

## GAMMA-BASED CLUSTERING VIA ORDERED MEANS WITH APPLICATION TO GENE EXPRESSION ANALYSIS

BY MICHAEL A. NEWTON AND LISA M. CHUNG

*University of Wisconsin*

It can be useful to know the probabilities that  $K$  independent Gamma-distributed random variables attain each of their  $K!$  possible orderings. Each ordering event is equivalent to an event regarding independent negative-binomial random variables, and this finding guides a dynamic-programming computation. Gamma-rank probabilities are central to a model-based clustering method for multi-group gene expression analysis, which is evaluated, demonstrated, and compared to alternative strategies. The structuring of model components according to inequalities among latent means leads to strict concavity of the mixture log likelihood. The clustering method applies to expression data collected by microarrays or by next-generation sequencing.

**1. Introduction.** A common problem in statistical genomics is how to organize expression data from genes that have been determined to exhibit differential expression relative to various cellular states. Cells in a time-course experiment may exhibit such genes, as may cells in any sort of designed experiment or observational study where expression alterations are being examined (e.g., Parmigiani *et al.*, 2003; Speed, 2004). In the event that the error-rate-controlled list of significantly altered genes is small, the post-processing problem amounts to inspecting observed patterns of expression, investigating what is known about the relatively few genes identified, and planning follow-up experiments as necessary. However, it is all too common that hundreds or even thousands of genes are detected as significantly altered in their expression pattern relative to the cellular states. Post-processing these non-null genes presents a substantial statistical problem. Difficulties are compounded in the multi-group setting because a gene can be non-null in many different ways (Jensen *et al.* 2009).

Ever since Eisen *et al.* (1998), clustering methods have been used to organize expression data. Thalamuthu *et al.* (2006) provides a recent perspective. Clustering methods are often applied in order to partition non-null genes which have been identified in differential expression analysis (e.g., Campbell *et al.* 2006; Grasso *et al.* 2008). Popular approaches are informative but not completely satisfactory. There are idiosyncratic problems, like how to select the number of clusters, but there is also the subtle issue that the clusters

identified by most algorithms are anonymous: each cluster is defined only by similarity of its contents rather than by some external pattern that its genes may be approximating. Anonymity may contribute to technical problems, such as that the objective function being minimized is not convex, and that realized clusters have a more narrow size distribution than is warranted by the biological system.

Model-based clustering treats data as arising from a mixture of component distributions, and then forms clusters by assigning each data point to its most probable component (e.g., Titterton *et al.* 1985; McLachlan and Basford, 1988). For example, the `mclust` procedure is based on mixtures of Gaussian components (Fraley and Raftery, 2002); the popular `K-means` algorithm is implicitly so based (Hastie *et al.*, 2001, page 463). There is considerable flexibility in model-based clustering, though technical challenges have also affected its development: the likelihood function is often multimodal; identifiability can be difficult to establish (e.g., Redner and Walker, 1984; Holzmann *et al.*, 2006); and even where constraints may create identifiability, there can be a problem of label-switching during Bayesian inference (Stephens, 2000). Some sophisticated model-based clustering methods have been developed for gene expression (e.g., Medvedovic *et al.*, 2004). Beyond empirical studies it is difficult to determine properties of such approaches, and their reliance on Monte Carlo computation is somewhat limiting.

Here a model-based clustering method is developed that aims to support multi-group gene expression data analysis and possibly other applications. The method, called gamma ranking, places genes in a cluster if their expression patterns commonly approximate one element from a finite catalog of possible structures, in contrast to anonymous methods (Section 3). Under certain conditions the component distributions are linearly independent functions – each one associated with a structure in the finite catalog – and this confers favorable computational characteristics to the gamma ranking procedure (Sections 4,5). The cataloged structures record patterns of equality and inequality among latent expected values. Where normal-theory specifications seem to be intractable, a gamma-based mixture model produces closed formulas for all necessary component densities, thanks to an embedding of the relevant gamma-distributed variables in a set of Poisson processes (Section 2). The formulation also extends to Poisson-distributed responses that are characteristic of gene expression measured by next-generation sequencing (Section 6).

**2. Gamma-rank probabilities.** The following statistical computing problem must be solved in order to implement gamma ranking. Specifically,

it is required to calculate the probability  $P(E)$  of the event

$$(1) \quad E = \{Z_1 > Z_2 > \dots > Z_K\}$$

where  $\{Z_k : k = 1, 2, \dots, K\}$  are mutually independent gamma distributed random variables with possibly different shape and rate parameters. Here, the marginal density of  $Z_k$  is

$$p(z) = \frac{\lambda_k^{a_k} z^{a_k-1} \exp\{-z\lambda_k\}}{\Gamma(a_k)}, \quad z > 0$$

and  $a_k, \lambda_k > 0$ . In the special case  $K = 2$ , the event in two gamma-distributed variables is equivalent to the  $E' = \{B > \lambda_2/(\lambda_1 + \lambda_2)\}$ , where  $B$  is a Beta( $a_1, a_2$ ) distributed variable. Thus  $P(E) = P(E')$  can be computed by standard numerical approaches for the Beta distribution. Although a similar representation is possible for Dirichlet-distributed vectors when  $K > 2$ , a direct numerical approach is not clearly indicated. In modeling permutation data, Stern (1990) presented a formula for  $P(E)$  for any value  $K$ , but assuming common shape parameters  $a_k = a$ . Sobel and Frankowski (1994) calculated  $P(E)$  for  $K < 5$  and assuming constant rates  $\lambda_k = \lambda$ , but to our knowledge a general formula has not been developed. A Monte Carlo approximation is certainly feasible, but a fast and accurate numerical approach would be preferable for computational efficiency: target values may be small, and  $P(E)$  may need to be recomputed for many shape and rate settings.

There is an efficient numerical approach to computing  $P(E)$  when shapes  $a_k$  are positive integers. The approach involves embedding  $\{Z_k\}$  in a collection of independent Poisson processes  $\{\mathbb{N}_k\}$ , where  $k = 1, 2, \dots, K$ . Specifically, let  $\mathbb{N}_k$  denote a Poisson process on  $(0, \infty)$  with rate  $\lambda_k$ . So  $\mathbb{N}_k(0, t] \sim \text{Poisson}(t\lambda_k)$ , for example. Of course, gaps between points in  $\mathbb{N}_k$  are independent and exponentially distributed, and the gamma-distributed  $Z_k$  can be constructed by summing the first  $a_k$  gaps,

$$Z_k = \min \{t > 0 : \mathbb{N}_k(0, t] \geq a_k\}.$$

Next, form processes  $\{\mathbb{M}_k\}$  by accumulating points in the originating processes:  $\mathbb{M}_k = \sum_{j=1}^k \mathbb{N}_j$ . Marginally  $\mathbb{M}_k$  is a Poisson process with rate  $\Lambda_k = \sum_{j=1}^k \lambda_j$ , but over  $k$  the processes are dependent owing to overlapping points. To complete the construction define count random variables  $M_1, M_2, \dots, M_{K-1}$  by

$$(2) \quad M_k = \mathbb{M}_k(0, Z_{k+1}].$$

It is immediate that each  $M_k$  has a marginal negative binomial distribution: the gamma distributed  $Z_{k+1}$  is independent of  $\mathbb{M}_k$ ; conditioning on  $Z_{k+1}$  in (2) gives a Poisson variable which mixes out to the negative binomial (Greenwood and Yule, 1920). Specifically,

$$M_k \sim \text{NB}(\text{shape} = a_{k+1}, \text{scale} = \Lambda_k / \lambda_{k+1}),$$

which corresponds to the probability mass function

$$(3) \quad p_k(m) = \frac{\Gamma(m + a_{k+1})}{\Gamma(a_{k+1}) \Gamma(m + 1)} \left( \frac{\lambda_{k+1}}{\Lambda_{k+1}} \right)^{a_{k+1}} \left( \frac{\Lambda_k}{\Lambda_{k+1}} \right)^m$$

for integers  $m \geq 0$ . The first main finding is

**THEOREM 1.** *With  $E$  as in (1),  $M_k$  as in (2), and  $p_k$  as in (3),  $P(E)$  equals*

$$(4) \quad \sum_{m_1=0}^{a_1-1} \sum_{m_2=0}^{m_1+a_2-1} \cdots \sum_{m_{K-1}=0}^{m_{K-2}+a_{K-1}-1} p_1(m_1) p_2(m_2) \cdots p_{K-1}(m_{K-1}).$$

It does not seem to be obvious that  $E$  in (1) is equivalent to an event in the  $\{M_k\}$ . We also find it striking that the  $M_k$  variables are independent considering that they are constructed from highly dependent  $\mathbb{M}_k$  processes. Proof of (4) and the related distribution theory are presented in Appendix A.

A redistribution of products and sums allows a numerically efficient evaluation of (4), as in the sum-product algorithm (e.g., Kschischang *et al.* 2001). For instance with  $K = 4$ ,

$$(5) \quad P(E) = \sum_{m_1=0}^{a_1-1} p_1(m_1) \left\{ \sum_{m_2=0}^{m_1+a_2-1} p_2(m_2) \left[ \sum_{m_3=0}^{m_2+a_3-1} p_3(m_3) \right] \right\}.$$

Here one would evaluate  $P(E)$  by first constructing for each  $m_2 \in \{0, 1, \dots, a_1 + a_2 - 2\}$  an inner sum  $P(M_3 \leq m_2 + a_3 - 1)$ . This vector in  $m_2$  values is used to process the second inner sum, for each value  $m_1 \in \{0, 1, \dots, a_1 - 1\}$ . Indeed the computation is completely analogous to the Baum-Welch backward recursion (e.g., Rabiner, 1989), although, interestingly, there seems to be no hidden Markov chain in the system. A version of the Viterbi algorithm identifies the maximal summand and thus provides an approach to computing  $\log P(E)$  in case  $P(E)$  is very small.

