

Distance-based differential analysis of gene curves

Christopher Minas¹, Simon J. Waddell² and Giovanni Montana^{1,*}

¹Statistics Section, Department of Mathematics, Imperial College London, London SW7 2AZ and ²Clinical and Laboratory Investigation, Brighton and Sussex Medical School, University of Sussex, Brighton BN1 9PX, UK

Associate Editor: Janet Kelso

ABSTRACT

Motivation: Time course gene expression experiments are performed to study time-varying changes in mRNA levels of thousands of genes. Statistical methods from functional data analysis (FDA) have recently gained popularity for modelling and exploring such time courses. Each temporal profile is treated as the realization of a smooth function of time, or curve, and the inferred curve becomes the basic unit of statistical analysis. The task of identifying genes with differential temporal profiles then consists of detecting statistically significant differences between curves, where such differences are commonly quantified by computing the area between the curves or the l_2 distance.

Results: We propose a general test statistic for detecting differences between gene curves, which only depends on a suitably chosen distance measure between them. The test makes use of a distance-based variance decomposition and generalizes traditional MANOVA tests commonly used for vectorial observations. We also introduce the *visual* l_2 distance, which is shown to capture shape-related differences in gene curves and is robust against time shifts, which would otherwise inflate the traditional l_2 distance. Other shape-related distances, such as the curvature, may carry biological significance. We have assessed the comparative performance of the test on realistically simulated datasets and applied it to human immune cell responses to bacterial infection over time.

Availability: R scripts are freely available for download at <http://www2.imperial.ac.uk/~gmontana/dbf.htm>.

Contact: g.montana@imperial.ac.uk

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Received on June 17, 2011; revised on August 11, 2011; accepted on September 15, 2011

1 INTRODUCTION

In longitudinal microarray experiments, the temporal evolution of expression levels in thousands of genes are monitored in an attempt to understand the dynamic processes that regulate them (Storey *et al.*, 2005). Repeated measurements of mRNA are taken at a relatively small number of time-points for all the available biological replicates. A common aim of such studies is to compare gene expression profiles observed in different populations or under different experimental conditions, and to identify those genes whose temporal profiles differ significantly. The data produced by such longitudinal studies, however, present several challenges for

statistical analysis. Statistical tests developed for cross-sectional data, such as the t -test and its many modifications, are inadequate because they are only able to detect differential expression at individual time-points and they ignore the temporal dependencies that are typical of the experimental data (Storey *et al.*, 2005). Models from classical time series analysis are also limited in scope as the time courses are generally very short, sampled at irregularly spaced time-points, and often contain missing data.

Over the last few years, these issues have led to an increasing interest in the application of functional data analysis (FDA) techniques (Ramsay and Silverman, 2005). In FDA, the longitudinal observations are treated as noisy realizations of an underlying smooth curve, which represents the true continuous evolution of a gene's expression level over time. Once these underlying curves have been inferred from the observed data, they become the basic unit of statistical analysis, on which a wide variety of tasks can be performed such as dimensionality reduction and visualization (Hyndman and Shang, 2010), clustering (Heard *et al.*, 2006; Wang *et al.*, 2008) and differential analysis (Bar-Joseph *et al.*, 2003; Behseta and Kass, 2005). A recent review of developments in this area can be found in Coffey and Hinde (2011).

We consider the problem of detecting differences between gene curves representative of two or more groups when several biological replicates are available for each group. A number of test statistics have been proposed to test for equality between curves, including maSigPro (Conesa *et al.*, 2006), EDGE (Storey *et al.*, 2005), PACE (Liu and Yang, 2009) and functional Bayes approaches (Angelini *et al.*, 2007; Hong and Li, 2006). The underlying problem of detecting differences between curves has also been extensively considered in non-parametric statistics, and various tests have been developed that could be used for the analysis of gene expression temporal profiles, including the functional F -statistic (Shen and Faraway, 2004) and the more recent TN statistic (Zhang *et al.*, 2010). The underlying assumption shared by these methods is that, under the null hypothesis of equality between curves, the area between them is zero.

Our work is motivated by the observation that although the area between two curves, as captured by the l_2 distance, often provides an intuitive and easily interpretable measure of difference between them, existing methods only capture this specific type of difference. Thus, potentially interesting differences between gene curves that may carry biological relevance may be missed. For instance, the l_2 distance may fail to capture specific shape-related features that would otherwise be detected by the expert eye, such as the number and location of 'peaks' and 'troughs' or their respective curvatures. Furthermore, time-shifted curves, considered different when using the standard l_2 distance, might be considered similar using an

*To whom correspondence should be addressed.

alternative distance measure. This would mitigate the use of special alignment approaches to capture the similarity of such response patterns (Aach and Church, 2001; Smith *et al.*, 2009).

In this article, we propose a general distance-based test statistic for detecting differences between functional representations of gene expression temporal profiles that only relies on pairwise distances between replicate gene curves. The test statistic is obtained using a generalized distance-based variance decomposition and inferences are drawn using a permutation approach. The proposed test can be interpreted as a generalization of multivariate analysis of variance (MANOVA) methods to non-vectorial data. We also demonstrate that, when the area between gene curves is the preferred difference, the proposed test has the same statistical power as other functional tests, such as EDGE and TN. The major contribution of our approach consists of providing a unified statistical framework for detecting differential gene curves with respect to any notion of distance, which in turn enables the detection of a wide range of patterns in functional data.

