Integrated Genomic and Proteomic Analyses of Gene Expression in Mammalian Cells*

Qiang Tian‡§¶, Serguei B. Stepianants§¶, Mao Mao§¶, Lee Weng¶, Megan C. Feetham‡, Michelle J. Doyle§, Eugene C. Yi‡, Hongyue Dai, Vesteinn Thorsson‡, Jimmy Eng‡, David Goodlett‡, Joel P. Berger¶, Bert Gunter¶, Peter S. Linseley¶, Roland B. Stoughton¶, Ruedi Aebersold‡, Steven J. Collins**, William A. Hanlon¶, and Leroy E. Hood‡

Using DNA microarrays together with quantitative proteomic techniques (ICAT reagents, two-dimensional DIGE, and MS), we evaluated the correlation of mRNA and protein levels in two hematopoietic cell lines representing distinct stages of myeloid differentiation, as well as in the livers of mice treated for different periods of time with three different peroxisome proliferative activated receptor agonists. We observe that the differential expression of mRNA (up or down) can capture at most 40% of the variation of protein expression. Although the overall pattern of protein expression is similar to that of mRNA expression, the incongruent expression between mRNAs and proteins emphasizes the importance of posttranscriptional regulatory mechanisms in cellular development or perturbation that can be unveiled only through integrated analyses of both proteins and mRNAs. Molecular & Cellular Proteomics 3:960–969, 2004.

Genome-wide mRNA expression profiling by means of DNA microarrays has proven to be a powerful approach in characterizing the changes in biological processes such as disease states, developmental stages, and responses to drugs or genetic perturbations (1). However, DNA arrays measure only the changes at the mRNA level. Most biological functions are executed by the proteins rather than mRNAs. While the expression of many genes is controlled at the transcriptional level, other genes also employ posttranscriptional regulation processes involving mRNA stability, translation initiation, and protein stability. An important issue is the extent to which the changing expression patterns of mRNAs reflect corresponding changes in their cognate proteins. Recent advances in quantitative proteomics, especially the application of ICAT reagents in conjunction with MS, have made possible simultaneous quantitative comparison of hundreds of proteins between two complex mixtures (2). Integrated analyses of mRNA and protein expression data by concurrent measurement of both have revealed moderate to poor correlation in yeast and Halobacteria (3–5). Discordant expression of protein and mRNA was also observed in lung adenocarcinomas (6). However, these analyses examined only one aspect of a biological system, i.e. the steady-state levels of mRNAs and proteins. Another important aspect that concerns the kinetic process of perturbation and how the correlation of mRNA and protein evolves during this process was not addressed. Here, we evaluated the correlation of mRNA and protein expression in mammalian systems under two experimental conditions. In the first, we compared steady-state levels of mRNAs and proteins between two related cell lines representing distinct hematopoietic stages, i.e. multipotent myeloid precursors versus lineage-committed promyelocytic cells. In the second condition, we used a mouse model to demonstrate the kinetic changes in liver mRNA and protein levels in response to treatment with three different drugs. In both cases, we observed a moderate correlation between mRNA and protein levels with the expression of mRNA reflecting at most 40% of the variation of protein expression. This integrated approach not only allows us to identify genes that are regulated at the transcription level, but also unveils important posttranscriptional regulatory mechanisms during both hematopoiesis and drug responses that would not be evident by examining either mRNA or protein expression alone. In addition, in the current study we describe a newly developed statistical error model for ICAT data analysis.

EXPERIMENTAL PROCEDURES

Cell Lines and Tissue Cultures—EML cells were maintained in Iscove’s modified Dulbecco’s medium (Invitrogen, Carlsbad, CA) containing 20% horse serum and 10% BHK/MKL-conditioned media as previously described (7). The BHK/MKL cell line was established by transfecting the BHK cell line with an expression vector containing a cDNA encoding the secretory form of mouse c-Kit ligand (stem cell factor). MPRO cells (8) were cultured in Dulbecco’s modified Eagle medium (Invitrogen) containing 10% FBS and 5% BHK-HM5-conditioned medium that produces murine granulocyte/macrophage colony-stimulating factor (GM-CSF).1

