Supporting Text

Overview of the Protocols Used in Massively Parallel Signature Sequencing (MPSS) Transcription Profiling

The protocols for MPSS are described in detail in refs. 1 and 2. However, to make this paper self-contained, we will review the steps and methods used in the MPSS process that are most relevant to the present work. These methods are cDNA signature/tag conjugate library construction, microbead loading, and signature sequencing (see also Fig. 5).

cDNA Signature/Tag Conjugate Library. Poly(A)$^+$ mRNA was extracted from the tissue of interest and reverse-transcribed into a biotinylated first cDNA strand on which the second strand was synthesized to form a double-stranded biotinylated cDNA. The cDNA was digested with *Dpn*II restriction enzyme, and the resulting 5' digestion overhang was used to ligate an adaptor containing the *Mme*I (a type IIs endonuclease) recognition site adjacent to the *Dpn*II site. The resulting biotinylated oligomers were purified with streptavidin beads. *Mme*I was used to digest a fragment of only 20 bases off the beads, which includes the sequence GATC that is derived from the anchoring *Dpn*II site. In this way, a 20-mer fragment containing the *Dpn*II site most proximal to the poly(A) tail is isolated for each transcript. The 13-mer segment contiguous to the GATC, along with the GATC, forms a 17-mer sequence that will eventually be sequenced, and that we call the signature. In this way, for each tissue of interest, a signature library containing 20-base signatures of every transcript is constructed. After ligating primer-containing adaptors on the 3' end of the fragments, the signature library is PCR-amplified. The PCR product is then digested with cloning restriction enzymes (*Dpn*II and *Sfn*I), PAGE-purified, and cloned into the tag vectors that contain a complexity of 8^8 different unique tags, as described (1, 2). The complexity of the resulting signature/tag library is determined by transformation of the ligation reactions to *Escherichia coli* and the subsequent tittering, also as described (1, 2).

Microbead Loading. After the complexity was determined by the tittering, multiple pools of bacteria corresponding to 640,000 signatures/tags (≈4% of the tag complexity) were cultured and their plasmids purified. The signature/tag conjugates were amplified from the plasmids. After removing the bottom strand, the signature/tag conjugate is combined under stringent hybridization conditions with microbeads that have all the different complementary tags (antitags) attached. Because the tags in the sample of conjugates make up a small percentage (≈4%) of the total number of tags, only that percentage of the microbeads will be loaded with the template molecules. This fraction is concentrated into a library of loaded microbeads with a FACS. Each microbead in such a library has a population of $\approx 10^4$–10^5 identical copies of a single kind of signature molecule attached. Fixed aliquots of $\approx 1.5 \times 10^6$ signature-containing microbeads are assembled in a flow cell that constrains these microbeads to form a closely packed planar array that remains fixed as the signature sequences in each bead were determined by serial enzymatic reactions, as described below.

MPSS. The sequencing proceeds by the parallel identification of sets of four bases by ligation of encoded adaptors, followed by hybridization of labeled encoders and then removal of that set of four bases by *Bbv*I, a type IIs restriction enzyme whose recognition site is contained in the encoded adaptors, and then repetition of the process. All sequencing reagents are pumped through the flow cell. These fluorescent reactions occur underneath an automated microscope connected to a charge-coupled device detector used for the collection and imaging of fluorescent signals generated by the entire microbead array. These images are processed to yield the number of beads that have a given signature sequence. The procedure is completely parallel, facilitating large-scale high-quality sequencing of 17 nucleotides. Different initiating adaptors, whose type IIs restriction sites were offset by two bases, were ligated to two separate sets of microbeads containing a replicate of the same signature library to reduce signature losses from self-ligation of ends of signatures produced when cleavage with *Bbv*I fortuitously exposes palindromic (or other self-ligating) overhangs. Each of these two alternative sequencing reactions is called a stepper, and the two reactions are referred to as two-stepper (TS) and four-stepper (FS) sequencing. Each signature library was sequenced at least twice with the TS frame and at least twice with the FS frame.