The structure of the article is as follows. In Section 2.1, we introduce some notation and provide a brief overview of the FDA approach for the analysis of time course data where some limitations of the l_2 distance are also highlighted. In Section 2.2, we describe two shape-based distances for gene curves; the visual l_2 and curvature distances. In Section 2.3, we present our distance-based test statistic whose performance is assessed in Section 3 using realistically simulated curves. The potential benefits of our approach are discussed in Section 4 where it is applied to a longitudinal transcriptional dataset of human immune cell responses. A final discussion and conclusions are provided in Section 5.

2 METHODS

2.1 FDA

We consider a setting in which the expression levels of K genes are observed over a time period τ . We assume that each gene is modelled independently, and that the expression levels for each gene are observed for G different groups (e.g. representing different biological conditions) at S possibly irregularly spaced design time-points $t = (t_1, \dots, t_S)^T$. A random sample of size N_g is observed for group g , for $g = 1, \dots, G$, and consists of expression levels of biological replicates denoted by $y_i^{(g)}(t) = (y_i^{(g)}(t_1), \dots, y_i^{(g)}(t_S))^T$ for $i = 1, \dots, N_g$. Within each group g , these data points are assumed to be noisy observations taken from an underlying smooth function of time, $\mu^{(g)}$, observed at the design time-points. Several statistical models for estimating the underlying gene curves exist, including semiparametric mixed effect models using different parametric curves such as B-splines and smoothing splines (Aryee *et al.*, 2009; Berk and Montana, 2009; Ramsay and Silverman, 2005; Stegle *et al.*, 2010; Storey *et al.*, 2005); any of these methods could be used in our framework.

The hypothesis of equality between curves is commonly stated as

$$H_0: \mu^{(i)}(t) = \mu^{(j)}(t) \text{ versus } H_1: \mu^{(i)}(t) \neq \mu^{(j)}(t), \quad (1)$$

for all $t \in \tau$, where $i \neq j = 1, \dots, G$. Several existing statistical tests for functional data, including EDGE and TN, have been developed to test this null hypothesis of equality. A major limitation of these approaches is the implicit assumption that, in order to be considered equal, the curves must have the same amplitude over the entire time range τ . In other words, the area between them, or their l_2 distance, denoted d_L and given by

$$d_L(\mu^{(i)}, \mu^{(j)}) \equiv \left(\int_{\tau} (\mu^{(i)}(t) - \mu^{(j)}(t))^2 dt \right)^{\frac{1}{2}},$$

must be zero under the null. Thus, current methods either explicitly or implicitly test for zero d_L distances by construction. However, since d_L

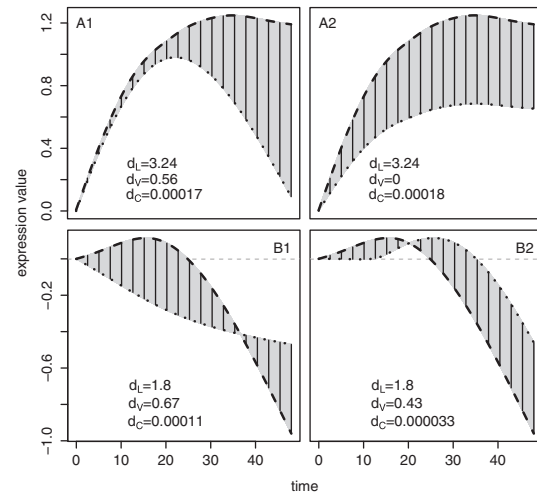


Fig. 1. Four different comparisons between two gene curves illustrating the effects of using the d_L , d_V and d_C distances. The curves in A1 and A2 have the same d_L distances (represented by the shaded regions with vertical lines) despite clearly visible differences in the temporal gene expression patterns. Similarly, the curves in B1 and B2 have the same d_L distances, although the curves in B1 have quite different time-varying behaviour while those in B2 have the same shape but are time-delayed. These shape-related differences are better captured by d_V and d_C .

is only concerned with *vertical* distances between points taken on each curve, expression profiles may not necessarily exhibit differences in the time-varying patterns of mRNA abundance even when having large d_L distances.

This is illustrated in Figure 1 where we consider two groups ($G = 2$). Each of the four plots shows a comparison between two gene curves, one curve for each group. The solid vertical lines indicate the vertical distances considered when computing d_L . For both the A1 and A2 comparisons, the d_L distance is the same ($d_L = 3.24$), indicating that the area between the two curves (shaded region) is the same in both cases. However, in these examples there is a clearly visible difference in their respective shapes. Specifically, halfway through the time course, the expression levels in A1 increasingly diverge as time progresses, whereas the expression levels in A2 both plateau. For both the B1 and B2 comparisons (bottom row), the d_L distance between curves is also the same ($d_L = 1.8$), indicating again that the areas between the curves are the same. However, whereas the two curves in B2 have very similar shapes and have only undergone a time-shift, the curves in B1 have different time-varying patterns, resulting in shapes more different than just being time-shifted. Clearly the d_L distance is unable to identify similar temporal profiles that only differ due to delays on the time scale. Such time-shifts can be representative of expression responses that may be slower in one group than another, due to a time lag in their transcription control for example, Qian *et al.* (2001).

