Appendix I: Assigning Significance to Global Difference

Recall that our goal is to solve the following maximization problem:

$$max_{Y_C}(p(Y_C|Y_1,\sigma^2)) \qquad \text{such that} \qquad D(C,C_1)=e^2\ . \tag{1}$$

To show how to solve this problem we introduce additional notation. Recall that the expression curves consist of sets of cubic splines. As mentioned in $Methods$, splines can be fully specified by a set of control points. Let F_1 be the set of control points for C_1, and F_C be the set of control points for a curve C. Because F_1 and F_C fully specify C_1 and C, we can use the notation $D(F_C,F_1)$ instead of $D(C,C_1)$. In addition, based on our model we can write $Y_1 = SF_1$ [*], and $Y_C = SF_C$. Since individual noise terms are normally distributed, it can be shown (by taking the log and ignoring constant terms) that maximizing $p(Y_C|Y_1,\sigma^2)$ is equivalent to minimizing $(Y_1-Y_C)^T(Y_1-Y_C)$. Thus, the above maximization problem can be written as the following quadratic minimization problem:

$$min_{F_C}(S(F_1-F_C))^T(S(F_1-F_C)) \qquad \text{such that} \qquad D(F_C,F_1)=e^2 \qquad . \tag{2}$$

This problem can be rewritten as a quadratic minimization problem, as shown in the following lemma.

Lemma 0.1 *The minimization problem*

$$min_{F_C}(S(F_1-F_C))^T(S(F_1-F_C)) \qquad \textit{such that} \qquad D(F_C,F_1)=e^2$$

can be written as

$$min_{\delta}(S\delta)^T(S\delta) \qquad \textit{such that} \qquad \delta^T A\delta = 1\ ,$$

where A is a positive definite matrix.

Proof: First, we need to explicitly represent $D(C,C_1)$ using F_C and F_1. This could be done as follows

$$D(C,C_1)=D(F_C,F_1)=\frac{\int_{v_s}^{v_e}(s(t)[F_1-F_C])^2dt}{V}\ ,$$

where $s(t)$ is the spline basis function evaluated at time t. Set $\delta = F_1 - F_C$, then we have $(S(F_1-F_C))^T(S(F_1-F_C)) = (S\delta)^T(S\delta)$. As for the integration, denote the number of splines (or the number of piecewise polynomials) by q, and let $p_0 \ldots p_q$ be the set of start and end points for the individual splines (that is, the first spline starts at $p_0 = v_s$ and ends at p_1 and so on). Thus, the integration part can be rewritten

$$\int_{v_s}^{v_e}(s(t)[F_1-F_C])^2dt = \sum_{i=0}^{q-1}\int_{p_i}^{p_{i+1}}(s(t)\delta)^2dt = \sum_{i=0}^{q-1}\delta^T(\int_{p_i}^{p_{i+1}}(s(t)s(t)^T)dt)\delta\ .$$

Because $s(t)$ is continuous polynomial between p_i and p_{i+1}, we can evaluate the integral in the above equation. Note that this integral is actually a matrix, and as just mentioned, each entry in that matrix can be evaluated. Set $B_i = (\int_{p_i}^{p_{i+1}}(s(t)s(t)^T)dt$ and let $B=\sum_i B_i$, then $\int_{v_s}^{v_e}(s(t)\delta)^2dt = \delta^T B\delta$. Note that B is positive semidefinite, since the integral on the right hand side measures the squared distance between two splines. To show that B is positive definite, we note that if $\delta \neq 0$, then the two sets of control points differ in at least one entry. Using a B-spline basis function, each polynomial is only supported by four control

[*] Note that we omit the noise term from this equation since we are now using the predicted reference splines.

points. Since the basis for a cubic polynomial is of degree four, two identical cubic polynomials cannot be represented by two different sets of four control points. Thus, at least in the segment corresponding to this polynomial, the two curves differ, and since the integral measures the squared distance between the two curves, the result will be > 0. Setting $A = B/(V + e^2)$ proves the lemma because both V and e^2 are positive values. ■

Next, we use Lagrange multipliers to show that the solution is a vector δ such that δ is the eigenvector with the smallest eigenvalue for the matrix $A^{-1}S^TS$, as we prove in the following lemma.