**3. Mixture of structured components.** Gamma ranking entails clustering genes according to the fit of a specific mixture model of gene-level data. With  $x_g = (x_{g,1}, x_{g,2}, \dots, x_{g,n})$  denoting the vector of expression measurements on gene  $g$  (*i.e.*, across all samples in all groups), let  $p(x_g)$  be the joint probability density for  $x_g$  induced by mixing over discrete *structures*  $\eta$ . Specifically,

$$(6) \quad p(x_g) = \sum_{\eta} p(x_g|\eta) \pi_{\eta},$$

where  $\pi_{\eta}$  is a mixing proportion and the component density  $p(x_g|\eta)$  is determined through modeling. Each expression measurement  $x_{g,i}$  is treated as a positive, continuous variable representing a fluorescence intensity from a microarray, after pre-processing has adjusted for various systematic effects not related to the groupings of interest. Recent technological advances allow expression to be measured instead as an explicit abundance count. The mixture model adapts readily to this case (Section 6).

Posterior structure probabilities are  $p(\eta|x_g) = p(x_g|\eta)\pi_{\eta}/p(x_g)$  and these determine gene clusters by Bayes's rule assignment. Alternatively the cluster contents can be regulated by a threshold parameter  $c$ , and

$$\text{cluster}(\eta) = \{g : p(\eta|x_g) \geq c\},$$

though some genes may go unassigned in this formulation.

The component density  $p(x_g|\eta)$  takes a special form in gamma ranking. It is derived from integration in a hierarchical model in which gene-level data are gamma distributed conditionally on latent expected values, and in which these latent values also fluctuate, but respecting the structure  $\eta$  (Appendix B). The gamma observation component is often supported empirically; there is theoretical support from stochastic models of population abundance (Dennis and Patil, 1984; Rempala and Pawlikowska, 2008); and there are practical considerations that a gamma-based model may be the only one for continuous data in which ordering calculations are feasible.

In gamma ranking  $p(x_g|\eta)$  is structured according to characteristics of  $\eta$ , which indicates a pattern of equality and inequality constraints among latent expected expression levels. The two-group comparison entails two latent means and thus three structures, denoted  $\{(1)(2), (12), (2)(1)\}$ . The notation conveys both the partition of the group means and the ordering of subsets within the partition. For instance, in  $\eta = (2)(1)$  the expected expression level in group 2 is less than that of group 1; while  $\eta = (12)$  indicates equality of these two latent means. With  $p = 3$  groups there are

13 structures,

$$(123), (12)(3), (3)(12), (13)(2), (2)(13), (1)(23), (23)(1), \\ (1)(2)(3), (2)(1)(3), (1)(3)(2), (2)(3)(1), (3)(1)(2), (3)(2)(1),$$

and the number grows rapidly with the number of groups (Table 1). A simple way to think about  $H_{\text{ord}} = \{\eta\}$ , the set of these ordered structures on  $p$  groups, is to imagine  $p$  real values  $y = (y_1, y_2, \dots, y_p)$  and the possible vectors you would get by ranking  $y$ . Of course there are  $p!$  rankings if ties are not permitted, but generally there are far more rankings, and  $H_{\text{ord}}$  is in 1-1 correspondence with the set of rankings of  $p$  numbers, allowing ties.

TABLE 1

*The number of ordered structures, Bell+, as a function of the number of groups. This is  $\sum_{k=1}^p (k!)S(p, k)$ , where  $S(p, k)$  are Stirling numbers of the second kind. The Bell number of partitions of  $1, \dots, p$  is included for comparison.*

$p$	Bell+	Bell
2	3	2
3	13	5
4	75	15
5	541	52
6	4683	203

An ordered structure  $\eta$  also dictates an association between sample labels  $i \in \{1, 2, \dots, n\}$  and levels of latent expected values; this association is a partition of the samples together with an ordering of the subsets within the partition. In the null structure the partition has a single element containing all samples, but otherwise there are  $K_\eta > 1$  disjoint subsets, say  $\sigma(\eta, k)$ ,  $k = 1, 2, \dots, K_\eta$ , with index  $k$  ordering the expected values. For example, suppose that samples  $\{1, 2, \dots, 6\}$  constitute two replicate samples in each of three groups, and  $\eta = (13)(2)$  is considered to relate the group-specific expected values. Then  $K_\eta = 2$ ,  $\sigma(\eta, 1) = \{1, 2, 5, 6\}$  and  $\sigma(\eta, 2) = \{3, 4\}$ . Subset  $\sigma(\eta, k)$ , has size  $n_k$  and gene-level statistic

$$s_{g,k} = \sum_{i \in \sigma(\eta,k)} x_{g,i}.$$

The second main finding is a formula for the density of observations for each structure  $\eta$ .

**THEOREM 2.** *With the model defined above for data  $x_g = (x_{g,1}, \dots, x_{g,n})$*

from  $p$  groups, the joint density  $p(x_g|\eta)$  equals

$$(7) \quad c_\eta \left( \prod_{i=1}^n x_{g,i}^{\alpha-1} \right) \underbrace{\prod_{k=1}^{K_\eta} \left( s_{g,k} + \frac{\alpha_0 \nu_0}{\alpha} \right)^{-a_k}}_{\text{center}(\eta)} \underbrace{P(Z_1 > Z_2 > \dots > Z_{K_\eta})}_{P_{\text{ord}}(\eta)}$$

where the  $Z_k$ 's are mutually independent gamma-distributed random variables with shapes  $a_k = \alpha_0 + \alpha n_k$  and rates  $\lambda_k = \alpha_0 \nu_0 + \alpha s_{g,k}$ , and where the normalizing constant is

$$(8) \quad c_\eta = \frac{K_\eta!}{[\Gamma(\alpha)]^n [\Gamma(\alpha_0)]^{K_\eta}} \left( \frac{\alpha_0 \nu_0}{\alpha} \right)^{\alpha_0 K_\eta} \prod_{k=1}^{K_\eta} \Gamma(a_k).$$

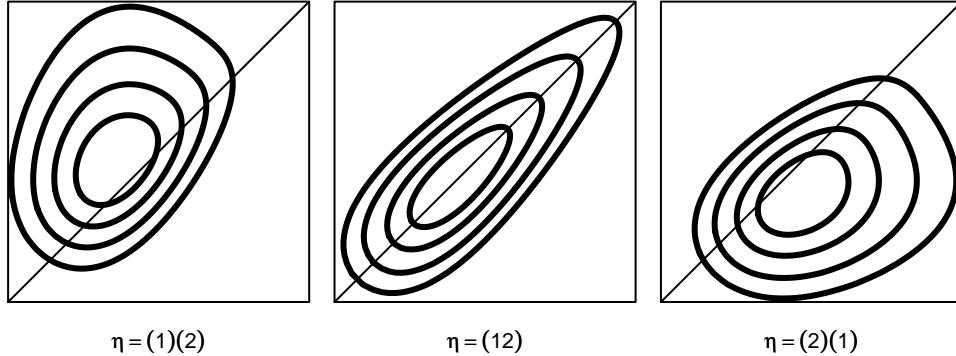
Here  $\alpha, \alpha_0$  and  $\nu_0$  are positive parameters related to within-group variation, between-group variation, and the overall scale of measurement, respectively (Appendix B). When  $K_\eta = 1$ ,  $P_{\text{ord}}(\eta) = 1$ .

In the null structure, for any  $p$ , (*i.e.*,  $K_\eta = 1$  and  $\eta = (12 \dots p)$ ), the distribution is exchangeable and equals a multivariate compound gamma (Hutchinson, 1981). Inspection also confirms that if the random  $X = (X_1, \dots, X_n)$  has density  $p(x|\eta)$  in (7), and if  $b > 0$ , then  $Y = (bX_1, \dots, bX_n)$  has a density of the same type, with shape parameters  $\alpha_0$  and  $\alpha$  unchanged, but with scale parameter  $b\nu_0$ . Special cases of the density (7) have been reported: Newton *et al.* (2004) presented the case  $p = 2$ ; Jensen *et al.* (2009) presented the case  $p = 4$ . See also Yuan and Kendzioriski (2006a). Evidently an algorithm for gamma-rank probabilities, such as the one in Section 2, is required in order to evaluate the component mixing densities. Beyond the  $p = 2$  case, previous reports have evaluated the factor  $P_{\text{ord}}(\eta)$  by Monte Carlo.

Figure 1 displays contours of the three structured components when  $n = 2$  and  $p = 2$ . Clearly the components distribute mass quite differently from one another, and in a way that reflects constraints encoded by  $\eta$ . Note that the densities from different structures  $\eta$  have the same support; the constraints restrict latent expected values rather than observables. In this way the approach shares something with generalized linear modeling wherein responses are modeled by generic exponential family densities and covariate information constrains the expected values (McCullagh and Nelder, 1989).

**4. Linear independence.** The component densities (7) seem to have the useful property of being linearly independent functions on  $\mathbb{R}^n$ . Linear independence of the component density functions is equivalent to identifiability of the mixture model (Yakowitz and Spragins, 1968). It is necessary for strict concavity of the log-likelihood, but it is not routinely established.

FIG 1. Three structured components in  $\mathbb{R}^2$ . Here  $\alpha = 10$  and  $\alpha_0 = 3$ . Contours cover 50%, 80%, 95%, and 99% probability. For convenience the density is shown for log-transformed pairs.



Let  $a = (a_\eta)$  denote a vector of real numbers indexed by structures  $\eta$ . Recall that the finite catalog of functions  $\{p(x|\eta)\}$  is linearly independent if

$$T_a(x) = \sum_{\eta} a_{\eta} p(x|\eta) = 0 \quad \text{for all } x \text{ implies } a_{\eta} = 0 \text{ for all } \eta.$$

It is plausible that this property holds generally, but we have been able to establish a rigorous proof only in the following special case.

**THEOREM 3.** *In a balanced experiment where  $m$  replicate samples are measured in each of  $p = 2$  or  $p = 3$  groups, the component densities  $p(x_g|\eta)$  in (7) are linearly independent functions on  $\mathbb{R}^{mp}$ .*

A proof proceeds by finding a multivariate polynomial  $\phi(x) > 0$  such that  $\phi(x)T_a(x)$  is itself a multivariate polynomial. A close study of the degrees and coefficients of this polynomial leads us to the result (Appendix C). That such a  $\phi(x)$  exists follows from (7): the center is a rational function, and the factor  $P_{\text{ord}}(\eta)$  is also rational, being a linear combination of rational functions, as established in (4).