1 The abbreviations used are: GM-CSF, granulocyte-macrophage colony-stimulating factor; PPAR, peroxisome proliferative activated receptor; 2D, two-dimensional.
Expression Profiling Using DNA Microarray—The mouse oligonucleotide microarrays (Agilent Technologies, Palo Alto, CA) that contain 23,574 oligonucleotides (80-mer) corresponding to 22,788 mouse genes were hybridized with cDNAs labeled with Cy3 or Cy5 (CyDye; Amersham Pharmacia Biotech, Piscataway, NJ) as described previously (9). After hybridization, slides were washed and scanned using a confocal laser scanner (Agilent Technologies). Fluorescence intensities on scanned images were quantified, corrected for background noise, and normalized.

ICAT Labeling and μLC-MS/MS Analysis—A total of 4 × 10⁶ cells were lysed in hypotonic buffer A (15 mM KCl, 10 mM HEPEs, pH 7.6, 2 mM MgCl₂, 0.1 mM EDTA, 0.1% Nonidot P-40). After centrifugation, the supernatant was saved and used as non-nuclear fractions. The remaining nuclei were lysed in buffer B (1 mM KCl, 25 mM HEPEs, pH 7.6, 0.1 mM EDTA) and used as nuclear proteins. For ICAT labeling, 1 mg of either non-nuclear or nuclear proteins from EML and MPRO cells were denatured with 6 M urea and 0.05% SDS and immediately reduced with 5 mM tributylphosphine. The cysteine residues were labeled with strong cation exchange HPLC (2.1 mm × 200 mm poly-sulfonethyl A; PolyLC, Columbia, MD). Each cation exchange fraction was purified over monomeric avidin cartridge (Applied Biosystems) and analyzed by μLC-MS/MS as described previously (10). Tandem mass spectra for selected peptide ions were searched against the NCI mouse protein sequence database by using the SEQUEST algorithm and Peptide Prophet computer program (National Institutes of Systems Biology). Relative expression ratios were calculated by using the EXPRESS software tool (National Institutes for Systems Biology) and confirmed manually.

Mice for Kinetic Drug Responses—Male C57BL/6J mice (11–12 weeks old) were housed four per cage and given ad libitum access to rodent chow (Teklad 7012) and water during the study. Animals were housed alone or binned daily in the moring by oral gavage with vehicle (0.25% methylcellulose), the peroxisome proliferative activated receptor agonist WY-14643 (30 mg/kg) for 7.6, 0.1 mM EDTA) and used as nuclear proteins. For ICAT labeling, 1 mg of either non-nuclear or nuclear proteins from EML and MPRO cells were denatured with 6 M urea and 0.05% SDS and immediately reduced with 5 mM tributylphosphine. The cysteine residues were labeled with strong cation exchange HPLC (2.1 mm × 200 mm poly-sulfonethyl A; PolyLC, Columbia, MD). Each cation exchange fraction was purified over monomeric avidin cartridge (Applied Biosystems) and analyzed by μLC-MS/MS as described previously (10). Tandem mass spectra for selected peptide ions were searched against the NCI mouse protein sequence database by using the SEQUEST algorithm and Peptide Prophet computer program (National Institutes of Systems Biology). Relative expression ratios were calculated by using the EXPRESS software tool (National Institutes for Systems Biology) and confirmed manually.

**RESULTS**

Correlation Analysis of mRNA and Protein Levels During the Early Myeloid Differentiation Process—To gain further insight into the molecular mechanism governing the early myeloid
The correlation coefficient for the 150 signature genes between mRNA and protein is \( r_{\text{sig}} = 0.64 \), and the overall correlation for all the genes (425 in total) in the analysis is \( r_{\text{all}} = 0.59 \).