Lemma 0.2 *Let δ be the eigenvector with the smallest eigenvalue for the matrix $A^{-1}S^TS$, then δ (appropriately scaled), is the solution to the following minimization problem:*

$$min_\delta (S\delta)^T(S\delta) \qquad \textit{such that} \qquad \delta^T A\delta = 1\,. \tag{3}$$

Proof: Using Lagrange multipliers we can write $L = (S\delta)^T(S\delta) - \lambda\delta^T A\delta - \lambda$. Taking the derivative with respect to δ and setting $\frac{\partial L}{\partial \delta} = 0$ we get: $S^TS\delta = \lambda A\delta \Rightarrow A^{-1}S^TS\delta = \lambda\delta$. Thus, δ is an eigenvector of $A^{-1}S^TS$. Multiplying both sides of the above equation by $\delta^T A$ we get: $\delta^T S^TS\delta = \lambda\delta^T A\delta = \lambda$, where the last equality results from our constraint ($\delta^T A\delta = 1$). Because $\delta^T S^TS\delta$ is the quantity we wish to minimize, λ must be the smallest eigenvalue of $A^{-1}S^TS$, and δ should be the eigenvector corresponding to that eigenvalue, appropriately scaled so that $\delta^T A\delta = 1$. ■

Now, set $F_\delta = F_1 - \delta$ and $Y_\delta = SF_\delta$. Based on the discussion above, we can now compute $p(C_2|C_1, H_0) = p(Y_\delta|Y_1, \sigma^2)$.

Using Y_δ we can now perform the hypothesis testing in the following way. First, as discussed in $Methods$ we set $p(C_2|C_1, H_1) = p(Y_2'|\sigma^2, H_1)$, which can be written as $p(Y_2'|Y_2', \sigma^2)$ since under H_1, C_2 represents the mean curve for the second experiment. For H_0 we have $p(C_2|C_1, H_0) = p(Y_\delta|Y_1, \sigma^2)$, and thus the log likelihood ratio evaluates to:

$$2\log\frac{p(C_2|C_1,H_1)}{p(C_2|C_1,H_0)} = 2\log\frac{e^{-\frac{(Y_2'-Y_2')^T(Y_2'-Y_2')}{2\sigma^2}}}{e^{-\frac{(Y_1-Y_\delta)^T(Y_1-Y_\delta)}{2\sigma^2}}} = \frac{(Y_1-Y_\delta)^T(Y_1-Y_\delta)}{\sigma^2}\,. \tag{4}$$

Next, to perform a significance test, we use the χ^2 distribution with q degrees of freedom (where q is the number of spline control points used by the curves).

Appendix II: The Symmetric Version of our Algorithm

So far we have assumed a fixed referenced curve. That is, when we compute the area between the reference and test curves we do not consider the fact that the reference curve might be a noisy realization of the true underlying curve. Under the null hypothesis we assume that both data sets were generated from the same underlying profile, and thus a symmetric version of our algorithm seems more appropriate for this hypothesis. As we show in this section, the algorithm presented in $Methods$ can also be described as a symmetric test on both the reference and test curve (with appropriately scaled P values), and thus our algorithm is suitable for the null hypothesis as well.

Instead of relying on the reference curve, we assume that both C_1 and C_2 are realizations of the same underlying curve C. We reformulate the comparison in terms of joint probabilities over the two curves (making the comparison symmetric) but rely on the distance between the curves rather than the points

directly. As in $Methods$, let $e^2 = D(C_1, C_2)$. For the null hypothesis we solve the following maximization problem,

$$max_{C,C',C''} P(Y_{C'}, Y_{C''}|Y_C, \sigma^2) \qquad \text{such that} \qquad D(C', C'') = e^2 \ ,$$

where C' and C'' are new (arbitrary) versions of the reference and test curves and C denotes the common underlying true curve. Due to the Gaussian noise we assume for individual measurements, the maximization over C yields $Y_C = (Y_{C'} + Y_{C''})/2$; therefore, ignoring constant terms, we have

$$\log P(Y_{C'}, Y_{C''}|Y_C, \sigma^2) = -(1/2\sigma^2)(||Y_{C'} - Y_C||^2 + ||Y_{C''} - Y_C||^2) = -(1/4\sigma^2)||Y_{C'} - Y_{C''}||^2 \ .$$

If we now set $Y_\delta = Y_{C'} - Y_{C''} = SF' - SF'' = S\delta$, then we end up solving the same optimization problem as before:

$$min_\delta (S\delta)^T (S\delta) \qquad \text{such that} \qquad \delta^T A \delta = 1 \ .$$

The only difference between the result obtained by this method (symmetric) and the result discussed in $Methods$ (asymmetric) is that the value of the likelihood ratio test using the symmetric test is half the value obtained using the asymmetric method. Thus, for every P value cutoff used by the asymmetric method, there is a corresponding P value for the symmetric method which yields the same results (i.e. the same set of genes are determined to be significantly changing). Because our P value is tuned using synthetic data, changing from the asymmetric to the symmetric method would not have changed the results presented in this paper. Since the asymmetric method is somewhat easier to explain, we focused on it $Methods$.