These examples support the argument that the d_L distance may be unable to capture clearly visible differences in the *shape* of the expression profiles, and so existing tests that focus on this distance are expected to have very little statistical power in detecting certain shape-related differences. Below, we introduce alternative distances that enable the identification of shape-related patterns of variation for which no statistical procedures are currently available.

2.2 Shape-based distance measures

As already noted, the area between curves as computed by d_L only relies on vertical distances between the gene curves, and therefore does not truly capture the notion of shape. An alternative distance that overcomes this

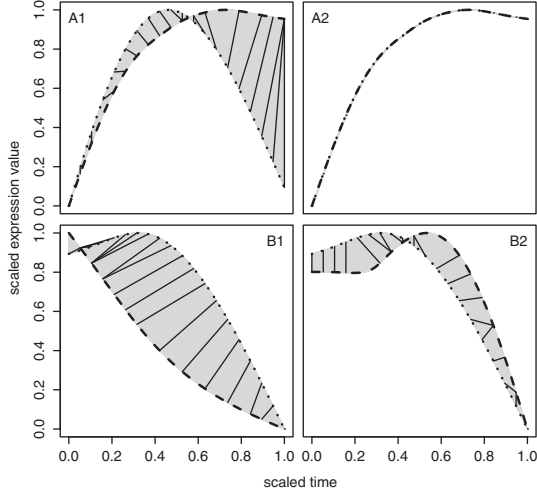


Fig. 2. Illustration of the d_V distance between two curves. Each plot refers to the corresponding comparison presented in Figure 1. In each case, the curves are scaled in both the x - and y -axes, with the shaded areas quantifying the region considered in the computation of the d_V distances. The solid lines connecting the two curves (no longer just vertical) represent some points on the curves deemed close by the naked eye.

limitation is the *visual* l_2 measure, denoted here as d_V , which was originally proposed in the context of qualitative smoothing (Marron and Tsybakov, 1995). In order to compute the distance between curves $\mu^{(i)}$ and $\mu^{(j)}$, scaling effects are initially removed by rescaling both axes to the $[0, 1]$ interval, thus ensuring that any differences detected between curves are truly differences in shape. The curves are then considered as infinite sets of points in the \mathbb{R}^2 plane so that rescaled curve $\mu^{(i)}$ is represented by a set of 2D points collected in

$$p_i = \{(t, \mu^{(i)}(t)) \mid t \in [0, 1]\}.$$

For a given time-point t , the distance between point $\mu^{(i)}(t)$ and curve $\mu^{(j)}$ is found as the shortest Euclidean distance between the 2D coordinate $(t, \mu^{(i)}(t))$ and all coordinates in p_j .

$$\delta(i, j) \equiv d_E^{\min}((t, \mu^{(i)}(t)), p_j).$$

The integral of this quantifies the difference between the two curves analogous to scanning all points on $\mu^{(i)}$ by the naked eye and detecting the corresponding closest points on $\mu^{(j)}$. That is, the features of $\mu^{(i)}$ such as peaks and troughs, for example, are compared with the closest features of $\mu^{(j)}$. The difference between $\mu^{(i)}$ and $\mu^{(j)}$ is then quantified in the same way, and these two comparisons are combined into the overall d_V distance

$$d_V(\mu^{(i)}, \mu^{(j)}) \equiv \left(\int_0^1 \delta(i, j)^2 dt + \int_0^1 \delta(j, i)^2 dt \right)^{\frac{1}{2}}.$$

The fact that d_V takes into account both vertical and horizontal directions yields a distance measure more representative of the distance detected by the naked eye. This is illustrated in Figure 2, in which we have shown how the d_V distance of the curves presented in Figure 1 are computed. The curves in the A1 comparison have a larger visual distance ($d_V = 0.56$) than those in A2 ($d_V = 0$) showing that once the difference in amplitude has been removed, the two curves have exactly the same shape. Also, d_V is smaller for the curves represented in the B2 comparison, than those in B1; similar conclusions would have been reached by visual exploration of the curves, which clearly shows that the two temporal profiles are time-shifted, but their shapes are otherwise very similar.

In some studies, the speed of differential gene expression over time is of interest, regardless of the changes in direction. In this case, a suitable

measure should quantify the difference in curvatures, and a suitable distance would be

$$d_C(\mu^{(i)}, \mu^{(j)}) \equiv \left| \int_{\tau} (\mu^{(i)''}(t))^2 dt - \int_{\tau} (\mu^{(j)''}(t))^2 dt \right|,$$

where μ'' indicates the curve's second derivative. This measure quantifies the difference in smoothness of the curves, and unlike d_L and d_V , will yield a zero value if the curves are perfect reflections of each other in the time axis (for example, one having a peak and the other having a trough). Such inverted temporal profiles can indicate inhibitory relationships between the groups (Shi *et al.*, 2007). In this case, on using d_C , the gene curves are considered similar. In Figure 1, d_C is smaller in B2 than in B1, thus also declaring the time-shifted curves more similar to each other than those in B1.