differentiation program and to identify stage-specific developmental markers, we compared mRNA and protein expressions of two murine bone marrow-derived cell lines, EML and MPRO, using oligonucleotide arrays and high-throughput quantitative proteomic analysis via ICAT-MS. EML cells are kit-ligand (stem cell factor, SCF)-dependent multipotent progenitors that can be induced to generate lineage-committed GM-CSF-dependent promyelocytic MPRO cells (7, 8). DNA microarray analysis identified 1,199 mRNAs (about 5% of the total genes on the array) that are differentially expressed by the EML and MPRO cells (\( p < 0.01 \)) with 568 expressed highly in EML cells and 631 in MPRO cells. ICAT-MS analysis quantitatively compared 2,919 cysteine-containing peptides (with peptide probability score greater than 0.9; Ref. 15) originating from 672 unique proteins. Among these 672 proteins, 419 were identified by more than one peptide, 253 were identified by a single peptide. We were able to map the abundance ratios of 425 proteins (268 identified by two or more peptides, 157 by single peptide) to their corresponding mRNA expression levels as detected by DNA microarrays. We have also developed a statistical error model for ICAT data analysis that employs the more objective value to measure the changes of protein expression instead of using the arbitrary fold change as adopted in earlier studies (3–5) (see “Experimental Procedures” for more details). We identified a total of 150 signature genes, i.e. those with significant changes (\( p < 0.01 \)) at either the protein and/or the mRNA level. Although 113 signature genes (76%) exhibited changes for mRNAs and their cognate proteins in the same direction (Fig. 1, 1st and 3rd quadrants), only 29 of them changed significantly at both mRNA and protein levels and were thus dubbed correlated genes (red). In contrast, 67 genes showed significant changes at the mRNA but not the protein level (green), whereas 52 genes showed significant changes at the protein but not the mRNA level (blue). Another two genes showed opposite expression patterns of mRNA and protein (brown). The correlation coefficient between mRNA and protein is 0.64 for the signature genes and 0.59 for all the genes examined. A complete list of signature genes is shown in Supplemental Table I.

Among the 29 correlated genes, one encodes c-kit, a receptor kinase and widely used hematopoietic stem and progenitor cell-surface marker that mediates signals by its ligand, kit ligand (also known as stem cell factor, SCF) (16). The expression levels of both c-kit mRNA and c-kit protein are much higher in the multipotent progenitor EML cells, with the c-kit mRNA level about 9-fold more and the c-kit protein level about 7-fold more compared with the MPRO cells. Consistent with this, the kit ligand protein also exhibits 5-fold more expression in the EML cells than in the MPRO cells. However, we observed no change at the mRNA level for kit ligand between these two cell lines. Thus, the expression of c-kit and kit ligand during myeloid differentiation may be regulated by different mechanisms: while c-kit expression is regulated transcriptionally, kit ligand expression requires posttranscriptional processes. If only the mRNA levels were examined, the developmental regulation of important signaling molecules like kit ligand would be missed. On the other hand, our DNA microarray analysis revealed a 2-fold enhanced expression of RACK1 (receptor of activated protein kinase C1) mRNA in MPRO cells compared with EML cells, indicating that this protein may be involved in myeloid development. However, ICAT-MS analysis showed no significant change at the protein level (1.1:1), which is consistent with a recent report showing that RACK1 associates with the common \( \beta \)-chain of the IL-5/IL-3/GM-CSF constitutively and is not altered by cellular stimulation (17). Thus, mRNA expression analysis alone will inevitably generate information that does not reflect changes in protein expression, and the complementary proteomic analysis will provide the much needed extra confirmation for microarray data.

Of particular note are two anti-correlated genes (those with significant changes at both mRNA and protein levels but in opposite directions; brown in Fig. 1). One encodes the \( \text{O} \)-subunit of mitochondrial protein \( \text{H}^-\text{ATP} \) synthase (Atp5o). It has been reported previously that the \( \beta \) subunit of \( \text{H}^-\text{ATP} \) synthase (\( \beta\)-F1-\( \text{ATP} \)ase) exhibits a similar paradoxical expression pattern, and a translation-inhibitory protein that binds to the 3′-UTR of its mRNA was identified in fetal liver and hepatoma extracts (18). Remarkably, eight other mitochondrial genes also exhibited significant lower levels of proteins in MPRO cells but higher or similar levels of mRNAs compared with EML cells (Table I). These findings suggest the existence of a commonly shared posttranscriptional regulatory mechanism for the expression of mitochondrial proteins as the multipotent EML cells differentiate into myeloid-specific MPRO.
The second anti-correlated gene is HNRNP A0, which is involved in RNA processing. Intriguingly, we found five out of six RNA-processing genes fell in the 2nd quadrant of Fig. 1, indicating a negative correlation between mRNA and protein levels (Table I). This result is in line with our earlier finding of 12 yeast RNA-processing genes (Ref. 3 and unpublished data) and strongly supports the hypothesis that posttranscriptional regulation is a conserved mechanism for controlling the expression of this particular class of genes. Thus, none of the above two expression patterns can be uncovered by measuring either mRNA or protein expression only.

