \mathcal{R}_{\tilde{H}}^B$ when controlling the FWER with the Bonferroni correction as a multiple testing correction. When controlling the FDR with the Benjamini–Hochberg procedure, we report the hypotheses in the set $\mathcal{R}_H^{BH} \cap \mathcal{R}_{\tilde{H}}^{BH}$.

3. Results

This section presents the dataset under investigation to which we apply the proposed methodology (Section 3.1). An example of the observed *p*-value distributions for both H_j

and \tilde{H}_j is presented in Section 3.2. We conclude with an example of the actual significances observed with our methodology (Section 3.3).

3.1. Dataset under Investigation

Severe lower respiratory tract infections (LRTI) are a leading cause of hospitalization and preventable death in children worldwide [16]. From 2010 to 2022, accounting for the United States alone, an average of 22.2 out of 100,000 children less than 18 years old were hospitalized with severe influenza virus infection, resulting in 1358 deaths [17]. Multiple organ dysfunction syndrome (MODS) is an uncommon but life-threatening complication of severe influenza infection in children that may negatively impact their longer-term health [1].

We consider a dataset originally created by the Pediatric Intensive Care Influenza (PICFLU) investigators group [2,3], comprising children of less than 18 years of age with confirmed influenza infection which were admitted to Pediatric Intensive Care Units (PICU) at 30 sites between March 2010 and March 2017. Influenza was confirmed as previously described in [18]. Individuals with known immunodeficiencies, chronic lung disease, symptomatic cardiac disease, neuromuscular disease, malignancy, metabolic or mitochondrial disease, or individuals who received systemic immunosuppressive medications within six weeks prior to admission for this acute illness were excluded.

A custom gene panel was designed for $m = 469$ mRNA targets incorporating genes known for moderating inflammation, cytokines, and associated with influenza and sepsis. Based on similar studies, the dataset also includes seven housekeeping genes [19–21]. The gene panel of the PICFLU dataset is given in the Supplementary Material.

The Pediatric Sequential Organ Failure Assessment (pSOFA) score, ranging from 0 to 24, was used to identify MODS. It can also quantify MODS over time [22] and it is positively correlated with mortality. To be precise, organ dysfunction was defined by a pSOFA score of 2 or greater. The class “Prolonged MODS” was defined as multiple organ dysfunction and/or extracorporeal membrane support (ECMO), or death on or after PICU day 7. MODS was measured at the time of the initial sample collection. PICU survivors with MODS when the first sample was collected who did not have MODS on or after PICU day 7 were categorized as “MODS Recovery”.

The dataset under investigation contains $n = 45$ individuals, divided into two groups A and B with $n_1 = 22$ and $n_2 = 23$ individuals, respectively.

Since the aim of this contribution is to showcase the new methodology, a simulated dataset is generated from the above dataset of the Pediatric Intensive Care Influenza (PICFLU) investigators group. As the analysis of the real dataset requires an extensive discussion of the biological implications, it is deferred to a separate publication. Instead, the simulated dataset we use was generated by bootstrapping (sampling with replacement) from the real data described above. For this, when bootstrapping, we keep the original group sizes of $n_1 = 22$ and $n_2 = 23$ individuals for groups A and B , respectively. For each individual, we consider the fixed panel of $m = 469$ genes, for which we possess gene expression measurements at two endpoints.

The bootstrapping is conducted as follows. Since we are only interested in the contrasts, see Figure 1, we first compute the contrasts for both the group “Prolonged MODS” (group A) and “MODS Recovery” (group B) on the real data. While keeping fixed the original group sizes of $n_1 = 22$ and $n_2 = 23$ individuals for groups A and B , respectively, we pool all contrast measurements for the group “Prolonged MODS” (group A) and draw bootstrapped measurements from this pool to create a new panel of n_1 individuals and m genes. The same is carried out using the contrast data for the group “MODS Recovery” (group B). A vector of m new age measurements is created by sampling with replacement from the original m age measurements, which are all contained in the interval $(0, 16)$.