2.3 Distance-based hypothesis testing

We introduce a general hypothesis testing framework that can be flexibly adopted to work with any distance measure. Instead of (1), we reformulate the problem as

$$H_0: d(\mu^{(i)}, \mu^{(j)}) = 0 \quad \text{versus} \quad H_1: d(\mu^{(i)}, \mu^{(j)}) \neq 0, \quad (2)$$

where $i \neq j = 1, \dots, G$ and d is any suitable non-negative and symmetric dissimilarity measure between two curves computed over the entire time range τ . In this formulation, under H_0 , the curves do not have to be exactly equal to be similar, but are similar with respect to the chosen dissimilarity measure. In what follows, we develop a distance-based test statistic for (2).

As detailed in Section 2.1, for each given gene, we assume that a smoothing procedure has been applied to estimate all N_g curves available in group g , for every $g = 1, \dots, G$. Each estimated curve for a biological sample is denoted as $\hat{\mu}_i$. We denote by N the total number of replicate curves available across all G groups, and obtain an $N \times N$ distance matrix $\Delta = \{d(\hat{\mu}_i, \hat{\mu}_j)\}_{i,j=1}^N$ containing all pairwise distances among the N estimated curves $\{\hat{\mu}_i\}_{i=1}^N$, with elements altered if necessary so that the distances are metric (see Supplementary Material). In order to keep track of group membership, we introduce the indicator variable I_{gi} , which equals 1 if curve $\hat{\mu}_i$ is in group g and 0 otherwise.

Our objective is to test (2) using all the information stored in Δ and the grouping information. This is achieved by first decomposing the total sample variability into within- and between-group elements in a way that only depends on Δ and I_{gi} for all g, i , and then building a test statistic that makes use of this generalized decomposition.

In order to derive the variance decomposition, we first consider the case of N independent data points being represented by S -dimensional vectors, y_1, \dots, y_N , and arranged in an $N \times S$ matrix Y . We let T represent the $S \times S$ total sum of squares matrix,

$$T = \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^N (y_i - y_j)(y_i - y_j)^T.$$

As generally done in classical multivariate analysis of variance, this can be decomposed as $T = W + B$, where W is the within-group sum of squares matrix

$$W = \sum_{g=1}^G \sum_{i=1}^{N_g} (y_i - \bar{y}_g)(y_i - \bar{y}_g)^T I_{gi},$$

and

$$B = \sum_{g=1}^G N_g (\bar{y}_g - \bar{y})(\bar{y}_g - \bar{y})^T,$$

is the between-group sum of squares matrix, where we have denoted the group means by \bar{y}_g , and the overall mean by \bar{y} .

In order to quantify the variance in the sum of squares matrices, we take the trace operator, yielding $\text{tr}(T) = \text{tr}(W) + \text{tr}(B)$. This decomposition can be re-written in terms of the pairwise Euclidean distances between the N observations stored in the $N \times N$ distance matrix $\Delta_E = \{d_E(y_i, y_j)\}_{i,j=1}^N$.

This is done by taking the traces of \mathbf{T} , \mathbf{W} and \mathbf{B} and using the result that $\text{tr}((\mathbf{y}_i - \mathbf{y}_j)(\mathbf{y}_i - \mathbf{y}_j)^T) = d_E^2(\mathbf{y}_i, \mathbf{y}_j)$. On taking the trace of \mathbf{T} we define

$$\text{tr}(\mathbf{T}_{\Delta_E}) = \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^N d_E^2(\mathbf{y}_i, \mathbf{y}_j),$$

where $\text{tr}(\mathbf{T}) = \text{tr}(\mathbf{T}_{\Delta_E})$, showing that $\text{tr}(\mathbf{T})$ only depends on the pairwise Euclidean distances in Δ_E . Similarly, we obtain expressions for the traces of \mathbf{W} and \mathbf{B} in terms of Euclidean distances as

$$\text{tr}(\mathbf{W}_{\Delta_E}) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N d_E^2(\mathbf{y}_i, \mathbf{y}_j) \sum_{g=1}^G \frac{I_{gi} I_{gj}}{N_g},$$

and

$$\text{tr}(\mathbf{B}_{\Delta_E}) = \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^N d_E^2(\mathbf{y}_i, \mathbf{y}_j) \sum_{g=1}^G \left(I_{gi} - N \frac{I_{gi} I_{gj}}{N_g} \right),$$

where $\text{tr}(\mathbf{W}) = \text{tr}(\mathbf{W}_{\Delta_E})$ and $\text{tr}(\mathbf{B}) = \text{tr}(\mathbf{B}_{\Delta_E})$. The variance decomposition can be written in a distance-based equivalent form as $\text{tr}(\mathbf{T}_{\Delta_E}) = \text{tr}(\mathbf{W}_{\Delta_E}) + \text{tr}(\mathbf{B}_{\Delta_E})$, where all terms are non-negative as they are weighted sums of distances.