## 5. Numerical considerations.

5.1. *Estimation.* Parameters  $\alpha, \alpha_0, \nu_0$  in (7) reflect genome-wide properties of expression, and our current implementation estimates them separately from the relatively numerous set of mixing proportions. These mixing proportions  $\pi = \{\pi_\eta\}$  may be estimated by attempting to maximize the log

likelihood

$$(9) \quad l(\pi) = \sum_{g=1}^G \log \left\{ \sum_{\eta} \pi_{\eta} p(x_g|\eta) \right\}$$

where  $G$  is the number of genes with data, and where  $p(x_g|\eta)$  is as in (7). The approach is buttressed by the following finding.

**THEOREM 4.** *Suppose that the component densities are linearly independent functions in the mixture of structured components model. If  $G$  is sufficiently large, then the log likelihood  $l(\pi)$  in (9) is strictly concave on a convex domain, and thus admits a unique maximizer  $\hat{\pi} = \{\hat{\pi}_{\eta}\}$ . This property is almost sure in data sets.*

The expectation-maximization (EM) algorithm naturally applies to approximate  $\hat{\pi}$ . By strict concavity of  $l(\pi)$ , it is not necessary to re-run EM from multiple starting points. The final estimate and resultant clustering should be insensitive to starting position, as has been found in numerical experiments. This is a convenient but unusual property in the domain of mixture-based clustering (McLachlan and Peel, 2000, pg 44).

In a small simulation experiment we confirmed that our implementation of the EM algorithm was able to recover mixture proportions given sufficiently many draws from the marginal distribution (6) (data not shown).

**5.2. Example.** Edwards *et al.* (2003) studied the transcriptional response of mouse heart tissue to oxidative stress. Three biological replicate samples were measured using Affymetrix oligonucleotide arrays at each of five time points (baseline and one, three, five, and seven hours after a stress treatment) for several ages of mice. Considering the older mice for illustration, we have  $p = 5$  distinct groups,  $n = 15$  samples, and 10,043 genes (*i.e.*, probe sets, after pre-processing). Gene-specific moderated F-testing (Smyth, 2004) produced a list of  $G = 786$  genes that exhibited a significant temporal response to stress at the 10% false discovery rate (by q-value; Storey and Tibshirani, 2003). Gamma ranking involved fitting the mixture of structured components, which with  $p = 5$  mixes over 540 distinct components. (Since we worked with significantly altered genes, we did not include the null component in which all means are equal; other aspects of model fitting are provided in Appendix E). From the catalog of 540 possibilities, genes populated 23 clusters by gamma ranking, though only four clusters contained 10 or more of the  $G = 786$  stress-responding genes (Figure 2). Most expression changes occurred between baseline and the first time point, but 30 genes

(red cluster) showed significant up-regulation at all but one time point, for example.

Gamma-ranking gave different results than `K-means` or `mclust`, which respectively found 2 and 20 clusters in Edwards' data. ( $K$  was chosen according to guidelines in Hastie *et al.* 2001). The adjusted Rand index (Hubert and Arabie, 1985), which measures dissimilarity of partitions, was 0.09 comparing gamma-ranking and `K-means`, 0.16 for gamma-ranking and `mclust` and 0.16, while for `K-means` and `mclust` it was smaller, at 0.02. The biological significance of clusters identified by any algorithm may be worth investigating. For example, the cluster of 30 increasing expressors includes 2 genes (`Mgst1` & `Gsta4`) from among only 17 in the whole genome that are involved in glutathione transferase activity. Understanding the increased activity of this molecular function will give a more complete picture of the biology (e.g., Girardot *et al.* 2004). In isolation it is difficult to see how such investigation is supportive of a given clustering approach. The benefits become more apparent when we look at many data sets and many functional categories.

**5.3. Empirical Study.** Gamma ranking was applied to a series of 11 data sets obtained from the Gene Expression Omnibus (GEO) repository (Edgar *et al.*, 2002). These were all the data sets satisfying a specific and relevant query (Table 2). They represent experiments on different organisms and they exhibit a range of variation characteristics. In each case we applied the moderated F test and selected genes with q-value no larger than 5%. Gamma-ranking, and, for comparison, `mclust` and `K-means`, were applied in order to cluster genes separately for each data set. Basic facts about the identified clusters are reported in Table 2. Figure 3 shows that gamma ranking tends to produce smaller clusters than `mclust` and `K-means`, although it also has a wider size distribution; and there was a relatively low level of overlap among the three approaches.

The empirical study shows not only that gamma ranking produces substantially different clusters than popular approaches, but also that the identified clusters are significant in terms of their biological properties. Investigators often measure the biological properties of a gene cluster by identifying functional properties that seem to be over-represented in the cluster. Gene set enrichment analysis is most frequently performed by applying Fisher's exact test to each of a long list of functional categories, testing the null hypothesis that the functional category is independent of the gene cluster (e.g., Newton *et al.*, 2007). Functional categories from the Gene Ontology (GO) Consortium were used to assess the biological properties of all the clusters identified in the above calculation. Specifically, we computed for each cluster

FIG 2. Dominant patterns of differential expression in time course data from Edwards et al. (2003). Each panel summarizes data from one cluster identified by gamma ranking (the top nine clusters are shown). A digital code signifies the inferred ordering of the latent expected values (i.e.,  $\eta$ , in an alternative notation). Each gene is a single line trace; triplicate measurements were reduced by averaging and then standardized for display. Results are based on 100 cycles of EM to estimate mixing proportions followed by Bayes's rule assignment.

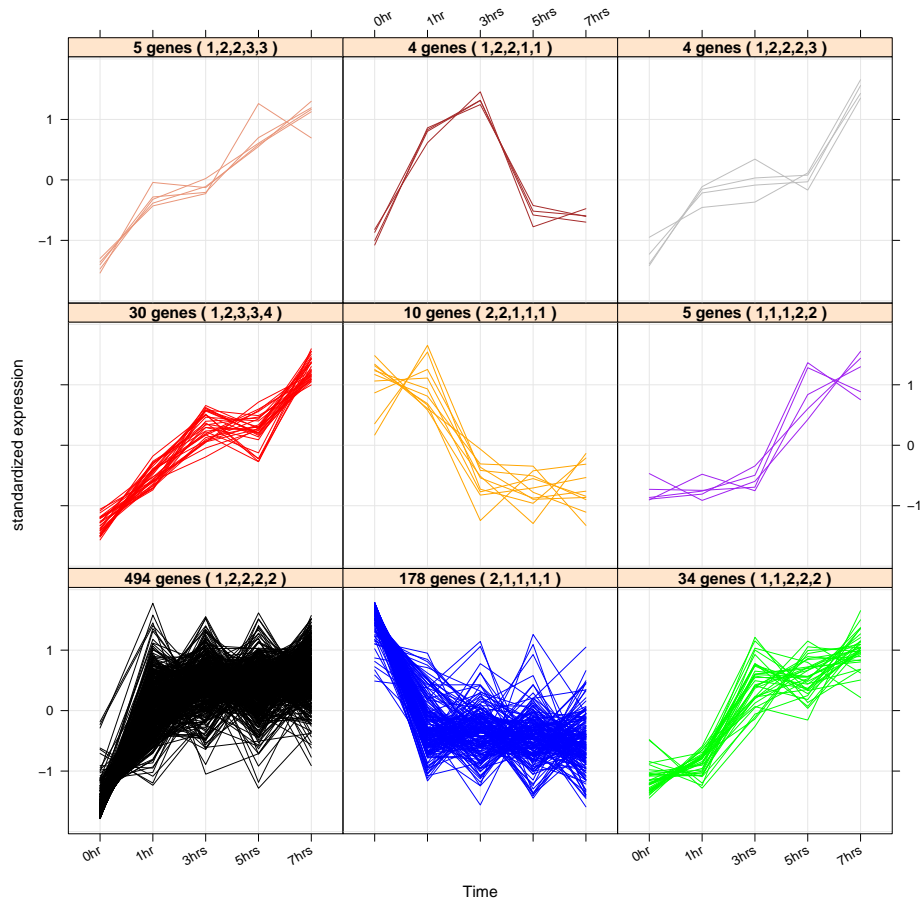
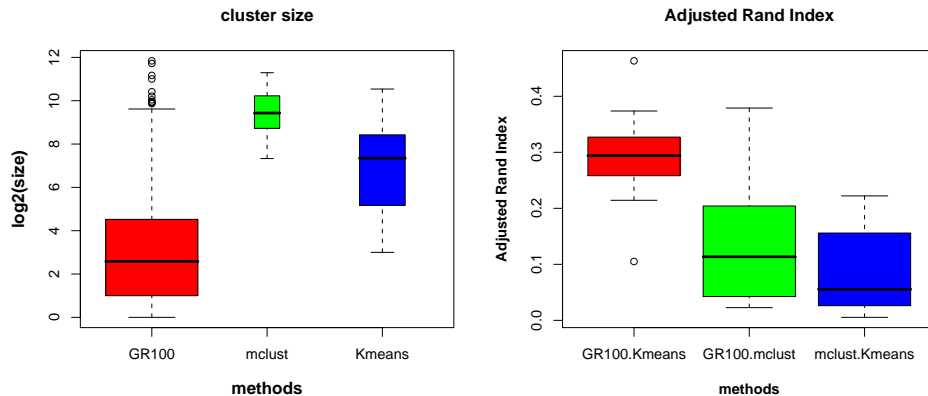


FIG 3. Characteristics of clusters from an empirical study of 11 data sets.