3.2. Example of p -Value Distributions for the Two Hypotheses

Figure 3 shows the two distributions obtained after calculating the p -values for both H_j and \tilde{H}_j according to Section 2.4. Note that the p -values in both subfigures are sorted,

meaning that it is not possible to immediately compare the p -values in the left and right subfigures. We observe that, for the test under the null, the p -values are quite conservative, with only very few significances. For the test under the alternative, we observe a step function behavior, in the sense that the level λ_j we determine for each test \tilde{H}_j with $j \in \{1, \dots, m\}$ seems to make the p -values either (almost) zero or one.

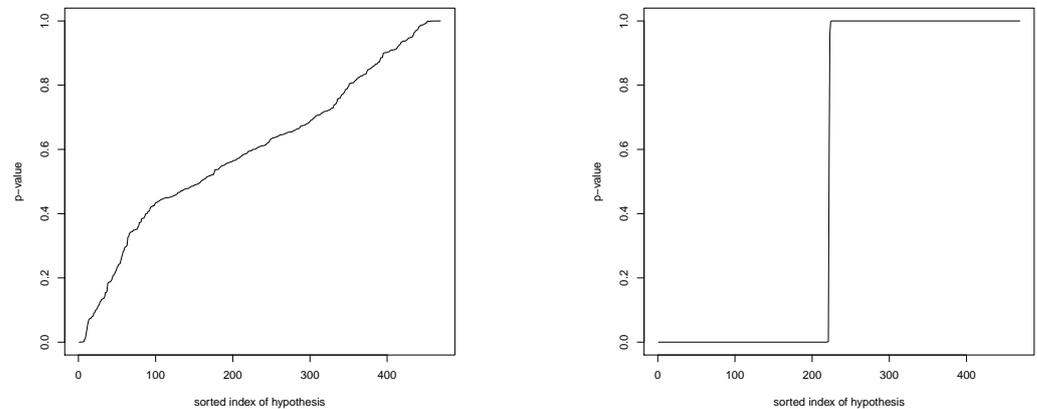


Figure 3. Sorted p -value distributions for H_j (left) and \tilde{H}_j (right).

3.3. Reported Genes

We evaluate all p -values considered in Section 3.2 by applying the Benjamini-Hochberg procedure to the ones calculated for the nulls H_j and the alternatives \tilde{H}_j separately at a level $\alpha/2$, where $\alpha = 0.05$. As described in Section 3.3, we then look for any indices $j \in \{1, \dots, m\}$ such that both H_j and \tilde{H}_j are rejected. For the above dataset, this procedure results in two genes to be reported which can be found in Table 1.

Table 1. p -values for H_j and \tilde{H}_j observed for the two rejections with respect to the FDR criterion.

| Name | p -Value for H_1 | p -Value for H_2 |
|-------|---------------------------|----------------------------|
| ALAS1 | 6.240291×10^{-5} | 1.553433×10^{-51} |
| CASP6 | 1.287339×10^{-4} | 1.127595×10^{-11} |

However, since the dataset we analyze here is created by bootstrapping from the PICFLU dataset (see Section 3.1), the *discovered* genes in Table 1 are actually random and thus not biologically meaningful. An application to real MODS data of [2,3], and the subsequent biological interpretation of the results, is deferred to a separate publication.

4. Discussion

This article considered the problem of testing contrasts for a gene expression application in the context of multiple organ dysfunction syndrome (MODS). The statistical challenge of the problem under consideration consists in the fact that we are interested in genes showing significances with respect to one group but *not* another group (denoted groups A and B). Although formulated as a problem with two endpoints, as a preprocessing step, the input data consisting of gene expression data collected at two different time points for the same set of genes and individuals in two different groups are converted to gene contrasts (differences in gene expression) per group. This effectively reduces the multiple endpoint problem to a single input.

Our proposed solution uses two hypothesis tests per gene under consideration, where m denotes the number of genes. Precisely, we conduct two linear regressions, where each linear regression allows us to determine if the contrasts can be explained by a single covariate alone (in our application this is the age covariate) and focus on the intercept to detect a departure from the baseline in gene expression. The two sets of m hypotheses

(leading to a total of $2m$ hypotheses which are being tested) consist of m hypotheses under the null (to detect a departure from the baseline in group A) and m hypotheses under the alternative (to model the condition that we are interested in genes showing significances in group A but not in group B). Special attention is paid to the formulation and calibration of an appropriate level under the alternative. The level we choose is essentially arbitrary, but as motivated in the literature (see Section 2.3), one option is to scale the intercept until it only explains a fraction (e.g., a fraction of 0.35) of the initial R^2 of the linear regression fit. To obtain p -values for both the null hypotheses and the alternative hypotheses, we give explicit formulas based on a t -distribution. We evaluate all p -values with the help of the FWER or FDR criterion to correct for multiple comparisons and report those genes as findings which are significant under both the null as well as the alternative.