Classification of Signatures from Genomic Sequence

The 17-mer signatures have a high specificity and typically match just one position in the transcripts of a complex genome (3). Each sequenced signature is assigned a class based on the position of the signature relative to annotated genes in the genome. The classification system was similar to that used for SAGE yeast data (4). Each signature was compared with the annotation and assigned a class depending on the position and strand relative to the poly(A) tail (≥13 As of a stretch of 15 bases within 50 bases away from the end of a sequence) and the poly(A) signal (AATAAA or ATTAAA) of annotated genes. Class 1, 2, or 3 refers to signatures that were located on the sense strand of a gene and correspond to the 3'-most *Dpn*II site of the sequence containing both the poly(A) tail and the poly(A) signal, the poly(A) signal alone, or the poly(A) tail alone, respectively. Class 4 refers to the sense-strand 3'-most signature derived from a sequence without the poly(A) tail or poly(A) signal information. Class 5 refers to the non-3'-most sense-stand signatures. Signatures of antisense contain corresponding classes of sense-strand signatures plus 10, whereas 20 is added to the classes of the signatures derived from the sequences of unknown orientation. During the classification routine, we also stored in our database the identifier of the gene in which the signature was found for all genomic signatures associated with a gene. Table 1 summarizes this information.

Statistical Significance of Differential Expression in Binary Comparisons

Given two replicate measurements θ_1 and θ_2 (with $\theta_2 > \theta_1$), we want to know whether the difference $\delta\theta = (\theta_2 - \theta_1)/\sqrt{2}$ is larger than a threshold $\delta\theta_t$ determined by a chosen P value P_0 according to

$$p_0 = \int_{|\delta\theta| \geq \delta\theta_t} d\delta\theta \, P(\delta\theta \mid \bar{\theta}),$$

where $P(\delta\theta \mid \bar{\theta})$ is the conditional probability of measuring $\delta\theta = (\theta_2 - \theta_1)/\sqrt{2}$ given that $(\theta_1 + \theta_2)/2 = \bar{\theta}$. This equation implicitly determines a function of $\delta\theta_t(\bar{\theta})$. A schematic representation of this function is plotted in red in Fig. 6, which also shows the geometric interpretation of the *P* value as the shaded area under the distribution for values of $\delta\theta$ larger than the measured one.

All vs. All Comparisons for Assignment of Significance to Time Traces: An Example

Measurements involving the time evolution of a group of cells are particularly important for understanding genes responsive to perturbations. It is usually the case that the time resolution of such measurements is limited, resulting in a small number of time points to be analyzed. Although this could be a problem in terms of the inability of the measurement to resolve small time scales, it can be used to our advantage, as will be discussed below. Suppose the aggregate transcripts per million (tpm) of a given signature is measured at *n* time points $t_0 = 0$ (i.e., unstimulated cells), t_1, t_2,…,t_{n-1}, yielding a vector of corresponding $\log_{10}$-tpms $(\theta_{t_0}, \theta_{t_1}, ..., \theta_{t_{n-1}})$. We are interested in determining whether the perturbation created any substantial change, beyond the level of noise. For this, we create a second vector whose entries are the *P* values corresponding to all the pair-wise comparisons at all time points. That is, we compute the vector with $n \times (n-1)/2$ entries

$$\boldsymbol{\pi} = \left(p(\theta_{t_0}, \theta_{t_1}), p(\theta_{t_0}, \theta_{t_2}), ..., p(\theta_{t_0}, \theta_{t_{n-1}}), p(\theta_{t_1}, \theta_{t_2}), ..., p(\theta_{t_{n-2}}, \theta_{t_{n-1}})\right).$$