**Significance of Protein/mRNA Correlation for the Early Myeloid Differentiation Process**—To further interpret the observed correlation coefficient, we tested it against two independent null hypotheses of no correlation ($r = 0$) and perfect correlation ($r = 1$) between the proteins and mRNAs. The rejection of the first null hypothesis is common and aims to show that the protein and mRNA expressions are not related by chance. The rejection of the second one indicates that the less than perfect correlation between the proteins and mRNAs is not simply a result of the noise in the data and thus reflects true biological differences. To assess the significance of the obtained correlation with respect to the null hypothesis of no correlation, we have estimated, using Fisher’s $z$-transform method (19), the $p$ value of the correlation of signature proteins/mRNAs to be $p \sim 10^{-20}$.

The alternative null hypothesis is that the observed correlation between protein and mRNA expressions could result from a perfect correlation $\sim 1$ corrupted by the noise in the data. This is based on the fact that even the correlation between different microarray slides hybridized with the identical sample could vary over a broad range depending on the expression signature and noise in the data. To test this hypothesis, we have computer-simulated protein and mRNA data that closely mimic our biological data in terms of the dynamic range and for which the correlation is 1. We have then performed 1,000 Montecarlo runs, perturbing these data by additive noise at each run. The noise in expression of each protein/mRNA was assumed to follow the Gaussian distribution with zero mean and width corresponding to the error bar of the protein/mRNA expression estimated from the error model. This assumption relies on the ability of the error model for microarrays and proteomics platforms to adequately capture the variability in the systems. We have developed the error models using multiple replicates of mRNA and protein data and have attempted to make them conservative. For example, at $p$ value of 0.01, we expect $\sim 250$ false positives from the same versus same hybridization experiment on the microarray with 25,000 genes. The error model falsely identifies only $\sim 10$ as significant, as a result of purposefully overestimating the measurement error. The distribution of correlation values resulting from Montecarlo simulations is given in Fig. 2. The median value of $p$ distribution is $\sim 0.9$, which is significantly larger than the observed $\sim 0.64$ correlation between the mRNA and protein expressions in EML and MPRO cells indicated by the red vertical line (Fig. 2). Thus, the significant difference between the observed biological correlation and simulated correlation is indicative of the true biological differences between mRNA and protein expressions.

**Correlation Analysis of Mouse Liver mRNA and Protein Levels During the Dynamic Process of Drug Response**—In another set of experiments, we evaluated how individual mRNA and protein levels correlate in response to multiple drug perturbations over time. Mice selected for proteomics and genomics analyses were treated with three PPAR agonists (WY-14653, TZD, and BRL-49653) daily for 1, 2, 3, and 7 doses. Six hours after the last dose administrated, mouse liver mRNA and protein levels were compared with that of vehicle control by oligonucleotide arrays and 2D-DIGE. For each spot on the gel, a two-way analysis of variance algorithm (20) was used to compute the $p$ value of variation between different drug treatments. Spots were rank-ordered

| Gene name                                                | LogRNA Ratio (MPRO/EML) | LogProtein Ratio (MPRO/EML) |
|----------------------------------------------------------|-------------------------|----------------------------|
| Mitochondrial                                            | 0.21                    | -0.24                      |
|                                          | 0.29                    | -0.57                      |
|                                          | 0.47                    | -0.4                       |
|                                          | -0.07                   | -0.53                      |
|                                          | -0.02                   | -0.52                      |
|                                          | -0.01                   | -0.5                       |
|                                          | 0.05                    | -0.4                       |
|                                          | 0.38                    | -0.32                      |
|                                          | -0.08                   | -0.23                      |
| RNA processing                                          | -0.19                   | 0.11                       |
|                                          | -0.12                   | 0.24                       |
|                                          | -0.04                   | 0.19                       |
|                                          | -0.04                   | 0.19                       |
|                                          | -0.02                   | 0.39                       |