There is no restriction on the number of gene contrasts that can be tested with our proposed methodology as long as the multiple testing correction is carried out correctly. This is due to the fact that the discovery of genes is essentially deferred to the discovery of significant hypotheses among the nulls H_j and the alternative hypotheses \tilde{H}_j , see Section 2.5. As long as a valid testing procedure is being used, such as the one in [14] to control the FWER or the one in [15] to control the FDR, any number of gene contrasts can be tested. In particular, if a type I error of α is desired, it is valid to split this error evenly among the m nulls and the m hypotheses under the alternative, thus applying the Bonferroni correction at threshold $\alpha/2$ to the m hypotheses per group.

The application which prompted the development of this methodology is in the area of multiple organ dysfunction syndrome (MODS). The dataset originally created for MODS stems from the Pediatric Intensive Care Influenza (PICFLU) investigators group [2,3] and serves as the basis of the experiments reported in this publication. However, the developed methodology is not tied to a certain type of data and, thus, we aimed to separate the presentation of the methodology and the analysis of the PICFLU dataset. Therefore, we used a simulated dataset to showcase the methodology, which is based on the real PICFLU dataset with an equal number of individuals and mRNA targets (as in the original PICFLU dataset) created via bootstrapping. Therefore, the *discovered* genes reported in Section 3 are actually not biologically meaningful and an interpretation of the *discoveries* is not sensible. This is due to the fact that an application to real MODS data will involve a much more meticulous biological interpretation of the results, and it is therefore deferred to a separate publication.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14061134/s1>, The gene panel of the PICFLU dataset is given in the supplementary material.

Author Contributions: G.H. and C.L. developed the methodology. G.H. conducted all simulations and wrote the manuscript draft. T.N., J.C.C., A.G.R. and C.L. gave technical advice, provided the dataset, and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for this research was provided through the National Institutes of Health [1R01 AI 154470-01; 2U01 HG 008685; R01 HG 008976; U01 HL 089856; U01 HL 089897; P01 HL 120839; P01 HL 132825; R21 HD 095228], the National Science Foundation [NSF PHY 2033046; NSF GRFP 1745302], the National Institutes of Allergy and Infectious Disease [R01 AI 154470], and the NIH Center grant P30-ES002109.