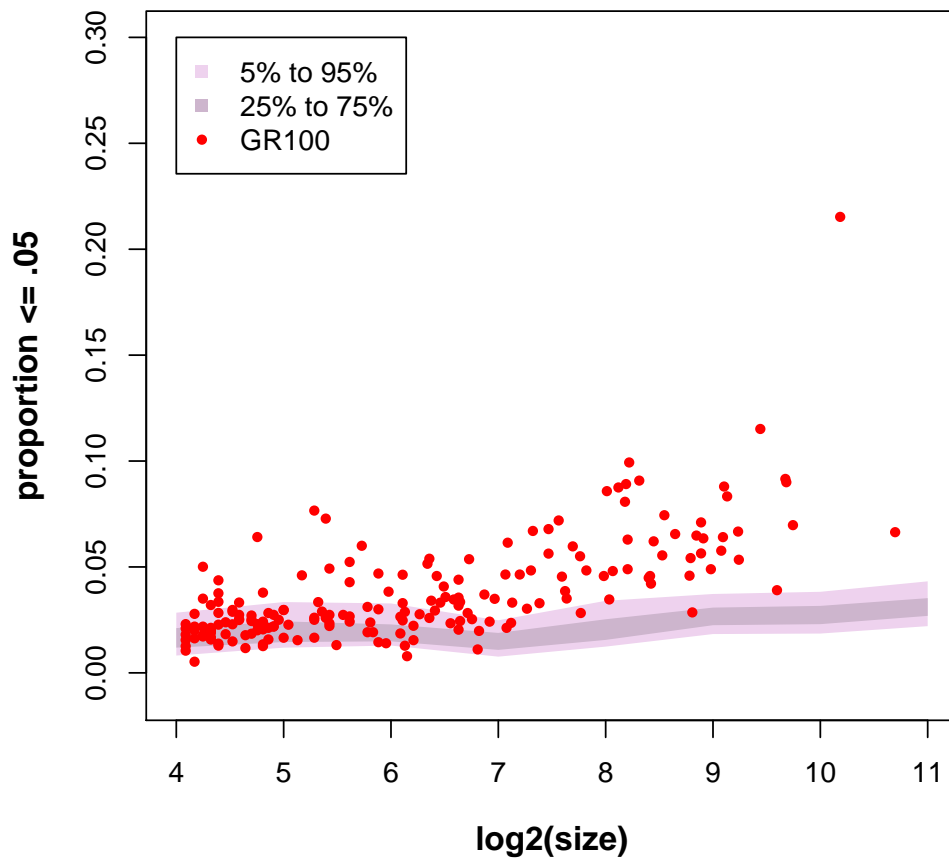


a vector of p-values across the GO data base. Figure 4 shows the proportion of these p-values smaller than 0.05, stratified by cluster size and in comparison to results on random sets of the same size. Evidently, the clusters identified by gamma-ranking contain substantial biological information.

**6. Count data.** Microarray technology naturally leads to continuous measurements of gene expression, as modeled in Section 3, but technological advances allow investigators essentially to count the number of copies of each molecule of interest in each sample (e.g., Mortazavi *et al.* 2008). Poisson distributions are central in the analysis of such data (e.g., Marioni *et al.* 2008), and gamma ranking extends readily to this case.

Briefly, data at each gene (or tag) is a vector  $x_g = (x_{g,1}, \dots, x_{g,n})$  as before, but  $x_{g,i}$  is now a count from the  $i$ th library (rather than an expression level on the  $i$ th microarray). There may be replicate libraries within a given cellular state, and comparisons of interest may be between different cellular states. Library sizes  $\{N_i\}$ , say, are additional but known design parameters. Important parameters are expected counts relative to some common library size. Adopting the notation from Section 3 and Appendix C, a cluster of libraries  $\sigma(\eta, k)$  may share their size-adjusted expected values, and so for any  $i \in \sigma(\eta, k)$  the observed count  $x_{g,i}$  arises from the Poisson distribution with mean  $N_i \mu_k$ . Further, the structure  $\eta$  on test puts an ordering constraint  $\mu_1 < \mu_2 \dots < \mu_{K_\eta}$  on these latent expectations. The key is to integrate away these latent expected values using a conjugate gamma prior, but conditionally on the ordering. Prior to conditioning, the  $\mu_k$ 's are independent and identical gamma variables with (integer) shape  $\alpha_0$  and rate  $\nu_0$ . Then, analogously to

FIG 4. Empirical study of the association between gamma-ranking clusters and biological function. For every gene cluster identified by gamma ranking in the data sets in Table 2, plotted is the proportion of small enrichment  $p$ -values (vertical) versus the set size (horizontal). The enrichment  $p$ -values are Fisher-exact-test  $p$ -values and the proportion is computed over a database of GO and KEGG pathways (Table 7). Bands indicate similar proportions computed for random sets.



Theorem 2, the predictive distribution for the vector of conditionally Poisson responses is:

$$(10) \quad p(x_g|\eta) = c_\eta \left( \prod_{i=1}^n \frac{1}{x_{g,i}!} \right) \underbrace{\left( \prod_{k=1}^{K_\eta} u_{g,k} \Gamma(s_{g,k} + \alpha_0) \right)}_{\text{center}(\eta)} P_{\text{ord}}(\eta)$$

where

$$P_{\text{ord}}(\eta) = P(Z_1 < Z_2 < \dots < Z_{K_\eta})$$

with the  $Z_k$ 's mutually independent gamma-distributed random variables with (gene-specific) shapes  $a_k = \alpha_0 + s_{g,k}$  and rates  $\lambda_k = \alpha_0\nu_0 + n_k$ . In (10), the normalizing constant is

$$c_\eta = \frac{K_\eta! (\alpha_0\nu_0)^{\alpha_0 K_\eta}}{[\Gamma(\alpha_0)]^{K_\eta}} \prod_{k=1}^{K_\eta} (\alpha_0\nu_0 + n_k)^{-\alpha_0}$$

and, further,  $s_{g,k} = \sum_{i \in \sigma(\eta,k)} x_{g,i}$ ,  $n_k = \sum_{i \in \sigma(\eta,k)} N_i$ , and

$$u_{g,k} = \prod_{i \in \sigma(\eta,k)} \left( \frac{N_i}{\alpha_0\nu_0 + n_k} \right)^{x_{g,i}}.$$

Notice that in  $P_{\text{ord}}(\eta)$  the event refers to an increasing sequence of gamma's, rather than a decreasing sequence, as in Theorem 2. This arises because for Poisson responses the conjugate prior involves a gamma distribution for the means, whereas for gamma responses the conjugate is inverse gamma on the means. For computations to work out the key thing is that some monotone transformation of each latent mean has a gamma distribution. In the null structure (all means equal),  $P_{\text{ord}}(\eta) = 1$  and (10) reduces to the negative-multinomial distribution. It will be important to study the practical utility of (10) and overdispersed extensions (*cf.* Robinson and Smyth, 2007), but such investigation is not within the scope of the present paper. The main reason to present the finding here is to show that gamma rank probabilities (Section 2) arise in multiple probability models.

**7. Concluding Remarks.** Calculations presented here consider a discrete mixture model and the resulting clustering for gene-expression or similar data types. The discrete mixing is over patterns of equality and inequality among latent expected values (*ordered structures*). In examples the method was applied after a round of feature selection, although it could have been

applied to each full data set (*i.e.*, by including the null structure in the mix) and it could have been the basis of more comprehensive analysis, going beyond clustering and towards hypothesis testing and error-rate-controlled gene lists. Our more conservative line is attributable in part to an incomplete understanding of the method's robustness. Relaxing the fixed-coefficient-of-variation assumption, as in Lo and Gottardo (2007) or Rossell (2009), could be considered to address the problem. The focus on clustering, however, is motivated largely by its practical utility in the context of genomic data analysis.

By cataloging ordered structures, rather than the smaller set of unordered structures, the mixture model produces readily interpretable clusters in the multi-group setting. Jensen *et al.* (2009) argues similarly. For example, the largest cluster of temporally responsive genes in Edwards' data are upregulated immediately after treatment and show no significant fluctuations thereafter. The development of calculations for ordered structures has been more challenging than for unordered structures, which were presented in Kendzierski *et al.* (2003) and implemented in the Bioconductor package **EBarrays** ([www.bioconductor.org](http://www.bioconductor.org)). Mixture calculations are simplified in the unordered case because component densities reduce by factorization to elementary products (*i.e.*, the last factor in (7) is not present). The requirement to compute gamma rank probabilities had limited a fuller development.

The mixture framework from Kendzierski *et al.* (2003) has supported a number of extensions to related problems in statistical genomics: Yuan and Kendzierski (2006b) (time-course data), Kendzierski *et al.* (2006) (mapping expression traits) and Keles (2007) (localizing transcription factors). The ability to monitor ordered structures may have some application in these problems. Further, the ability to compute gamma rank probabilities may have application in distinct inference problems (e.g., Doksum and Ozeki, 2008).

Plots like Figure 4, but for **K-means** and **mclust**, show a similar level of biological information in their clusters (not shown). Considering size distributions and overlap patterns, it is evident that gamma ranking provides distinct and potentially useful clusters. We also note that gamma ranking utilizes information on which samples are in which groups, while simpler clustering schemes are completely unsupervised.

TABLE 2

Summary of 11 data sets from the Gene Expression Omnibus (GEO). GDS is the GEO data set accession number. These sets satisfied the search query from August 2008 having subset variable type time or development stage or age and having a single factor with three to eight levels.  $p$  indicates the number of groups.  $G$  indicates the number of genes deemed significantly altered by one-way moderated  $F$  test and 0.05 FDR (limma). The remaining columns show how many clusters are found by gamma ranking with 100 EM iterations (GR100), and mclust and K-means.

<i>GDS</i>	<i>citation</i>	<i>organism</i>	$p$	$G$	GR100	mclust	K-means
2323	Coser <i>et al.</i>	Homo Sapiens	3	1409	11	5	13
1802	Tabuchi <i>et al.</i>	Mus musculus	4	3433	49	7	10
2043	Tabuchi <i>et al.</i>	Mus musculus	4	3001	51	8	18
2360	Ron <i>et al.</i>	Mus musculus	4	8714	50	8	30
599	Vemula <i>et al.</i>	Rattus norvegicus	5	673	42	2	40
812	Zeng <i>et al.</i>	Mus musculus	5	10982	135	7	15
1937	Pilot <i>et al.</i>	Drosophila	5	7733	88	8	10
568	Welch <i>et al.</i>	Mus musculus	6	3737	134	4	25
2431	Keller <i>et al.</i>	Homo Sapiens	6	8505	137	9	12
587	Tomczak <i>et al.</i>	Mus musculus	7	860	50	2	20
586	Tomczak <i>et al.</i>	Mus musculus	8	5211	118	5	20