If the perturbation did not affect the expression of this signature at any time point, then all the pair-wise comparisons would be as comparisons between replicate measurements, and we may expect a relatively large nonsignificant *P* value for each of the comparisons. However, if the perturbation significantly affected the expression of our signature at some time point beyond the level of variability between replicates, then we should expect a small *P* value for at least 1 of all the comparisons. It is reasonable to ask why we would like to compare all against all measurements and not just consecutive measurements, such as θ_{t_0} with θ_{t_1}, θ_{t_1} with θ_{t_2} and so on until $\theta_{t_{n-2}}$ with $\theta_{t_{n-1}}$. The reason for this is that there are cases when consecutive comparisons may not be beyond the level of noise. For example, even if the changes between θ_{t_0} and θ_{t_1} and that between θ_{t_1} and θ_{t_2} are not significant, it could be that the accumulated change between θ_{t_0} and θ_{t_2} is significant and enough to show that the signature is actually responding to the perturbation.

To exemplify this point, let us consider the case of one of the genes important in the transduction of the LPS signal. It is known (5, 6) that CD14, a differentiation antigen of monocytes, plays a role in LPS recognition. The time course of CD14 in our MPSS data is shown in Fig. 7. The *P* values corresponding to pair-wise comparisons between consecutive time points are all larger than 0.01; therefore, if we had set our significance threshold at 0.01 and were considering only adjacent time points, then CD14 would not be statistically significant. However, even if the pair-wise changes between times 4 h and 8 h (*P* value = 0.09) and between times 8 h and 24 h (*P* value = 0.03) are not significant

to a level of 0.01, those between times 4 h and 24 h (P value = 0.002) and between times 2 h and 24 h (P value = 0.003) are statistically significant.

As discussed in the main text, the pair-wise comparison vector $\boldsymbol{\pi}$ will be assigned an index of significance given by the minimum of the components in $\boldsymbol{\pi}$, that is

$$SI = \min_i \left(\boldsymbol{\pi} |_i \right) = \min_{i,j} \left(p(\theta_{t_i}, \theta_{t_j}) \right).$$

We call the significance index (SI) an index, because it is not strictly speaking a P value representing the probability of some event pertaining to the all-vs.-all comparison. It would, in principle, be possible to define a null hypothesis based on the maximum possible range in the time trace. However, this formalism would make the mathematics cumbersome and not so intuitive. Therefore, we prefer to sacrifice rigor in favor of clarity. In this way, a significant SI means at least one of the comparisons in the time trace has a P value smaller than the chosen threshold P_0. In the case of the time trace of CD14 discussed here, SI = 0.002. It is important to note that the most significant P value does not necessarily correspond to the largest fold change, given that the significance of a fold change depends of the expression level. In this way, a 5-fold change at small expression levels could be less significant than a 1.5-fold change at a high expression level.

Summary of Signature Libraries Obtained from Our MPSS Measurements

The MPSS measurements at all time points (as well as the replicates) of our LPS-stimulated macrophages resulted in seven signature libraries (see Table 2). The total number of MPSS runs incorporated into each library was four (two for the TS and two for the FS), except for the 8-h measurement, in which four TS and four FS runs were performed. The complete set of 32 sequencing runs comprised 130,150 17-base signatures, including both the TS and FS runs. These signatures were filtered by using two criteria, which we called reliability and significance, and which were designed to remove noisy sequences that may have resulted from erroneous processes or systematic errors. Briefly, the reliability filter removes signatures observed at least 50 times in only one sequencing run across all libraries (i.e., inconsistent expression). The significance filter removes signatures never observed above three tpm in any library (i.e., very low expression). Of the 130,150 distinct signatures, 47,841 were both significant and reliable (see Table 2). The genome partially served as an additional filter to remove erroneous (nonmatching) signatures. Most of our analyses focused on the 47,841 significant and reliable signatures.