Data Availability Statement: All data analysed in this study are included in published articles of the Pediatric Intensive Care Influenza (PICFLU) investigators group [2,3].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Watson, R.S.; Crow, S.S.; Hartman, M.E.; Lacroix, J.; Odetola, F.O. Epidemiology and Outcomes of Pediatric Multiple Organ Dysfunction Syndrome. *Pediatr. Crit. Care Med.* **2017**, *18*, S4–S16. [CrossRef] [PubMed]
2. Randolph, A. Pediatric Acute Lung Injury and Sepsis Investigators (PALISI) Research Network. Available online: www.palisi.org (accessed on 22 May 2023).
3. Randolph, A. Pediatric Intensive Care Influenza Network (PICFLU). Available online: <https://picflu.org> (accessed on 22 May 2023).
4. Park, T.; Yi, S.G.; Lee, S.; Lee, S.Y.; Yoo, D.H.; Ahn, J.I.; Lee, Y.S. Statistical tests for identifying differentially expressed genes in time-course microarray experiments. *Bioinformatics* **2003**, *19*, 694–703. [CrossRef] [PubMed]
5. Storey, J.D.; Xiao, W.; Leek, J.T.; Tompkins, R.G.; Davis, R.W. Significance analysis of time course microarray experiments. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12837–12842. [CrossRef] [PubMed]
6. Sun, W.; Wei, Z. Multiple Testing for Pattern Identification, With Applications to Microarray Time-Course Experiments. *J. Am. Stat. Assoc.* **2011**, *106*, 73–88. [CrossRef]
7. Yuan, M.; Kendzioriski, C. Hidden Markov Models for Microarray Time Course Data in Multiple Biological Conditions. *J. Am. Stat. Soc.* **2006**, *101*, 1323–1332. [CrossRef]
8. Wu, S.; Wu, H. More powerful significant testing for time course gene expression data using functional principal component analysis approaches. *BMC Bioinform.* **2013**, *14*, 1–13. [CrossRef] [PubMed]
9. Trapnell, C.; Cacchiarelli, D.; Grimsby, J.; Pokharel, P.; Li, S.; Morse, M.; Lennon, N.J.; Livak, K.J.; Mikkelsen, T.S.; Rinn, J.L. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **2014**, *32*, 381–386. [CrossRef]
10. Shalek, A.K.; Satija, R.; Shuga, J.; Trombetta, J.J.; Gennert, D.; Lu, D.; Chen, P.; Gertner, R.S.; Gaublot, J.T.; Yosef, N.; et al. Single cell RNA Seq reveals dynamic paracrine control of cellular variation. *Nature* **2014**, *510*, 363–369. [CrossRef] [PubMed]
11. Varoquaux, N.; Purdom, E. A pipeline to analyse time-course gene expression data [version 1; peer review: 2 approved with reservations]. *F1000Research* **2020**, *9*, 1–45. [CrossRef]
12. Smith, M. University of Texas—Inference for Contrasts (Chapter 4). Available online: <https://web.ma.utexas.edu/users/mks/384Esp08/infcontrast.pdf> (accessed on 22 May 2023).
13. National Institute of Standards and Technology (NIST). Assessing the Response from any Factor Combination. Available online: <https://www.itl.nist.gov/div898/handbook/prc/section4/prc436.htm> (accessed on 22 May 2023).
14. Bonferroni, C. Teoria statistica delle classi e calcolo delle probabilità. *Pubbl. Ist. Super. Sci. Econ. Commer. Firenze* **1936**, *8*, 3–62.
15. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [CrossRef]
16. Ruf, B.; Knuf, M. The burden of seasonal and pandemic influenza in infants and children. *Eur. J. Pediatr.* **2014**, *173*, 265–276. [CrossRef] [PubMed]
17. FluSurv-NET. Laboratory-Confirmed Influenza Hospitalizations. Available online: <https://gis.cdc.gov/grasp/fluview/fluhosprates.html> (accessed on 22 May 2023).
18. Randolph, A.G.; Agan, A.A.; Flanagan, R.F.; Meece, J.K.; Fitzgerald, J.C.; Loftis, L.L.; Truemper, E.J.; Li, S.; Ferdinands, J.M.; Lung, I.P.A.; et al. Optimizing Virus Identification in Critically Ill Children Suspected of Having an Acute Severe Viral Infection. *Pediatr. Crit. Care Med.* **2016**, *17*, 279–286. [CrossRef] [PubMed]
19. Kangelaris, K.N.; Prakash, A.; Liu, K.D.; Aouizerat, B.; Woodruff, P.G.; Erle, D.J.; Rogers, A.; Seeley, E.J.; Chu, J.; Liu, T.; et al. Increased expression of neutrophil-related genes in patients with early sepsis-induced ARDS. *Am. J. Physiol. Lung. Cell Mol. Physiol.* **2015**, *308*, L1102–1113. [CrossRef] [PubMed]
20. Sweeney, T.E.; Wong, H.R.; Khatri, P. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Sci. Transl. Med.* **2016**, *8*, 346ra391. [CrossRef] [PubMed]
21. Wong, H.R.; Cvijanovich, N.Z.; Anas, N.; Allen, G.L.; Thomas, N.J.; Bigam, M.T.; Weiss, S.L.; Fitzgerald, J.C.; Checchia, P.A.; Meyer, K.; et al. Improved Risk Stratification in Pediatric Septic Shock Using Both Protein and mRNA Biomarkers: PERSEVERE-XP. *Am. J. Respir. Crit. Care Med.* **2017**, *196*, 494–501. [CrossRef] [PubMed]
22. Matics, T.J.; Sanchez-Pinto, L.N. Adaptation and Validation of a Pediatric Sequential Organ Failure Assessment Score and Evaluation of the Sepsis-3 Definitions in Critically Ill Children. *JAMA Pediatr.* **2017**, *171*, e172352. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.