**Appendix A.** *Proof of Theorem 1:* Let  $g_k(z)$  denote the density of a Gamma distribution with shape  $a_k$  and rate  $\lambda_k$ . By definition

$$\begin{aligned} P(E) &= \int_0^\infty \int_{z_K}^\infty \cdots \int_{z_2}^\infty \left[ \prod_{k=1}^K g_k(z_k) \right] dz_1 \cdots dz_{K-1} dz_K \\ &= \int_0^\infty g_K(z_K) \int_{z_K}^\infty g_{K-1}(z_{K-1}) \cdots \int_{z_2}^\infty g_1(z_1) dz_1 \cdots dz_{K-1} dz_K \end{aligned}$$

where in the second line we move factors in the integrand as far as possible to the left. With this in mind we construct functions  $f_k(z)$ ,  $z \geq 0$ , recursively as  $f_0(z) = 1$  and, for  $k = 1, 2, \dots, K$ ,

$$(11) \quad f_k(z) = \int_z^\infty f_{k-1}(u) g_k(u) du,$$

and we observe that  $P(E) = f_K(0)$ . Evaluating these functions further, we see

$$\begin{aligned} f_1(z) &= \int_z^\infty g_1(z_1) dz_1 \\ &= P(Z_1 \geq z) \\ &= P(M_1 < a_1 | Z_2 = z) \\ &= \sum_{m_1=0}^{a_1-1} \text{po}(m_1; \lambda_1 z). \end{aligned}$$

Here  $M_1 = \mathbb{M}_1(0, Z_2)$  is  $\text{Poisson}(\lambda_1 z)$  distributed conditionally upon  $Z_2 = z$ , and  $\text{po}()$  indicates the Poisson probability mass function with the indicated parameter. The equivalence in the second and third lines above stems from basic relationships between objects in the underlying Poisson processes. As long as  $M_1$  is small, it means that the  $\mathbb{N}_1$  process has not accumulated many points up to time  $Z_2 = z$ , and hence the  $Z_1$  value must be relatively large. More basically,

$$(12) \quad P(U > u) = P(X < a)$$

when  $U \sim \text{Gamma}(a, \lambda)$  and  $X \sim \text{Poisson}(\lambda u)$ , for integer shapes  $a$ .

Proceeding to  $f_2(z)$ ,

$$\begin{aligned} f_2(z) &= \int_z^\infty f_1(z_2) g_2(z_2) dz_2 \\ &= \sum_{m_1=0}^{a_1-1} \int_z^\infty \text{po}(m_1; \lambda_1 z_2) g_2(z_2) dz_2 \\ &= \sum_{m_1=0}^{a_1-1} p_1(m_1) \int_z^\infty \frac{\text{po}(m_1; \lambda_1 z_2); , g_2(z_2)}{p_1(m_1)} dz_2 \end{aligned}$$

Here  $p_1(m_1)$  is the probability mass function of a negative-binomial distribution, as in (3). Indeed we have reorganized the summand above to highlight that integrand on the far right is precisely the density function of a Gamma distributed variable with shape  $a_2 + m_1$  and rate  $\lambda_1 + \lambda_2$ . This represents the Poisson-Gamma conjugacy in ordinary Bayesian analysis (e.g., Gelman *et al.* 2004, pages 52-53). The integral evaluates to 1 if  $z = 0$ , and hence we have proved the case  $K = 2$ . But furthermore the integral represents the chance that a Gamma distributed variable is large, and so by (12)

$$\begin{aligned} f_2(z) &= \sum_{m_1=0}^{a_1-1} \sum_{m_2=0}^{m_1+a_2-1} p_1(m_1) \text{po}(m_2; (\lambda_1 + \lambda_2)z) \\ &= \sum_{m_1=0}^{a_1-1} \sum_{m_2=0}^{m_1+a_2-1} p_1(m_1) \text{po}(m_2; \Lambda_2 z) \end{aligned}$$

The baseline result of an induction proof has been established. Assume that for some  $k \geq 3$ ,

$$(13) f_{k-1}(z) = \sum_{m_1=0}^{a_1-1} \cdots \sum_{m_{k-1}=0}^{m_{k-2}+a_{k-1}-1} \left( \prod_{j=1}^{k-2} p_j(m_j) \right) \text{po}(m_{k-1}; \Lambda_{k-1} z)$$

and then evaluate (11) to obtain

$$\begin{aligned} f_k(z) &= \int_z^\infty f_{k-1}(z_k) g_k(z_k) dz_k \\ &= \int_z^\infty \sum_{m_1=0}^{a_1-1} \cdots \sum_{m_{k-1}=0}^{m_{k-2}+a_{k-1}-1} \left( \prod_{j=1}^{k-2} p_j(m_j) \right) \text{po}(m_{k-1}; \Lambda_{k-1} z_k) g_k(z_k) dz_k \\ &= \sum_{m_1=0}^{a_1-1} \cdots \sum_{m_{k-1}=0}^{m_{k-2}+a_{k-1}-1} \left( \prod_{j=1}^{k-2} p_j(m_j) \right) \int_z^\infty \text{po}(m_{k-1}; \Lambda_{k-1} z_k) g_k(z_k) dz_k \\ &= \sum_{m_1=0}^{a_1-1} \cdots \sum_{m_{k-1}=0}^{m_{k-2}+a_{k-1}-1} \left( \prod_{j=1}^{k-1} p_j(m_j) \right) \int_z^\infty \frac{\text{po}(m_{k-1}; \Lambda_{k-1} z_k) g_k(z_k)}{p_{k-1}(m_{k-1})} dz_k \\ &= \sum_{m_1=0}^{a_1-1} \cdots \sum_{m_{k-1}=0}^{m_{k-2}+a_{k-1}-1} \left( \prod_{j=1}^{k-1} p_j(m_j) \right) \sum_{m_k=0}^{m_{k-1}+a_k-1} \text{po}(m_k; \Lambda_k z) \\ &= \sum_{m_1=0}^{a_1-1} \cdots \sum_{m_k=0}^{m_{k-1}+a_k-1} \left( \prod_{j=1}^{k-1} p_j(m_j) \right) \text{po}(m_k; \Lambda_k z) \end{aligned}$$

which establishes that (13) is true for *all*  $k$ . Evaluating at  $k = K$  and  $z = 0$  establishes the theorem.  $\square$

**Coda:** Further insight is gained by realizing from the definition of the counts that

$$\begin{aligned} \mathbb{M}_k(Z_k) &= \mathbb{M}_{k-1}(Z_k) + a_k \\ &= M_{k-1} + a_k. \end{aligned}$$

But also  $\mathbb{M}_k$  has a jump at  $Z_k$ , and so we see the equivalence

$$(14) \quad Z_k > Z_{k+1} \iff M_k < M_{k-1} + a_k.$$

The event  $E$  is an intersection of these pairwise events, and this is manifested in the ranges of summation in (4). In contrast to (4), these event considerations give  $P(E)$  equal to

$$(15) \quad \sum_{m_1=0}^{a_1-1} \sum_{m_2=0}^{m_1+a_2-1} \cdots \sum_{m_{K-1}=0}^{m_{K-2}+a_{K-1}-1} p_{\text{joint}}(m_1, m_2, \dots, m_{K-1}).$$

The implication seems to be that  $M_1, M_2, \dots, M_{K-1}$  are mutually independent, though Theorem 1 does not confirm this because the factorization into negative binomials is required for all arguments, beyond what is shown. It is a conjecture that the  $\{M_k\}$  are mutually independent. A proof by brute force evaluation in the special cases  $K = 3$  and  $K = 4$  is available (not shown), but we have not found a general proof. The finding is somewhat surprising because the  $\{\mathbb{M}_k\}$  processes are highly positively dependent. The independence seems to emerge as a balance between this positive dependence and the negative association created by  $Z_k$  being inversely related to  $\mathbb{M}_k(0, t]$ .  $\square$

**Appendix B.** *Details of the structured component  $p(x_g|\eta)$ , and proof of Theorem 2.* Reviewing notation, samples are indexed by  $i = 1, 2, \dots, n$ , a given structure  $\eta$  induces a partition of the samples into  $K_\eta$  distinct subsets  $\sigma(\eta, k)$ , for  $k = 1, 2, \dots, K_\eta$ , with the ordering by  $k$  associated with increases in the latent expected value associated with each subset. Data at one gene are  $x_g = (x_{g,1}, \dots, x_{g,n})$ . Sum statistics  $s_{g,k}$  and set sizes  $n_k$  are as above (7); also introduce product statistics  $t_{g,k} = \prod_{i \in \sigma(\eta,k)} x_{g,i}$ .

The latent expected values,  $\{\mu_k\}$ , are constrained by  $\eta$  to the order  $\mu_1 < \mu_2 \dots < \mu_{K_\eta}$ . Equivalently the inverses  $\psi_k = 1/\mu_k$  are in descending order. The conjugate pairing involves the a joint density on the latent means

$$(16) \quad p_\eta(\psi_1, \dots, \psi_{K_\eta}) = K_\eta! \left[ \prod_{k=1}^{K_\eta} \frac{(\alpha_0 \nu_0)^{\alpha_0} \psi_k^{\alpha_0-1} \exp\{-\alpha_0 \nu_0 \psi_k\}}{\Gamma(\alpha_0)} \right]$$

$$(17) \quad \times 1 [\psi_1 > \psi_2 \dots > \psi_{K_\eta}]$$

which reflects independent and identically distributed Gamma( $\alpha_0, \alpha_0 \nu_0$ ) components, but constrained to one ordering, and the observation model

$$(18) \quad p(x_g|\psi_1, \dots, \psi_{K_\eta}, \eta) = \prod_{k=1}^{K_\eta} \prod_{i \in \sigma(\eta,k)} \frac{(\alpha \psi_k)^\alpha x_{g,i}^{\alpha-1} \exp\{-x_{g,i} \psi_k \alpha\}}{\Gamma(\alpha)}$$

$$= \prod_{k=1}^{K_\eta} \frac{(\alpha \psi_k)^{\alpha n_k} t_{g,k}^{\alpha-1} \exp\{-s_{g,k} \psi_k \alpha\}}{(\Gamma(\alpha))^{n_k}}.$$

Equivalently, with sample  $i \in \sigma(\eta, k)$ , measurement  $x_{g,i}$  is distributed Gamma( $\alpha, \alpha \psi_k$ ), all conditionally on the latent values and  $\eta$ , and independently across samples. The structured component  $p(x_g|\eta)$  arises by integrating (18) against the continuous mixing distribution (16). Specifically,

$$p(x_g|\eta) = \int p(x_g|\psi_1, \dots, \psi_{K_\eta}, \eta) p_\eta(\psi_1, \dots, \psi_{K_\eta}) d\psi_1 \dots d\psi_{K_\eta}.$$