Molecular & Cellular Proteomics 3.10 963
in the ascending order of their p value, such that the topmost spots showed consistent large variation between different drug treatments. The top 70 spots from the ranked list were further examined for their quality and brightness, which resulted in ~30 candidates to be sequenced by MS. Spots that contain peptides derived from multiple proteins were discarded, whereas different spots that correspond to the same peptide were consolidated. We identified a total of 12 candidate genes. The overall correlation coefficient between mRNAs and proteins for all 144 data points (12 genes × 3
drugs × 4 time points) is 0.54 (Fig. 3a). The expression ratios of mRNAs and proteins for the 12 candidate genes at each time point are plotted in Fig. 3b, where the individual protein/mRNA correlations have been computed using Equation 1. As shown in Fig. 3b, eight genes showed correlation of mRNA and protein levels above 0.65 (p < 0.01), three exhibited correlation between 0 and 0.65, and esterase 1 (ES1), a carboxylesterase, demonstrated negative correlation (r = −0.70). Carboxylesterases have shown a paradoxical expression between mRNA and protein levels in rat liver (21). These findings suggest that PPAR agonists affect gene expressions at both transcriptional and posttranscriptional levels. Thus, proper evaluation of drug responses requires the assessment of both.

**Comparison of Protein and mRNA Expression Patterns**

With the availability of robust mRNA profiling and rapidly maturing protein profiling technologies, an important question is how these two expression data relate to each other and provide deeper insight into drug action in the cell. We addressed this question by comparing changes of protein and mRNA expression patterns in response to PPARα and -γ agonists. Since in our study, only 12 significantly regulated mRNA/protein pairs have been identified, and a detailed comparison of mRNA and corresponding protein expression is not feasible for the majority of genes. We therefore performed a general comparison of the mRNA/protein expression patterns induced or repressed by the PPAR compounds. We selected signature mRNAs and proteins with abs(logRatio) > 0.2 and values of p < 0.05 in any condition and performed a one-dimensional k-means gene clustering using k = 2. For both mRNAs and proteins, the clustering revealed up- and down-
regulated expression patterns in clusters 1 and 2, respectively (Fig. 4). mRNA clustering clearly showed that the effect of the α agonist (WY-14653) was stronger than that of the two γ agonists (TZD and BRL-49653) and thus WY-14653 stood out as a separate group. This has also been validated by clustering analysis in the compound dimension (data not shown). Protein clustering data exhibited a similar but slightly lesser degree of separation of α and γ agonists (Fig. 4).

As shown in Fig. 4, the numbers of regulated proteins/mRNAs under similar statistical cuts are ~250 and ~2,500, respectively. Given that the total number of spots on the gel and genes on the microarray is ~1,000 and ~22,000, these numbers translate into relative percentages of 25% for proteins and 11% for mRNAs, respectively. Thus the relative proportion of the signature proteins/mRNAs is roughly similar across two sets. Furthermore, all signature proteins/mRNAs are almost equally divided into two up- and down-regulated patterns given by clusters 1 and 2 (Fig. 4), with the down-regulated cluster being slightly larger. Identification of the 12 matched protein/mRNA pairs on the protein and mRNA k-means clusters in Fig. 4 shows that 10 out of 12 pairs (~80%) reside in the corresponding up/down-regulated clusters and only two (~20%) show discordant regulations.

To further quantitatively assess how similar or different the three PPAR agonists are by using protein or mRNA expression data, we compared protein and mRNA expression patterns within clusters 1 or 2. For each drug and dose combination, we averaged gene expression ratios of all the proteins or mRNAs in each cluster to generate two (12 × 1) column vectors for the protein or mRNA data, respectively. We then compared each protein vector with corresponding mRNA vector in the up-regulated cluster 1 or the down-regulated cluster 2 (Fig. 5, a and b, respectively). Each point on these figures corresponds to one of the 12 drug and dose combinations to demonstrate how protein and mRNA expression patterns group the compounds in the study. Fig. 5 (a and b) exhibited a Pearson correlation of ~0.7 between the protein and mRNA responses to drug treatments. This indicates that protein and mRNA expression data as a whole carry similar information about the drug treatments and provide roughly similar grouping of the compounds. The separation of the PPARα agonist (WY-14653) from the two γ agonists (TZD and BRL-49653) was better in the mRNA data than the protein data. One explanation for this is that microarray profiling is a more mature technology with higher sensitivity than 2D-DIGE.