When the data are functional rather than vectorial, such as gene curves, a generalized distance-based variance decomposition is obtained simply by replacing d_E with any curve-based distance measure d , yielding

$$\text{tr}(\mathbf{T}_{\Delta}) = \text{tr}(\mathbf{W}_{\Delta}) + \text{tr}(\mathbf{B}_{\Delta}).$$

Using this expression, we then define the distance-based statistic

$$\text{DBF}_{\Delta} = \frac{\text{tr}(\mathbf{B}_{\Delta})}{\text{tr}(\mathbf{W}_{\Delta})},$$

which is non-negative due to $\text{tr}(\mathbf{B}_{\Delta})$ and $\text{tr}(\mathbf{W}_{\Delta})$ being non-negative. Large values of the statistic are due to a large between- and low within-group variance, and support the hypothesis that the true group measurements differ with respect to d .

The null distribution of DBF_{Δ} is dependent on the specific distance measure being used, and is generally unknown. Permutation testing can be performed by simultaneously permuting the rows and columns of Δ (so the zero diagonal elements are maintained) and subsequently re-evaluating DBF_{Δ} to assess whether or not an observed statistic is significant. This procedure requires a set of N_{π} permutations Π such that each $\pi \in \Pi$ is a one-to-one mapping of the vector $(1, \dots, N)^T$ to itself. On denoting the permuted distance matrix Δ_{π} for a given permutation π , the set $\{\text{DBF}_{\Delta_{\pi}}\}_{\pi \in \Pi}$ is generated giving the permutation distribution of DBF_{Δ} under the null from which the empirical P -value is found as

$$P = \frac{\#\text{DBF}_{\Delta_{\pi}} \geq \text{DBF}_{\Delta}}{N_{\pi}}.$$

Our proposed test statistic encompasses several known statistical tests proposed in multivariate analysis as special cases. When the observations are scalars, the DBF statistic applied with Euclidean distances is equivalent to the standard ANOVA F -statistic. In this case, under the null hypothesis of no difference in mean between the groups, DBF_{Δ} follows an F distribution with $G-1$ and $N-G$ degrees of freedom (after multiplying by suitable constants). When the observations are vectorial, MANOVA techniques are usually applied to test for equality of the true group means (Rencher, 2002). On using the Mahalanobis distance for vectors, the DBF test is related to the Hotelling's T^2 , Wilks' Lambda, Pillai trace and Lawley-Hotelling trace tests in the two group case, and is related to Pillai and Lawley-Hotelling trace tests when $G > 2$ (analytical derivations are provided in the Supplementary Material).

3 EXPERIMENTAL RESULTS

The performance of the proposed DBF test was assessed using artificially simulated curves for 250 genes with 225 under H_0 and

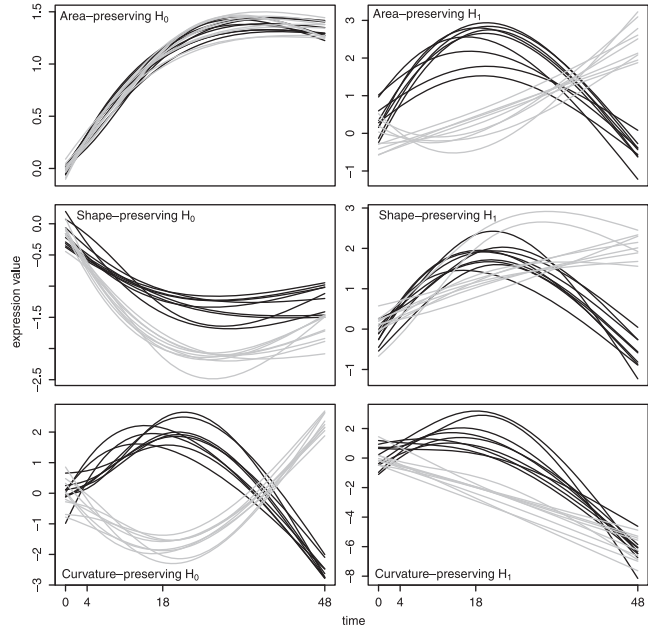


Fig. 3. Examples of simulated gene curves for each of the three experiments under the null and alternative hypotheses. Curves in Group 1 are black and those in Group 2 are gray.

25 under H_1 in a two group comparison. Independently for each gene we simulated group curves $\mu^{(g)}$ for $g=1,2$ over $\tau=[0,48]$ and took these as the known population curves. In order to create the observed dataset, each curve was observed at the design time-points $t = (0, 4, 18, 48)^T$. Finally, $N_g = 9$ replicates were assumed, for $g=1,2$, to emulate the real data described in Section 4. The salient features of the artificial population curves in each group depend on whether they are generated under the null or alternative hypotheses, and on a specific notion of distance. We performed three different experiments: (i) area-preserving, in which the null curves have zero d_L distances; (ii) shape-preserving, in which the null curves have zero d_V distances and (iii) curvature-preserving, in which the null curves have zero d_C distances. The challenge here consists of simulating population curves that preserve the prescribed distances and resamples experimental data. We used Bezier curves [see Farin (1992), for example] whose parameters can be randomly generated. The specific simulation settings are described in the Supplementary Material, and examples of simulated curves in each experiment are provided in Figure 3. It can be noted that the sample curves drawn under the null hypothesis (left plots) preserve the zero distances.