Moving allowable factors from the integral,

$$p(x_g|\eta) = \frac{K_\eta! (\alpha_0 \nu_0)^{K_\eta \alpha_0} \alpha^{\alpha n}}{\Gamma(\alpha_0)^{K_\eta} \Gamma(\alpha)^n} \left( \prod_{k=1}^{K_\eta} J_k t_{g,k}^{\alpha-1} \right)$$

$$\times \int \prod_{k=1}^{K_\eta} \frac{\psi_k^{\alpha_0 + \alpha n_k - 1} \exp\{-\psi_k (\alpha_0 \nu_0 + \alpha s_{g,k})\}}{J_k} d\psi_1 \dots d\psi_{K_\eta}$$

where the integral is over the ordered space of  $\psi_k$ 's, and where  $J_k$  represents any cluster-specific quantity which does not depend on  $\psi_k$ . Choosing

$$J_k = \frac{\Gamma(\alpha_0 + \alpha n_k)}{(\alpha_0 \nu_0 + \alpha s_{g,k})^{\alpha_0 + \alpha n_k}}$$

provides just the right normalization, because then the integrand becomes the joint density of independent Gamma-distributed variables, with variable  $k$  having shape  $a_k = \alpha_0 + \alpha n_k$  and rate  $\lambda_k = \alpha_0 \nu_0 + \alpha s_{g,k}$ . Result (7) follows immediately after further simplification.  $\square$

**Appendix C.** *Linear independence and proof of Theorem 3.* Consider the three-dimensional case, and initially consider a single replicate in each of the three groups. Data on each gene form the vector  $(x, y, z)$ , say, of three positive reals. Thirteen component densities  $p(x, y, z|\eta)$  constitute the mixture model (Table 3). For a vector  $a = (a_\eta)$  of reals, the test function is  $T_a(x, y, z) = \sum_\eta a_\eta p(x, y, z|\eta)$ . It needs to be shown that if  $T_a(x, y, z) = 0$  for all  $x, y, z > 0$ , then  $a_\eta = 0$  for all structures  $\eta$ . Specializing (7) to this case, and eliminating the positive factor  $(xyz)^{\alpha-1}$ , we see that  $T_a(x, y, z) = 0$  is equivalent to

$$(19) \quad \sum_{\eta} a_{\eta} c_{\eta} \text{center}(\eta) P_{\text{ord}}(\eta) = 0.$$

TABLE 3

*Thirteen structured components  $p(x, y, z|\eta) = c_{\eta}(xyz)^{\alpha-1} \text{center}(\eta) P_{\text{ord}}(\eta)$  in the three-dimensional, no-replicate case. The forms have been simplified, WLOG, by taking the scale  $\nu_0 = 1$ , by writing  $\beta = \alpha_0 + \alpha$  and  $\xi = \alpha_0/\alpha$ . Normalizing constants  $c_{\eta}$  are as in (8). Note that the  $e_m$  and  $e_{m,n}$  stand for constants (not involving  $x, y, z$ ), but possibly differing among rows.*

structure $\eta$	$[\text{center}(\eta)]^{-1}$	$P_{\text{ord}}(\eta)$
(123)	$(x + y + z + \xi)^{\beta+2\alpha}$	1
(12)(3)	$(x + y + \xi)^{\beta+\alpha} (z + \xi)^{\beta}$	$\sum_{m=0}^{\beta+\alpha-1} \frac{e_m (z+\xi)^{\beta} (x+y+\xi)^m}{(x+y+z+2\xi)^{\beta+m}}$
(3)(12)	“	$\sum_{m=0}^{\beta-1} \frac{e_m (z+\xi)^m (x+y+\xi)^{\beta+\alpha}}{(x+y+z+2\xi)^{\beta+\alpha+m}}$
(13)(2)	$(x + z + \xi)^{\beta+\alpha} (y + \xi)^{\beta}$	$\sum_{m=0}^{\beta+\alpha-1} \frac{e_m (y+\xi)^{\beta} (x+z+\xi)^m}{(x+y+z+2\xi)^{\beta+m}}$
(2)(13)	“	$\sum_{m=0}^{\beta-1} \frac{e_m (y+\xi)^m (x+z+\xi)^{\beta+\alpha}}{(x+y+z+2\xi)^{\beta+\alpha+m}}$
(1)(23)	$(y + z + \xi)^{\beta+\alpha} (x + \xi)^{\beta}$	$\sum_{m=0}^{\beta-1} \frac{e_m (x+\xi)^m (y+z+\xi)^{\beta+\alpha}}{(x+y+z+2\xi)^{\beta+\alpha+m}}$
(23)(1)	“	$\sum_{m=0}^{\beta+\alpha-1} \frac{e_m (x+\xi)^{\beta} (y+z+\xi)^m}{(x+y+z+2\xi)^{\beta+m}}$
(1)(2)(3)	$[(x + \xi)(y + \xi)(z + \xi)]^{\beta}$	$\sum_{m=0}^{\beta-1} \sum_{n=0}^{m+\beta-1} \frac{e_{m,n} (x+\xi)^m [(y+\xi)(z+\xi)]^{\beta} (x+y+2\xi)^n}{(x+y+2\xi)^{\beta+m} (x+y+z+3\xi)^{\beta+n}}$
(2)(1)(3)	“	$\sum_{m=0}^{\beta-1} \sum_{n=0}^{m+\beta-1} \frac{e_{m,n} (y+\xi)^m [(x+\xi)(z+\xi)]^{\beta} (x+y+2\xi)^n}{(x+y+2\xi)^{\beta+m} (x+y+z+3\xi)^{\beta+n}}$
(1)(3)(2)	“	$\sum_{m=0}^{\beta-1} \sum_{n=0}^{m+\beta-1} \frac{e_{m,n} (x+\xi)^m [(y+\xi)(z+\xi)]^{\beta} (x+z+2\xi)^n}{(x+z+2\xi)^{\beta+m} (x+y+z+3\xi)^{\beta+n}}$
(2)(3)(1)	“	$\sum_{m=0}^{\beta-1} \sum_{n=0}^{m+\beta-1} \frac{e_{m,n} (y+\xi)^m [(z+\xi)(x+\xi)]^{\beta} (y+z+2\xi)^n}{(y+z+2\xi)^{\beta+m} (x+y+z+3\xi)^{\beta+n}}$
(3)(1)(2)	“	$\sum_{m=0}^{\beta-1} \sum_{n=0}^{m+\beta-1} \frac{e_{m,n} (z+\xi)^m [(x+\xi)(y+\xi)]^{\beta} (x+z+2\xi)^n}{(x+z+2\xi)^{\beta+m} (x+y+z+3\xi)^{\beta+n}}$
(3)(2)(1)	“	$\sum_{m=0}^{\beta-1} \sum_{n=0}^{m+\beta-1} \frac{e_{m,n} (z+\xi)^m [(x+\xi)(y+\xi)]^{\beta} (y+z+2\xi)^n}{(y+z+2\xi)^{\beta+m} (x+y+z+3\xi)^{\beta+n}}$

A strictly positive multivariate polynomial  $\phi(x, y, z)$  is required that can convert the left-hand-side of (19) into a polynomial by the canceling of denominator factors. Specifically,  $\phi = \phi_1 \phi_2$  where  $\phi_1(x, y, z)$  controls factors in  $\text{center}(\eta)$  and  $\phi_2(x, y, z)$  controls factors in  $P_{\text{ord}}(\eta)$ . Inspection suggests taking  $\phi_1(x, y, z)$  equal to

$$(x + y + z + \xi)^{\beta+2\alpha} [(x + y + \xi)(x + z + \xi)(y + z + \xi)]^{\beta+\alpha} [(x + \xi)(y + \xi)(z + \xi)]^{\beta}$$

and  $\phi_2(x, y, z)$  equal to

$$(x + y + z + 2\xi)^{2\beta+\alpha-1} [(x + y + 2\xi)(x + z + 2\xi)(y + z + 2\xi)]^{2\beta-1} \times (x + y + z + 3\xi)^{3\beta-2}$$

Observe that the degree of  $x$  in the polynomial  $\phi = \phi_1\phi_2$  is  $13\beta + 5\alpha - 5$ . Indeed this is also the degree of  $y$  and the degree of  $z$  by symmetry. These degrees are reduced in the polynomial  $f_\eta = \phi(x, y, z) \text{center}(\eta) P_{\text{ord}}(\eta)$  by factors in the denominators of  $\text{center}(\eta)$  and  $P_{\text{ord}}(\eta)$ . For example, if  $\eta = (12)(3)$ , then

$$\begin{aligned} f_\eta &= (x + y + z + \xi)^{\beta+2\alpha} [(x + z + \xi)(y + z + \xi)]^{\beta+\alpha} [(x + \xi)(y + \xi)]^\beta \\ &\times [(x + y + 2\xi)(x + z + 2\xi)(y + z + 2\xi)]^{2\beta-1} (x + y + z + 3\xi)^{3\beta-2} \\ &\times \sum_{m=0}^{\beta+\alpha-1} e_m (z + \xi)^\beta (x + y + \xi)^m (x + y + z + 2\xi)^{\beta+\alpha-1-m}, \end{aligned}$$

which is a polynomial of degree  $11\beta + 4\alpha - 5$ , in both  $x$  and  $y$ , and of degree  $12\beta + 5\alpha - 5$  in  $z$ . A similar construction is possible for all structures; Table 4 records the degrees of  $x$ ,  $y$ , and  $z$  in all component polynomials  $f_\eta$ .