Discussion on the Effects of Zero Measurements in MPSS

We expand here our discussion of the analysis of signatures that have a measured observation value of zero in one or more (but not all) of the MPSS sequencing runs constituting a given signature library. In this analysis, we consider basal measurements (unperturbed macrophages, i.e., $t = 0$ measurements) and always consider the TS measurements. Exemplary measurements are shown in the following table:

Signature	B1:1	B1:2	B2:1	B2:2	Nomenclature
1	8	17	15	25	Nonzeros in B1 – nonzeros in B2
2	0	4	5	11	One zero in B1 – nonzeros in B2
3	3	8	2	0	Nonzeros in B1 – one zero in B2
4	22	42	0	0	Nonzeros in B1 – two zeros in B2
5	0	0	15	7	Two zeros in B1 – nonzeros in B2
6	0	476	0	0	One zero in B1 – two zeros in B2
7	0	0	625	0	Two zeros in B1 – one zero in B2
8	0	23	0	115	One zero in B1 – one zero in B2
9	0	0	0	0	Two zeros in B1 – two zeros in B2

In this table, B1:1 indicates MPSS run 1 for biological replicate 1 of macrophages in their basal state (B, basal), and B1:2 indicates MPSS run 2 for the same sample. Likewise, B2:1 and B2:2 indicate the same for the second biological replicate. One conspicuous feature of the data is the appearance of many signatures for which there is one zero in one of the MPSS runs (e.g., B1:1; signature 2 in the table above) and a nonzero in the companion MPSS run (B1:2). Continuity arguments may suggest that if the count of a signature is zero in one run, then it should be small in the other. However, we can see there are signatures that were not measured in three MPSS runs, but that were measured in 476 beads in one of the runs (B1:2, signature 6). In this case, because the signature is also observed elsewhere (samples outside of B1 and B2), it is regarded as a reliable signature and escapes the reliable filter. Likewise, continuity suggests that the statistical behavior observed for sequences for which we measure a zero count in one of the runs must be similar enough to the statistical behavior of the signatures observed at low but positive counts such 1, 2, etc. Fig. 8*a* (for which the axes are in bead counts rather than in tpm) shows that for comparisons between B1:1 and B1:2 and for comparisons between B2:1 and B2:2, the behavior at zero is quite different from that at 1 (in the log–log plot, the log of zero was replaced by the log of 0.5 for clarity). For example, there are 4,487 paired signatures from B1:1 and B1:2, for which exactly one of the pair of measurements is zero, whereas there are 277 pairs for which exactly one of the pair of measurements is 1. Furthermore, the distributions of the points paired with zero and one are quite different, as shown in Fig. 8*a*. For example, for the B1:1 count of zero, the ordinate reaches values of the order of 10^3, whereas for an abscissa value of 1 or 2, the ordinate reaches only values of 10.

Accepting that the zeros have a distinct behavior, it is worth studying them in further detail. First, we count the various combinations of zero and nonzero measurements for those signatures, which exhibit at least one nonzero measurement in the four basal MPSS runs. This yields the following table:

B1 and B2	Nonzeros in B1 (6,875)	One zero in B1 (4,487)	Two zeros in B1 (8,209)
Nonzeros in B2	5,239	1,558	3,293

(10,090)			
One zero in B2 (6,839)	817	1,106	4,916
Two zeros in B2 (2,642)	819	1,823	0

We can pose the previous table in terms of conditional probabilities of the zero count in B2 given the zero count in B1:

Probability (B2|B1)	Nonzeros in B1, %	One zero in B1, %	Two zeros in B1, %
Nonzeros in B2	76.2	34.7	40.1
One zero in B2	11.9	24.6	59.9
Two zeros in B2	11.9	40.6	

The table for the conditional probability of the zero count in B1 given the zero count in B2 has roughly the same characteristics as the table above. It is clear that having nonzeros in both MPSS runs in B1 is predictive of having nonzeros in B2. However, having one or two zeros in B1 does not clearly determine what to expect in B2.

Our conjecture (which can be refined by addressing sequence-specific characteristics) is that signatures with count numbers below some threshold have a higher probability of not being measured by the two MPSS runs (B1:1 and B1:2 in the B1 case), and thus their count in one of the MPSS runs is zero. In these instances, the count of zero does not have to be considered as a real count (i.e., as a possible measurement value given the actual count) but rather as an artifact to be interpreted as a failure in measuring a signature whose value is not necessarily small.