The DBF test is compared with EDGE and TN, which have been proposed to detect large differences in the area between curves, according to (1), and are directly comparable with the DBF test with the d_L distance. In this comparison, we also include the distance-based Mantel test (Mantel, 1967), defined as

$$\sum_{i < j}^N d(\hat{\mu}_i, \hat{\mu}_j) \sum_{g=1}^G I_{gi} I_{gj},$$

where $\{\hat{\mu}_i\}_{i=1}^N$ are the N simulated replicate curves across both groups and d is the selected distance measure. This statistic has been applied in ecology (Legendre and Legendre, 1998) and genetics

Table 1. Power (with SDs) of the DBF, Mantel, TN and EDGE test statistics for three false positive rates (FPRs) in all three simulation experiments

FPR (%)	Area-preserving			Shape-preserving			Curvature-preserving		
	1	5	10	1	5	10	1	5	10
DBF	0.93 (0.06)	0.97 (0.04)	0.97 (0.03)	0.64 (0.10)	0.74 (0.08)	0.79 (0.07)	0.86 (0.08)	0.96 (0.04)	0.98 (0.03)
Mantel	0.85 (0.08)	0.90 (0.07)	0.92 (0.06)	0.30 (0.11)	0.50 (0.12)	0.60 (0.10)	0.05 (0.05)	0.18 (0.09)	0.30 (0.11)
TN	0.93 (0.06)	0.97 (0.03)	0.98 (0.03)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
EDGE	0.96 (0.04)	0.98 (0.03)	0.98 (0.03)	0.05 (0.05)	0.15 (0.05)	0.20 (0.04)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

DBF with d_L distances is competitive with EDGE and TN for the area-preserving experiment. EDGE and TN have no power to detect shape-related differences between the groups. In all cases, DBF outperforms the Mantel test which only uses between-group distances and therefore has less power.

(Legendre and Fortin, 2010), with large values supporting the alternative hypothesis as large between-group distances are usually indicative of differences between groups. However, we note that while between-group distances may be large, the relative size of the within-group distances are ignored, potentially masking the true signal of differential expression.

The power of the tests are computed using 200 Monte Carlo simulations, and are reported in Table 1 for 1, 5 and 10% false positives rates. It is seen that EDGE and TN perform well for the area-preserving experiment where differentially expressed genes were designed to exhibit large differences in area between the two group profiles. Here, DBF is competitive with these approaches, and they all outperform Mantel. As expected, for the shape- and curvature-preserving experiments EDGE and TN have no power to detect the differential curves. Moreover, DBF always outperforms Mantel. This is because DBF considers both within- and between-group distances rather than just between-group distances. In particular, Mantel suffers a large reduction in power for the curvature-preserving experiment due to the low order of magnitude of the d_C distances masking the elements of the signal of differential expression provided by the between-group distances. The DBF statistic detects this signal due its ratio formulation of between- to within-group distances. (See Supplementary Material for more results).

4 APPLICATION TO HUMAN IMMUNE CELL DATA

Processes that may contribute to controlling *Mycobacterium tuberculosis* infection (Tailleux *et al.*, 2008) may be highlighted by comparing the transcriptional profiles of human phagocytic immune cells (macrophage and dendritic cells, denoted M/DCs) that differ in their ability to limit bacterial growth. Time course measurements of M/DC gene expressions were recorded at 0, 4, 18 and 48 h after infection using Affymetrix U133A high-density oligonucleotide arrays. The observation of each type of cell at each time-point was repeated with human immune cells isolated from 9 healthy donors. After preprocessing and removal of missing data, 10995 probe sets remained for the differential analysis (Tailleux *et al.*, 2008). Temporal profiles of all genes were smoothed using cubic smoothing splines, after which transcriptional responses were normalized at baseline so that any differences detected were relative to pre-infection state. The DBF test was used with the d_L , d_V and d_C distances. Empirical P -values were obtained by using 24310 permutations, and were subsequently adjusted using a multiple testing correction to control the false discovery rates at 1% (Storey and Tibshirani, 2003).

As expected, each dissimilarity measure identified different subsets of genes as differentially expressed between M and DCs in response to *Mycobacterium tuberculosis* infection (see Venn diagram in the Supplementary Material). We computed the hypergeometric probabilities of each gene belonging to pre-defined gene ontology (GO) classifications of genes (Ashburner *et al.*, 2000). This resulted in a set of functional categories which were significantly enriched with each dissimilarity measure, providing an overview of the pathways that are likely to be reprogrammed in dendritic cells compared with macrophages after infection. Gene Ontology (GO) terms for membrane invagination (GO:0010324) and endocytosis (GO:0006897, GO:0016193, GO:0016196), the process whereby phagocytic cells (such as M/DCs) engulf foreign bodies (such as *M. tuberculosis* bacilli), significantly overlapped with genes recognized using the DBF test with d_L . Additionally, genes associated with the endosome (GO:0005768) and late endosome (GO:0005770), the membrane structures containing foreign bodies that are formed during endocytosis, were significantly enriched only using the DBF test with d_V dissimilarities. Thus, the d_V measure that discriminates gene curves by shape, identified subtle changes in gene expression between the cell types that did not rely on large differences in amplitude between time-points. This suggests that processes controlling engulfment of *M. tuberculosis* bacilli may differ between phagocytic cell types permissive and non-permissive for mycobacterial growth.