TABLE 4

*Degree of  $x, y$ , and  $z$  in the multivariate polynomials  $f_\eta = \phi(x, y, z) \text{center}(\eta) P_{\text{ord}}(\eta)$ . Recall  $\beta = \alpha_0 + \alpha$  and both  $\alpha$  and  $\alpha_0$  are positive integers.*

structure $\eta$	degree( $x$ )	degree( $y$ )	degree( $z$ )
(123)	$11\beta + 4\alpha - 4$	$11\beta + 4\alpha - 4$	$11\beta + 4\alpha - 4$
(12)(3)	$11\beta + 4\alpha - 5$	$11\beta + 4\alpha - 5$	$12\beta + 5\alpha - 5$
(3)(12)	$12\beta + 4\alpha - 5$	$12\beta + 4\alpha - 5$	$11\beta + 4\alpha - 5$
(13)(2)	$11\beta + 4\alpha - 5$	$12\beta + 5\alpha - 5$	$11\beta + 4\alpha - 5$
(2)(13)	$12\beta + 4\alpha - 5$	$11\beta + 4\alpha - 5$	$12\beta + 4\alpha - 5$
(1)(23)	$11\beta + 4\alpha - 5$	$12\beta + 4\alpha - 5$	$12\beta + 4\alpha - 5$
(23)(1)	$12\beta + 5\alpha - 5$	$11\beta + 4\alpha - 5$	$11\beta + 4\alpha - 5$
(1)(2)(3)	$10\beta + 5\alpha - 5$	$11\beta + 5\alpha - 5$	$12\beta + 5\alpha - 5$
(2)(1)(3)	$11\beta + 5\alpha - 5$	$10\beta + 5\alpha - 5$	$12\beta + 5\alpha - 5$
(1)(3)(2)	$10\beta + 5\alpha - 5$	$12\beta + 5\alpha - 5$	$11\beta + 5\alpha - 5$
(2)(3)(1)	$12\beta + 5\alpha - 5$	$10\beta + 5\alpha - 5$	$11\beta + 5\alpha - 5$
(3)(1)(2)	$11\beta + 5\alpha - 5$	$12\beta + 5\alpha - 5$	$10\beta + 5\alpha - 5$
(3)(2)(1)	$12\beta + 5\alpha - 5$	$11\beta + 5\alpha - 5$	$10\beta + 5\alpha - 5$

Having introduced the multiplier  $\phi$ , the linear independence (19) is equivalent to the assertion that polynomial equation

$$(20) \quad \sum_{\eta} a_{\eta} c_{\eta} f_{\eta}(x, y, z) = 0 \quad \text{for all } x, y, z > 0$$

implies  $a_\eta = 0$  for all  $\eta$ . Fixing any  $y$  and  $z$ , the left-hand-side of equation (20) is a polynomial in  $x$ , with degree  $12\beta + 5\alpha - 5$ , according to Table 4. Indeed terms associated with structures  $\eta = (23)(1)$ ,  $(2)(3)(1)$ , and  $(3)(2)(1)$  all contribute monomials with that highest power in  $x$ . The coefficient of  $x^{12\beta+5\alpha-5}$ , denoted  $d = d(a, y, z)$ , equals

$$a_{(23)(1)}c_{(23)(1)}f'_{(23)(1)} + a_{(2)(3)(1)}c_{(2)(3)(1)}f'_{(2)(3)(1)} + a_{(3)(2)(1)}c_{(3)(2)(1)}f'_{(3)(2)(1)}$$

where  $f'$  indicates contributions from respective terms within  $f_\eta$ . This coefficient  $d$  must equal zero, for all  $y$  and  $z$ ; after all, a degree  $12\beta + 5\alpha - 5$  polynomial can equal zero in  $x$  for at most that many  $x$  values, unless the coefficient  $d$  is exactly zero; and we are asking that it equal zero at all values of  $x$ . From this study of the high-power coefficient in  $x$ , we have reduced consideration to three structures and are able to focus on  $d = d(a, y, z)$  as a bivariate polynomial in  $y$  and  $z$  (Table 5).

TABLE 5  
Degrees of  $y$ , and  $z$  in three terms of the bivariate polynomial  $d(a, y, z)$ . This is a subset of Table 4.

structure $\eta$	degree( $y$ )	degree( $z$ )
(23)(1)	$11\beta + 4\alpha - 5$	$11\beta + 4\alpha - 5$
(2)(3)(1)	$10\beta + 5\alpha - 5$	$11\beta + 5\alpha - 5$
(3)(2)(1)	$11\beta + 5\alpha - 5$	$10\beta + 5\alpha - 5$

The initial argument focusing on the degree of  $x$  can be adapted to study other variables in Table 5. With degree of  $y$  equal to  $11\beta + 5\alpha - 5$ , for instance, it must be that the coefficient  $d'(z)$ , say, of  $y^{11\beta+5\alpha-5}$  equals zero for all  $z$ . After all, the polynomial can equal zero at at most  $11\beta + 5\alpha - 5$   $y$  values, and we require it to be zero at all  $y$ . But all contributions to that coefficient are strictly positive, except possibly  $a_{(3)(2)(1)}$ , hence we conclude  $a_{(3)(2)(1)} = 0$ . By the same token, working with the degree  $11\beta + 5\alpha - 5$  term in  $z$ , it follows that  $a_{(2)(3)(1)} = 0$ , which then forces  $a_{(23)(1)} = 0$ , because we require  $d = 0$  overall. Three rows from Table 4 have been eliminated (i.e., forced  $a_\eta = 0$ ), all those in which the mean of the first variable is greater than the other two means. Next, return to the reduced table, and focus, say, on structures  $(13)(2)$ ,  $(1)(3)(2)$ , and  $(3)(1)(2)$ , in which the second variable has mean greater than the others. In doing so, three more coefficients  $a_{(13)(2)} = a_{(1)(3)(2)} = a_{(3)(2)(1)} = 0$  are forced, and Table 2 is further reduced to seven rows. Then the argument is repeated to get  $a_{(12)(3)} = a_{(1)(2)(3)} = a_{(2)(1)(3)} = 0$ , and it remains to assess coefficients  $a_\eta$  of the four structures in Table 6.

The argument is repeated in this domain, knowing that all but four terms in (20) have been eliminated. The degree of  $x$  is  $12\beta + 4\alpha - 5$ , and there are

TABLE 6  
*Final subtable*

structure $\eta$	degree( $x$ )	degree( $y$ )	degree( $z$ )
(123)	$11\beta + 4\alpha - 4$	$11\beta + 4\alpha - 4$	$11\beta + 4\alpha - 4$
(3)(12)	$12\beta + 4\alpha - 5$	$12\beta + 4\alpha - 5$	$11\beta + 4\alpha - 5$
(2)(13)	$12\beta + 4\alpha - 5$	$11\beta + 4\alpha - 5$	$12\beta + 4\alpha - 5$
(1)(23)	$11\beta + 4\alpha - 5$	$12\beta + 4\alpha - 5$	$12\beta + 4\alpha - 5$

contributions from both  $\eta = (3)(12)$  and  $\eta = (2)(13)$ . But then restricted to these rows we get  $a_{(3)(12)} = 0$  because the coefficient of  $x^{12\beta+4\alpha-5}$  as a polynomial in  $y$  has degree  $12\beta + 4\alpha - 5$ . The remaining constants  $a_\eta$  are similarly zero, completing the proof in the no-replicate ( $m = 1$ ), three group ( $p = 3$ ) case.

The balanced three group case follows suit, noting that now  $x$ ,  $y$ , and  $z$  are sums taken, respectively, across replicates in each of the three groups. The product statistic is not  $xyz$ , but anyway it is common to all components and thus cancels in the linear combination test function. The observation-related shape parameter  $\alpha$  is replaced by  $m\alpha$ . The two-dimensional ( $p = 2$ ) case is simpler and is left as an exercise.  $\square$

**Appendix D.** *Strict concavity of log-likelihood and proof of Theorem 4.*

Let  $q$  denote the number of non-null structures, and consider the log-likelihood  $l(\pi)$  in (9) to be on  $\mathbb{R}^q$ , with the null probability defined secondarily as  $\pi_0 = 1 - \sum_{\eta \neq \eta_0} \pi_\eta$ . This way we need not invoke Lagrange multipliers to compute derivatives of  $l(\pi)$ . By calculus, the  $q \times q$  Hessian  $H$  of negative 2nd derivatives of  $l(\pi)$  has  $(ij)$ th entry

$$\begin{aligned} H_{ij} &= \sum_g \frac{[p(x_g|\eta_i) - p(x_g|\eta_0)][p(x_g|\eta_j) - p(x_g|\eta_0)]}{[p(x_g)]^2} \\ &= \sum_g f_i(x_g) f_j(x_g) \end{aligned}$$

where  $p(x_g)$  is the marginal density obtained by mixing over structures, as in (6), and  $f_i(x) = [p(x|\eta_i) - p(x|\eta_0)]/p(x)$ . Now let  $a = (a_\eta)$  be a  $q$ -vector of constants. To determine curvature of the log-likelihood we consider the quadratic form

$$\begin{aligned} a^T H a &= \sum_{i=1}^q \sum_{j=1}^q a_i a_j \sum_g f_i(x_g) f_j(x_g) \\ &= \sum_g \left( \sum_{i=1}^q a_i f_i(x_g) \right)^2 \\ &= \sum_g [T_a(x_g)]^2 \end{aligned}$$

where  $T_a(x) = \sum_{i=1}^q a_i f_i(x)$ . Clearly  $a^T H a \geq 0$  regardless of  $a$ , and so  $H$  is non-negative definite and  $l(\pi)$  is concave. To establish *strict* concavity requires that we show  $T_a(x_g) = 0$  for all  $g$  if and only if  $a = 0$ . The following lemma shows that knowing  $T_a(x_g) = 0$  for all  $G$  values  $x_g$  is enough to force  $T_a(x) = 0$  for all  $x$ , as long as  $G$  is sufficiently large. But then  $a = 0$  by the linear independence assumption, completing the proof.

**LEMMA 1.** *Let  $\psi(x)$  be a multivariate polynomial in  $x \in \mathbb{R}^n$ , and let  $X_1, X_2, \dots, X_m$  denote a random sample from a continuous distribution on  $\mathbb{R}^n$ . If  $m$  is at least as large as the number of monomials in  $\psi$ , then, with probability one,  $\psi(X_i) = 0$  for  $i = 1, 2, \dots, m$  implies  $\psi(x) = 0$  for all  $x$ .*

*Proof of Lemma 1.* Every point  $X_i$  puts a linear condition on the space of coefficients of  $\psi$ . It needs to be verified that these conditions are linearly independent. Suppose that the first  $k$  conditions are linearly independent, so the space of  $\psi$ 's that are zero at  $X_1, \dots, X_k$  has dimension (number of

monomials in  $\psi$ ) minus  $k$ . Pick one such nonzero polynomial, and call it  $\phi$ . Since  $\phi = 0$  is a set with positive codimension, we may assume (with probability one) that  $\phi(X_{k+1})$  is not zero. Then, if we impose the additional condition  $\psi(X_{k+1}) = 0$ , the dimension of the solution space drops by at least one, hence it drops by one. Letting  $k$  increase from 1 to  $m$  completes the proof.  $\square$

**Appendix E.** *Further details of numerical examples.* The parameters  $\alpha$ ,  $\alpha_0$ , and  $\nu_0$  were fixed at values obtained by first fitting the unordered gamma-gamma mixture model in EBarrays, without a null structure but otherwise allowing all possible unordered structures. Shape parameters were then rounded to the nearest positive integer (Table 7) and all three parameters were plugged into the EM procedure to fit the proposed mixture model proportions  $\pi_\eta$ . To simplify EM calculations in the four examples having more than five groups, the full set of ordered structures was filtered to a reduced set based on the fitting of the unordered gamma-gamma model in EBarrays. Each ordered structure corresponds to exactly one unordered structure (a many to one mapping). If no gene had a high (greater than 0.5) probability of mapping to a given unordered structure, then we deemed all corresponding ordered structures to have  $\pi_\eta = 0$ .