Below, we will test a necessary condition that follows from the previous conjecture. Let us consider the signatures whose counts in B1 had exactly one zero in one of the two MPSS runs, but whose count in B2 had two nonzeros (such as, for example, signature 2 in the first table of this section). We will measure the distribution of counts for those signatures in B2 on the one hand and in B1 (only the nonzero measurement) on the other. If the measurement of zero in one of the two MPSS runs in B1 were because these signatures simply did not go through the whole process, then one should expect that the distribution of the counts of these signatures in any one of the MPSS runs in B2 (blue line in Fig. 8*b*) and in the nonzero MPSS runs in B1 (red line in Fig. 8*b*) coincide. This is exactly what Fig. 8*b* shows. The overlap of the red and blue curves shows that we should not take the zero of one of the MPSS runs as a real measure but rather as the absence of measurement. For the sake of completeness, we also plot in Fig. 8*b* the probability density functions (pdfs) of the counts in B1 and B2 of signatures, with nonzero counts in all the MPSS runs (green and black curves) superimposed on top of the pdfs corresponding to signatures with exactly one zero in one of the B1 MPSS runs. We see that the dynamic range of counts whose measurements consist of two nonzeros is considerably larger than that of measurements of signatures that have at least one MPSS run with a count of zero.

The consideration of whether a zero is to be considered a true measurement or the absence of a measurement is important in deciding what to do with the signatures for which we have one MPSS run claiming some value and the other claiming zero. In particular, it can be argued that if one of the MPSS runs yields zero, then it does not have to be averaged when we are computing the aggregate tpm for all runs. To see this, let us compare, signature by signature, the averages of signatures with two nonzeros in B2 and one nonzero in B1. If the zero in the B1 MPSS run were to be considered a real measurement, then after averaging the zero in B1, one would expect there will be roughly as many signatures with higher tpm in B1 than in B2 as signatures with higher tpm in B2 than in B1. If we do it this way, of the 1,558 signature with one nonzero in B1 and two nonzeros in B2, the average in B2 is larger than that in B1 in 1,363 of 1,558 cases. This can be explained simply by saying that we have lowered the averages in B1 artificially by adding a zero to signatures that actually were present at a higher level (the zero being an artifact). The global mean (which should be 0) of the differences between the logs of the averages in B1 and B2 was 0.4, which should be compared with the standard error of 0.01, giving a *z* score of 40. If instead we ignored the zero in one of the MPSS runs of B1, of the 1,558 signatures with one nonzero in B1 and two nonzeros in B2, the average in B2 is larger than the nonzero measurement in B1 in 900 cases (down 463 compared with the averaged B1 and closer to the ideal ≈780 of half and half), and the global mean (which should be 0) of the difference between the logs of the averages in B1 and B2 was 0.07, which should be compared with the standard error of 0.01, giving a *z* score of 7. This indicates that the trends obtained by ignoring the zero measurements are closer to the ideal unbiased comparison between B1 and B2.

Our conclusion is simply that when we measure a zero in one of the two MPSS runs and a nonzero in the other, then the zero has to be ignored (i.e., we take the maximum of the two runs). If the two runs in B1 measure zero, and one of the runs in B2 is nonzero, then the nonzero signature counts are distributed as in Fig. 8*b* (red and blue curves). This latter distribution can be used to determine the null hypothesis, when for a given signature the two MPSS runs comprising one of the replicates both yield values of zero.