We also looked at a subset of genes involved in intracellular trafficking, the RAB GTPases. These are a family of small guanosine triphosphatases found on the surface of intracellular membranes that play integral roles in regulating the movement of these vesicles around the cell (Brumell and Scidmore, 2007). The retention of RAB5 and the failure to recruit RAB7 has been used to characterize the stalled development of the *M. tuberculosis*-containing phagosome (Brumell and Scidmore, 2007). Genes encoding RAB7A and RAB7L1 were identified to be differentially regulated between M/DCs using the DBF tests with multiple measures (RAB7A with d_L , d_V and d_C , and RAB7L1 with d_L and d_V). Figure 4 displays the M/DC expression profiles for a selection of genes, of which RAB7A is the first. We see that there is a large difference in area between the two curves (d_L distance), large scale-invariant differences in shape (d_V distance) and the macrophage curve changes shape much faster than the dendritic curve (d_C distance).

RAB5B and RAB5C were only revealed to be divergently expressed using the DBF test with the d_L measure, as was also the case for RAB22A (Fig. 4B), which has been implicated in the reprogramming of *M. tuberculosis*-phagosome trafficking (Brumell

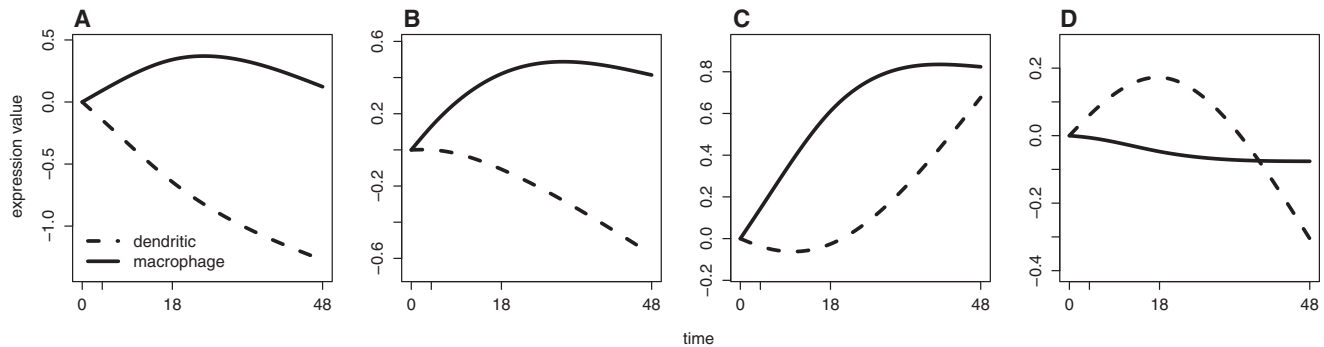


Fig. 4. Macrophage (solid) and dendritic (dashed) expression profiles for genes (A) RAB7A; (B) RAB22A; (C) RAB13; and (D) RND1, all identified by the DBF test with different dissimilarities; RAB7A was identified with d_L , d_V and d_C , RAB22A with d_L , RAB13 with d_V and RND1 with d_C .

and Scidmore, 2007). The differential expression of RAB13 (Fig. 4C) and RAB21 were only detected when considering the differences in the scale-invariant expression profiles as determined by d_V . For these genes, the divergent pattern of gene expression over time indicates that distinct processes are impacting upon intracellular trafficking in M compared with DC after mycobacterial infection, highlighting pathways of interest for further investigation and identifying inflection time-points where the interactions between host and pathogen appear to alter. RND1 (Fig. 4D) and RND3, Rho family GTPases, were only identified when considering the d_C measure between the expression profiles. Here, large differences in the speed with which the profiles changed over time were captured, whereas the direct time-point comparisons of the d_L measure did not detect significant differences in amplitude.

5 DISCUSSION AND CONCLUSIONS

In this work, we have presented a distance-based test for the detection of statistically significant differences between groups of curves representing gene expression temporal profiles. Compared with existing tests that have power to detect time-varying differences in expression profiles characterized by large areas, our approach provides additional flexibility by enabling the use of any distance measure that is most suitable for the specific application. In an exploratory study, a variety of distance measures could be used in order to detect alternative patterns. In particular, we have suggested a *visual* distance between curves that can identify differences in shape such as those detected by visual exploration. This distance has been shown to be robust to time shifts and other perturbations that do not alter the shape of the response curve.

Using realistically generated curves that closely resemble the experimental data, we have shown that DBF can achieve the same statistical power of other functional tests specifically designed to detect amplitude-driven differences between gene curves. In our case study, the identification of RAB13 and RAB21 genes in particular demonstrates that shape-based distances may be employed to discriminate genes by the geometric patterns of their transcriptional response over time. Such analyses may be desired in scenarios where large changes in amplitude of gene expression are not a pre-requisite, or where the differential actions of transcriptional regulators are of interest.

The emphasis of this article has been on functional data for longitudinal experiments, and gene expression profiles in particular. When the response profiles are described by vectorial rather than functional representations (e.g. a vector of expression values at all design time-points), the DBF test provides a distance-based alternative to common MANOVA tests for detecting mean differences using the Mahalanobis distance. However, apart from gene expression data, the DBF test has a much wider applicability and can be used to detect group differences involving any biological entity, regardless of the specific data structure, as long as a meaningful pairwise distance matrix can be defined. For entities best represented by non-vectorial data types such as sequences, protein 3D structures and networks, a variety of distance measures have been proposed in the literature and would be promptly available.