Appendix III: Value Specific Variance

So far we have assumed that all expression values have the same variance, σ^2. In practice, we have found that the variance of expression measurements depends on the magnitude of expression values or fold changes (see http://www.psrg.lcs.mit.edu/DiffExp/DiffExp.html). Taking this fact into account is especially important for time-series data, because small shifts in the magnitude of expression values can result in large global differences for genes with high fold change values. Our framework can be modified to use variances that depend on expression value magnitudes. Instead of maximizing $p(Y_C|Y_1, \sigma^2)$ we maximize $p(Y_C|Y_1, \sigma_1^2 \ldots \sigma_m^2)$ where $\sigma_1^2 \ldots \sigma_m^2$ are the m expression value specific variances for the samples in Y_1. Recall that the rows of S (the spline basis function matrix) correspond to the time points that were sampled in the reference experiment.

Denote by S_i the ith row of S. Let $S_i' = S_i/\sigma_i$. Then maximizing $p(Y_C|Y_1, \sigma_1^2 \ldots \sigma_m^2)$ is equivalent to minimizing $(S'(F_1 - F_C))^T (S'(F_1 - F_C))$, and we proceed by replacing S with S' in Eq. 3. This results in the differential weighting of the individual errors around Y_1, leading to a reduction in the effect that experimental artifacts and associated high variance can play in determining differential gene expression.

To compute the value-specific variance for a value x we use the following method. Let R_1 and R_2 be two repeats of the same experiment, and let θ denote a weighting coefficient (in this paper we use $\theta = 0.25$), which allows us to control the range of values that will contribute to the computation of the variance. For $r_1^i \in R_1$ let $r_2^i \in R_2$ be the corresponding repeat in the second experiment. Set

$$p(r_1^i) = \frac{1}{\sqrt{2\pi\theta^2}} e^{(r_1^i - x)^2/(2\theta^2)}$$

and $P = sum_i p(r_1^i)$. Then the value-specific variance for x, v_x is computed by setting

$$v_x = \sum_i \frac{(r_1^i - r_2^i)^2}{P} \ .$$

That is, v_x is computed by using a Gaussian bump around the selected value (x) and weighting the contribution of the different repeats based on their distance from x.

Appendix IV: Synthetic Data Results

To test the sensitivity of our algorithm, and to compare it to previous methods that work by comparing individual points, we generated four sets of samples $Y1 - Y4$ as follows (see Fig. 4). $Y1$ consisted of a set of uniform samples from a sinusoid between 0 and 4π. $Y2$ was generated by adding random noise (normally distributed with mean 0) to $Y1$. $Y3$ was generated by adding a positive value, b, to the values in $Y1$. Finally, we set $Y4 = aY1$ where $0 < a < 1$. The parameters a and b were selected such that the mean absolute error of all sets with respect to $Y1$ was equal. While $Y2$ is a noisy realization of $Y1$, $Y3$ and $Y4$ represent consistent additive or multiplicative differences. If the input set is normalized appropriately, such consistent differences over time might represent real biological change. For example, in cell cycle experiments, genes that show a reduced cycling profile are probably effected by the experimental condition and should be detected.

We repeated the process of generating $Y2 - Y4$ 1000 times for each of several sampling rates that were chosen to be similar to those used in actual expression experiments, and with various noise variances. For all the different sampling rates and noise models, our algorithm correctly identified the $Y2$ samples as a noisy realization of $Y1$. As the sampling rate increased, so did the ability of our algorithm to correctly detect consistent changes ($Y3$ and $Y4$) as differentially expressed. To compare our results with methods that work directly on the input samples, we have also performed the hypothesis testing with the actual samples by replacing Y_δ with $Y2, Y3, Y4$. In Fig. 4 we present the results of the two methods using 24 samples (similar to the sampling rates of (1)) and a noise model derived from time series repeats from Zhu $et\ al$ (2). While in all cases our algorithm correctly identified $Y2$ as a noisy realization of $Y1$, the sample based method identified 90% of the $Y2$ samples as differentially expressed. Since in time series experiments most genes do not change, such a high false-detection rate cannot be tolerated. In addition, our algorithm was able to detect more of the $Y3$ and $Y4$ curves, because the number of degrees of freedom it uses is smaller than the number used by the sample based method. Similar results where obtained with other sampling rates and different noise models.

References

[1] Spellman, P. T. and Sherlock, G. and Zhang, M.Q. and Iyer, V.R. and Anders, K. and Eisen, M.B. and Brown, P.O. and Botstein, D. and Futcher, B. (1998) *Mol. Biol. Cell* 9, 3273-3297

[2] Zhu, G. and Spellman, P. T. and Volpe, T. and Brown, P.O. and Botstein, D. and Davis, T.N. and Futcher, B. (2000) *Nature* 406, 90-94