Additional Details on the Three Null Hypotheses

At first sight, the results pertaining to a given signature from two MPSS experiments (each comprised of two or more MPSS runs) are the aggregate tpm values arising from each experiment. However, we obtain more information by considering the values of the counts of the MPSS runs used in the calculation of the aggregate tpm. As we show in the main text and in *Discussion on the Effects of Zero Measurements in MPSS*, if the individual MPSS runs yields zero, that gives us an indication that the measurements for that signature may be noisier. In that sense, it is convenient to generate null hypotheses using the information that goes into the computation of the final tpm values. In other words, the history of how we arrive at the aggregate tpm value gives us an indication of how reliable that value is. A measurement of the number of zero counts seems to be a particularly important factor within the history resulting in the final tpm value. If we decide to use one null hypothesis, then we will combine all signatures, and the final one

will be more lenient for the noisier signatures and more stringent for the less noisy ones . In other words, using a combination of null hypotheses allows us to have a uniform signal-to-noise ratio for all signatures.

Suppose we measure the same signature in two biological replicates, with MPSS run counts given by $\nu_1^{(1)}$, $\nu_2^{(1)}$, and $\nu_3^{(1)}$ in replica 1 and by $\nu_1^{(2)}$, $\nu_2^{(2)}$, and $\nu_3^{(2)}$ in replica 2. Depending on the pattern of zeros, we choose our conditional null hypotheses according to the following table:

	Replica 1		Replica 2	
Hypothesis	$\nu_1^{(1)}x\nu_2^{(1)}x\nu_3^{(1)}$	$\nu_1^{(1)}+\nu_2^{(1)}+\nu_3^{(1)}$	$\nu_1^{(2)}x\nu_2^{(2)}x\nu_3^{(2)}$	$\nu_1^{(2)}+\nu_2^{(2)}+\nu_3^{(2)}$
Nonzero	>0	>0	>0	>0
One zero	=0	>0	≥0	>0
One zero	≥0	>0	=0	>0
All zeros	=0	=0	≥0	>0
All zeros	≥0	>0	=0	=0

Fig. 9 *A* -*C* shows the results of each of these three null hypotheses for the aggregate $\log_{10}$(tpm) in replicas 1 (θ) and 2 (θ'). In all cases, we used replicates of the macrophage MPSS data after LPS stimulation at $t = 0$ and $t = 4$ h. All signatures were considered, measured by both the TS and FS sequencing. However, θ and θ' always correspond to either TS or FS. The deviation from the diagonal towards the lower *P* value range in Fig. 9*Ad* is mainly due to the two outliers highlighted by arrows in Fig. 9*Ac*. Fig. 9*Bd* shows a deviation from the diagonal at *P* values smaller than 0.05 because of the sparseness of the data and the consequent difficulty in obtaining a good estimation of the σ(μ) curve.

Statistics of Significant Signatures in the Signature Library

When considering only the reliable and significant signatures measured in at least one of the LPS runs, we found 47,841 signatures. Of these, only 4,980 (≈10%) have an index of significance <0.05. Of these 47,841, only 28,326 can be assigned a UniGene number at this time, and among these, 2,877 signatures (10%) have a SI <0.05. Each signature could have more than one assignment to a UniGene number. If among the many signatures corresponding to the same UniGene number, we pick the assignment that has the lowest SI, the list of signatures reduces to 12,657, of which 2,356 (20%) are statistically significant with a SI <0.05. Fig. 10*a* shows the histogram of the number of these signatures with a given SI. We can see that 80% of the signatures are above the threshold of 0.05, whereas the remaining 20% have SIs that reach values as small as 10^{-16}.

Let us now restrict our attention to only the 132 genes found by Nau *et al.* (7) (using Affymetrix Hu6800 GeneChips) as significantly induced after macrophage exposure to bacteria. We identified signatures corresponding to 127 of those 132 genes. Of the 127 genes identified in both datasets, 26 (20%) were nonsignificant, with SI >0.05. The remaining 101 significant genes have SIs that range from 10^{-13} to 0.05, as shown in Fig.

10*b*. These 101 genes, although statistically significant, are not the most significant genes in the LPS time course identified with MPSS.