ACKNOWLEDGEMENTS

The authors would like to thank M. Pelizzola, A. Mortellaro, P.R. Castagnoli and M. Foti who generated and analyzed the original M/DC dataset, and L. Tailleux and O. Neyrolles who developed the *M.tuberculosis* infection model. The datasets utilized in this study are available as detailed in the original article describing work funded by the EC (LSHP-CT-2004-012187).

Funding: EPSRC Grant (to C.M. and G.M.).

Conflict of Interest: none declared.

REFERENCES

- Aach, J. and Church, G. (2001) Aligning gene expression time series with time warping algorithms. *Bioinformatics*, **17**, 495–508.
- Angelini, C. et al. (2007) A Bayesian approach to estimation and testing in time-course microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **6**, Article 24, [Epub ahead of print, September 16, 2007].
- Aryee, M.J. et al. (2009) An improved empirical bayes approach to estimating differential gene expression in microarray time-course data: BETR (Bayesian Estimation of Temporal Regulation). *BMC Bioinformatics*, **10**, 409–418.
- Ashburner, M. et al. (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
- Bar-Joseph, Z. et al. (2003) Comparing the continuous representation of time-series expression profiles to identify differentially expressed genes. *Proc. Natl Acad. Sci. USA*, **100**, 10146–10151.
- Behseta, S. and Kass, R.E. (2005) Testing equality of two functions using BARS. *Stat. Med.*, **24**, 3523–3534.

- Berk, M. and Montana, G. (2009) Functional modeling of microarray time series with covariate curves. *Statistica*, **2**, 153–177.
- Brumell, J.H. and Scidmore, M.A. (2007) Manipulation of rab GTPase function by intracellular bacterial pathogens. *Microbiol. Mol. Biol. Rev.*, **71**, 636–652.
- Coffey, N. and Hinde, J. (2011) Analyzing time-course microarray data using functional data analysis - a review. *Stat. Appl. Genet. Mol. Biol.*, **10**, Article 23, [Epub ahead of print, May 3, 2011].
- Conesa, A. *et al.* (2006) maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics*, **22**, 1096–1102.
- Farin, G. (1992) *Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide*. Academic Press Inc., San Diego, USA.
- Heard, N.A. *et al.* (2006) A quantitative study of gene regulation involved in the immune response of Anopheline Mosquitoes: an application of Bayesian hierarchical clustering of curves. *J. Am. Stat. Assoc.*, **101**, 18–29.
- Hong, F. and Li, H. (2006) Functional hierarchical models for identifying genes with different time-course expression profiles. *Biometrics*, **62**, 534–544.
- Hyndman, R.J. and Shang, H.L. (2010) Rainbow plots, bagplots, and boxplots for functional data. *J. Comput. Graph. Stat.*, **19**, 29–45.
- Legendre, P. and Fortin, M.J. (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Mol. Ecol. Resour.*, **10**, 831–844.
- Legendre, P. and Legendre, L. (1998) Numerical ecology. *Dev. Environ. Model.*, **20**.
- Liu, X. and Yang, M.C.K. (2009) Identifying temporally differentially expressed genes through functional principal components analysis. *Biostatistics*, **10**, 667–679.
- Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res.*, **27**, 209–220.
- Marron, J.S. and Tsybakov, A.B. (1995) Visual error criteria for qualitative smoothing. *J. Am. Stat. Assoc.*, **90**, 499–507.
- Qian, J. *et al.* (2001) Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions. *J. Mol. Biol.*, **314**, 1053–1066.
- Ramsay, J.O. and Silverman, B.W. (2005) *Functional Data Analysis*. 2nd edn. Springer, New York, USA.
- Rencher, A. (2002) *Methods of Multivariate Analysis*. John Wiley & Sons, Inc., New York, USA.
- Shen, Q. and Faraway, J. (2004) An F test for linear models with functional responses. *Stat. Sin.*, **14**, 1239–1258.
- Shi, Y. *et al.* (2007) Inferring pairwise regulatory relationships from multiple time series datasets. *Bioinformatics*, **23**, 755–763.
- Smith, A.A. *et al.* (2009) Clustered alignments of gene-expression time series data. *Bioinformatics*, **25**, i119–i1127.
- Stegle, O. *et al.* (2010) A robust Bayesian two-sample test for detecting intervals of differential gene expression in microarray time series. *J. Comput. Biol.*, **17**, 355–367.
- Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.
- Storey, J.D. *et al.* (2005) Significance analysis of time course microarray experiments. *Proc. Natl Acad. Sci. USA*, **102**, 12837–12842.
- Tailleux, L. *et al.* (2008) Probing host pathogen cross-talk by transcriptional profiling of both *Mycobacterium tuberculosis* and infected human dendritic cells and macrophages. *PLoS One*, **3**, e1403.
- Wang, H. *et al.* (2008) Nonparametric clustering of functional data. *Stat. Interface*, **1**, 47–62.
- Zhang, C. *et al.* (2010) Two samples tests for functional data. *Commun. Stat. Theory Methods*, **39**, 559–578.