**DISCUSSION**

mRNA/protein expression, protein-protein interaction, and genetic data are keys to unlocking gene regulatory networks, biological pathways, and cellular responses to various perturbations, e.g. drug treatments. The ability to analyze these data synergistically is a gateway for understanding cell function and, ultimately, conducting model-driven systems biology. In this study, we attempted to deduce similarities and differences of the first two data types, i.e. mRNA and protein expressions. This is a first attempt that strives to provide a complete picture by comparing microarray data to two commonly used proteomics platforms, ICAT and 2D-DIGE. Even more importantly, we tried to provide a comprehensive comparison of both steady-state and dynamic-state correlations of proteins and mRNAs. Although our comparison is not perfect given the unavailability of complete matching set of proteins and mRNAs in the 2D-DIGE experiment and the inability to accurately assess cross-platform variations, our approaches for comparing mRNA and protein expression data provide a framework for future comparison of more comprehensive transcriptome and proteome data.

The ideal comparison of mRNA and protein data should employ two expression matrices $R_{ij}$ (for mRNAs) and $P_{ij}$ (for proteins) where $i = 1, \ldots, \text{number of conditions}$ and $j = 1, \ldots, \text{number of genes}$. The same conditions should be used for both protein and mRNA data with all the mRNA/protein sequences completely matched. In this ideal case, one needs to compare the rows of $R$ to the corresponding rows of $P$ for the number of genes across each condition, similar to what we did for the microarray/ICAT comparison; one shall also compare the columns of $R$ to the corresponding columns of $P$ for each gene across the number of conditions, similar to what we did for the microarray/2D-DIGE comparison. These analyses could be further complemented by various pattern-matching and clustering techniques in order to gain better insights into the similarities and differences of mRNA and protein expression and assess post-transcriptional regulation genome-wide.

Currently, mRNA expression profiles are often used as surrogates for protein expression. This will incur little problem for genes that are regulated at the transcriptional level that display correlated expressions of mRNAs and proteins. However, our examples of correlation analysis in mice indicate that the differential expression of mRNAs can be used to explain at most 40% ($\rho^2 = 0.6^2$) of the differential expression of proteins both in steady-state cell lines and under dynamic process of drug perturbations. This moderate correlation between mRNAs and proteins can be attributed either to the inherent true biological difference or to the variations within and between experimental platforms, i.e. measurement errors of microarray and quantitative proteomics. Based on our own experience, the within-platform noise of microarray can be estimated from the correlation between the replicates, i.e. the same samples hybridized to two microarray slides and correlation computed. Depending on the size of the signature, this correlation can range from 0 (same versus same) to ~0.9 (large signatures). Thus it is possible that even for this simple case the correlation can be corrupted by the noise in the platform. The within-platform error for ICAT technology is largely unknown, and no cross-platform error between microarray and ICAT technology has to date been assessed. To evaluate the effect of measurement errors from each platform, we applied the mRNA and protein error models and Monte-
carlo simulation to examine to what extent one can corrupt a perfect correlation solely by noise in the measurements. We found that the moderate correlation coefficient between mRNAs and proteins cannot be attributed solely to noise in the data and is more likely a reflection of the underlying biological mechanisms. More rigorous analysis should be performed in the future with larger datasets with good handle on both within and cross-platform variations and good linearity calibration within each platform. The comparison between protein and mRNA expression should also involve the assessment of false-negative and false-positive rates within each platform as well as the expression pattern-matching as described by He et al. (22).