The signatures measured using MPSS can be located in different regions with respect to the closest gene in the genome (see *Classification of Signatures from Genomic Sequence*), but by design, the method should find the 3'-most *Dpn*II site proximal to the poly(A) signal. Of all the unique UniGene ID nos. with the best SIs, signatures with SIs <0.05, 1,605 signatures (68%) corresponded to signatures with class values of 1, 2, or 3 (see Table 1). (At an SI threshold of 0.001, where we have 118 signatures, this fraction is 83%.) This is a considerable enrichment with respect to the 40% of 3'-most *Dpn*II site proximal to the poly(A) signal signatures contained among the 28,326 signatures associated with a UniGene entry. The point we wish to make in this respect is that this enrichment is achieved by subselecting the signatures with the same UniGene ID no. whose SI is the smallest, and among them, choosing the ones whose SI is smaller than a threshold.

Eisen Plot of the MPSS Statistically Nonsignificant Genes Among the Macrophage Activation Program Genes

It can be seen from the Eisen plot (Fig. 4) that the 101 genes deemed statistically significant according to our SI are indeed consistent in their behavior with Nau *et al.*'s (7) measurements. This is observed in the correct ordering of the temporal conditions: the first temporal condition measured at 1 h is the first in the dendrogram, followed by the 2-h measurement (with the MPSS and Affymetrix measurements at 2 h being quite similar), followed by the 4-h measurement (MPSS), followed by the Affymetrix 6 h, and so on. The pair of conditions measured in both platforms at 24 h are also quite similar. This consistency between the two platforms is due in part to the fact that the 101 genes chosen for this plot are the most significantly changing genes among the macrophage activation program genes defined by Nau *et al*. (7). That is, these are genes whose measurements have a high signal-to-noise ratio. This consistency is much less clear when the remaining 26 nonsignificant genes are considered in a similar plot, as shown in Fig. 11. In this case, the measurements at 2 h from the two platforms are dissimilar, as are measurements at 24 h. The 4-h measurement is farthest from the 2-h measurement, and the clear progression in time is absent, as is the clean separation between early and late responders. This disorganized arrangement in the temporal conditions is, in part, because these genes have a low signal-to-noise ratio in the MPSS measurements.

Categories of Earlier and Later Responders Among Genes Significant in Both Nau *et al.* (7) and MPSS Measurements

Table 3 lists the categories of 81 of the 101 significant genes (see Table 4) in both Nau *et al.*'s (7) and our MPSS data, as classified by Nau *et al.*, along with their time of response. The remaining 20 genes were classified as miscellaneous, unknown, or belonging to categories with less than five members, and we did not consider them. Antiapoptotic, adhesion, cytokines and chemokines, transcription-related, and signaling genes tend to be expressed soon after stimulation (i.e., active in the first temporal cluster of Fig. 4),

whereas enzymes (mostly associated with metabolism) and receptors are transcribed later (active in the second temporal cluster in Fig. 4).

1. Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., *et al*. (2000) *Nat. Biotechnol*. **18,** 630-634.
2. Brenner, S., Williams, S. R., Vermaas, E. H., Storck, T., Moon, K., McCollum, C., Mao, J. I., Luo, S., Kirchner, J. J., Eletr, S., *et al*. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 1665-1670.
3. Saha, S., Sparks, A. B., Rago, C., Akmaev, V., Wang, C. J., Vogelstein, B., Kinzler, K. W. & Velculescu, V. E. (2002) *Nat. Biotechnol*. **20,** 508-512.
4. Ball, C. A., Jin, H., Sherlock, G., Weng, S., Matese, J. C., Andrada, R., Binkley, G., Dolinski, K., Dwight, S. S., Harris, M. A., *et al*. (2001) *Nucleic Acids Res*. **29,** 80-81.
5. Kielian, T. L. & Blecha, F. (1995) *Immunopharmacology*. **29,** 187-205.
6. Wright, S. D., Ramos, R. A., Tobias, P. S., Ulevitch, R. J. & Mathison, J. C. (1990) *Science* **249,** 1431-1433.
7. Nau, G. J., Richmond, J. F., Schlesinger, A., Jennings, E. G., Lander, E. S. & Young, R. A. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 1503-1508.