TABLE 7

*Parameter estimates (not including mixing proportions) from the examples analyzed. The last column indicates the number of functional categories in GO and KEGG having at least five annotated genes, which were used in the development of Figure 4. KEGG was not available for GDS1937, and so this data set was not used in Figure 4.*

data set	$\alpha$	$\alpha_0$	$\nu_0$	# GO/KEGG
Edwards	113	1	586.5	
GDS2323	14	1	119.1	3849/184
GDS1802	17	1	46.8	3619/182
GDS2043	22	1	47.7	3619/182
GDS2360	8	1	15.8	3258/175
GDS599	12	1	0.01	3180/159
GDS812	5	1	15.4	3258/175
GDS1937	6	1	20.5	NA
GDS568	10	1	37.1	3258/175
GDS2431	4	1	67.6	4085/188
GDS587	8	1	9999.2	1876/127
GDS586	13	1	4566.2	3258/175

**Acknowledgements.** We thank Lev Borisov for the proof of Lemma 1 and Christina Kendzierski for comments that improved the development of this work, which was supported in part by grants R01 ES017400 and T32

GM074904 from the National Institutes of Health. We thank reviewers in advance for any useful comments.

## References.

- [1] CAMPBELL, E.A., O'HARA, L., CATALANO, R.D., SHARKEY, A.M., FREEMAN, T.C., AND JOHNSON, M.H. (2006). Temporal expression profiling of the uterine luminal epithelium of the pseudo-pregnant mouse suggests receptivity to the fertilized egg is associated with complex transcriptional changes. *Human Reproduction*, **21**, 2495–2513.
- [2] DAHL, D.B. AND NEWTON, M.A. (2007). Multiple hypothesis testing by clustering treatment effects. *Journal of the American Statistical Association*, **102**, 517–526.
- [3] DENNIS, B. AND PATIL, G.P. (1984). The gamma distribution and weighted multimodal gamma distributions as models of population abundance. *Mathematical Biosciences*, **68**, 187–212.
- [4] DOKSUM, K.A. AND OZEKI, A. (2008). Semiparametric models and likelihood-the power of ranks. In, *Proceedings of the Third Erich L. Lehmann Symposium*.
- [5] EDGAR, R., DOMRACHEV, M., AND LASH, A.E. (2002). Gene expression omnibus: NCBI gene expression and hybridization array data repository, *Nucleic Acids Research*, **30**, 207–210.
- [6] EDWARDS, M., SARKAR, D., KLOPP, R., MORROW, J., WEINDRUCH, R., AND PROLLA, T. (2003). Age-related impairment of the transcriptional response to oxidative stress in the mouse heart. *Physiological Genomics*, **13**, 119–127.
- [7] EISEN, MB, SPELLMAN, PT, BROWN, PO, AND BOTSTEIN, D (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*. **95**, 14863–14868.
- [8] FRALEY, C., AND RAFTERY, A.E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* **97**, 611–631.
- [9] GELMAN, A., CARLIN, J.B., STERN, H.S., AND RUBIN, D.B. (2004). *Bayesian data analysis*. 2nd edition. Chapman and Hall.
- [10] GIRARDOT, F., MONNIER, V., TRICOIRE, H. (2004). Genome wide analysis of common and specific stress responses in adult drosophila melanogaster. *BMC Genomics*, **5**, 74.
- [11] GRASSO, L.C., MAINDONALD, J., RUDD, S., HAYWARD, D.C., SAINT, R., MILLER, D.J., AND BALL, E.E. (2008). Microarray analysis identifies candidate genes for key roles in coral development. *BMC Genomics*, **9**, 540.
- [12] HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. (2001). *The elements of statistical learning*. Springer-Verlag, New York.
- [13] HUBERT, L. AND ARABIE, P. (1985). Comparing partitions. *Journal of Classification*, **2**, 193–218.
- [14] GREENWOOD, M. AND YULE, G.U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society*, **83**, 255–279.
- [15] HOLZMANN, H., MUNK, A., AND GNEITING, M. (2006). Identifiability of finite mixtures of elliptical distributions. *Scandinavian Journal of Statistics*, **33**, 753–763.
- [16] HUTCHINSON, T.P. (1981). Compound gamma bivariate distributions. *Metrika*, **28**, 263–271.
- [17] JENSEN, S.T., ERKAN, I., ARNARDOTTIR, E.S., AND SMALL, D.S. (2009). Bayesian testing of many hypotheses  $\times$  many genes: a study of sleep apnea. *Annals of Applied Statistics*, to appear.
- [18] KELES, S. (2007). Mixture modeling for genome-wide localization of transcription factors. *Biometrics*, **63**, 10–21.

- [19] KENDZIORSKI, C.M., NEWTON, M.A., LAN, H., AND GOULD, M.N. (2003). On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. *Statistics in Medicine*, **22**, 3899–3914.
- [20] KENDZIORSKI, C.M., CHEN, M., YUAN, M. LAN, H., AND ATTIE, A.D. (2006). Statistical methods for expression quantitative trait loci (eQTL) mapping. *Biometrics*, **62**, 19–27.
- [21] KSCHISCHANG, R., FREY, B.J., AND LOELIGER, H.A. (2001). Factor graphs and the sum-product algorithm. *IEEE Transactions on information theory*, **47**, 498–519.
- [22] LO, K. AND GOTTARDO, R (2007). Flexible empirical Bayes models for differential gene expression. *Bioinformatics*, **23**, 328–335.
- [23] MARIONI, J.C., MASON, C.E., MANE, S.M., STEPHENS, M., AND GILAD, Y. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, **18**, 1509–1517.
- [24] MCCULLAGH, P. AND NELDER, J.A. (1989). *Generalized linear models*, 2nd edition. Chapman and Hall.
- [25] MCLACHLAN, G.J. AND BASFORD, K.E. 1988. *Mixture models: inference and applications to clustering*. Marcel Dekker, New York.
- [26] MCLACHLAN, G.J. AND PEEL, D. (2000). *Finite mixture models*. Wiley, New York.
- [27] MEDVEDOVIC, M., YEUNG, K.Y., AND BUMGARNER, R.E. (2004). Bayesian mixture model based clustering of replicated microarray data. *Bioinformatics*, **20**, 1222–1232.
- [28] MORTAZAVI, A., WILLIAMS, B.A., MCCUE, K., SCHAEFFER, L. AND WOLD, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, **5**, 621–628.
- [29] NEWTON, M.A., NOUEIRY, A., SARKAR, D., AND AHLQUIST P. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Bio-statistics*, **5**, 155–176.
- [30] NEWTON, M.A., QUINTANA, F.A., DEN BOON, J.A., SENGUPTA, S., AND AHLQUIST, P. (2007). Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis. *Annals of Applied Statistics*, **1**, 85–106.
- [31] PARMIGIANI, G., GARETT, E.S., IRIZARRY, R.A., ZEGER, S.L. (EDS) (2003). *The analysis of gene expression data: methods and software* Springer.
- [32] RABINER, L.R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**, 257–286.
- [33] REDNER, R.A. AND WALKER, H.F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* **26**, 195–239.
- [34] REMPALA, G.A. AND PAWLIKOWSKA, I. (2008). Limit theorems for hybridization reactions on oligonucleotide microarrays. *Journal of Multivariate Analysis*, **99**, 2082–2095.
- [35] ROBINSON, M.D. AND SMYTH, G.K. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, **23**, 2881–2887.
- [36] ROSSELL, D. (2009). GaGa: a parsimonious and flexible model for high-throughput data analysis *Annals of Applied Statistics*, **3**, in press.
- [37] SMYTH, G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, **3**, no 1, article 3.
- [38] SOBEL, M. AND FRANKOWSKI, K. (1994). The 500th anniversary of the sharing problem (the oldest problem in the theory of probability). *American Mathematical Monthly*, **101**, 833–847.
- [39] SPEED, T. (ED) (2004). *Statistical analysis of gene expression microarray data*. Chapman and Hall/CRC.

- [40] STEPHENS, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society, Series B*, **62**, 795–809.
- [41] STERN, H. (1990). Models for distributions on permutations. *Journal of the American Statistical Association*, **85**, 558–564.
- [42] STOREY, J.D. AND TIBSHIRANI, R. (2003). Statistical significance for genome-wide studies. *Proceedings of the National Academy of Sciences*, **100**, 9440–9445.
- [43] THALAMUTHU, A., MUKHAPADHYAY, I., ZHENG, X., AND TSENG, G.C. (2006). Evaluation and comparison of gene clustering methods in microarray analysis. *Bioinformatics*, **22**, 2405–2412.
- [44] TITTERINGTON, D.M., SMITH, A.F.M., AND MAKOV, U.E. (1985). *Statistical analysis of finite mixture distributions*. Wiley, New York.
- [45] YAKOWITZ, S.J. AND SPRAGINS, J.D. (1968). On the identifiability of finite mixtures. *Annals of Mathematical Statistics*, **39**, 209–214.
- [46] YUAN, M. AND KENDZIORSKI, C.M. (2006a). A unified approach for simultaneous gene clustering and differential expression identification. *Biometrics*, **62**, 1089–1098.
- [47] YUAN, M. AND KENDZIORSKI, C.M. (2006b). Hidden Markov models for time course data in multiple biological conditions (with discussion). *Journal of the American Statistical Association*, **101**, 1323–1340.

E-MAIL: newton@stat.wisc.edu

E-MAIL: lchung@stat.wisc.edu