2D-DIGE was originally developed by Unlu et al. and validated by Davison and colleagues (12, 23). It has been applied recently for the identification of esophageal scans, cell cancer-specific protein markers, and proteomic analysis of a model breast cancer cell system (24, 25). One of the advantages of 2D-DIGE is its capability of revealing posttranslational modifications that represent an essential aspect of cell physiology, development, or disease. However, this is beyond the scope of our current analysis. Our pattern comparison in

![Correlation analysis of protein and mRNA expression patterns as a whole in response to PPAR agonists treatments.](image)
the microarray/2D-DIGE data reveals that on average mRNAs and proteins behave consistently and that in the present data mRNA profiling has better discriminatory power. This could be due to multiple reasons such as microarray profiling being more mature technology with higher sensitivity than 2D-DIGE, or due to the fact that protein profiling is more variable in time. As a consequence, this protein expression data, especially without a detailed mRNA/protein mapping, will not be able to add additional information to the existing mRNA expression data for the purpose of distinguishing the biological effects of different compounds.

Taken together, the discrepancy between protein and mRNA expression that we have observed is most likely a result of the biology of gene expression rather than the measurement errors. It suggests various levels of regulation during protein synthesis, e.g. posttranscriptional, translational, or posttranslational regulation. Although mRNA profiling alone provides a wealth of information for compound classification, target validation, and biomarker selection, we provide examples that some important molecular events during hematopoietic development or drug response can be either missed or misinterpreted by this unilateral measurement. Thus, integrated analysis of both mRNAs and proteins is crucial to gain further insights into complex biological systems.

Acknowledgments—We thank Peggy McCann, Robert Chang, John Thompson, and Eric Muise for mouse PPAR agonist treatments, liver RNA, and protein isolation, and Rosetta Gene Expression Laboratory for the microarray experiments. Special thanks to Eugene Tan for useful discussions of the PPAR microarray data.

This project is supported by National Institutes of Health (NIH) Grant P01 DK 53074 (to L. E. H.) and by the National Heart, Lung, and Blood Institute, NIH, under contract no. N01-HV-28179 and NCI R33 CA 093302 (to R. A.), by NIH Grant HL54881 (to S. J. C.) and by Merck & Co., Inc. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

†To whom correspondence should be addressed: Giang Tian, Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103-8904. Tel.: 206-732-1308; Fax: 206-732-1299; E-mail: gtian@systemsbiology.org, Serguei B. Stepaniants, Rosetta Inpharmatics, 401 Terry Avenue North, Seattle, WA 98109. Tel.: 206-802-7348; Fax: 206-802-6411; E-mail: serguei.stepaniants@merck.com.

REFERENCES
1. Young, R. A. (1995) Biomedical discovery with DNA arrays. Cell 102, 9–15
2. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 17, 994–999
3. Ideker, T., Thorsson, V., Rian, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292, 929–934
4. Griffin, T. J., Gygi, S. P., Ideker, T., Rist, B., Eng, J., Hood, L., and Aebersold, R. (2002) Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. Mol. Cell. Proteomics 1, 323–333
5. Baliga, N. S., Pan, M., Goo, Y. A., Yi, E. C., Goodlett, D. R., Dimitrov, K., Shannon, P., Aebersold, R., Ng, W. Y., and Hood, L. (2002) Coordinate regulation of energy transduction modules in Halobacterium sp. analyzed by a global systems approach. Proc. Natl. Acad. Sci. U. S. A. 99, 14913–14918
6. Chen, G., Gharib, T. G., Huang, C. C., Taylor, J. M., Misek, D. E., Kardia, S. L., Giordano, T. J., Iannettoni, M. D., Orringer, M. B., Hanash, S. M., and Beer, D. G. (2002) Discordant protein and mRNA expression in lung adenocarcinomas. Mol. Cell. Proteomics 1, 304–313
7. Tsai, S., Bartelmez, S., Sitnicka, E., and Collins, S. (1994) Lymphohematopoietic progenitors immortalized by a retroviral vector harboring a dominant-negative retinoic acid receptor can recapitulate lymphoid, myeloid, and erythroid development. Genes Dev. 8, 2831–2841
8. Tsai, S., and Collins, S. (1993) A dominant negative retinoic acid receptor blocks neutrophil differentiation at the promyelocyte stage. Proc. Natl. Acad. Sci. U. S. A. 90, 7153–7157
9. Hughes, T. R., Mao, M., Jones, A. R., Burchard, J., Marton, M. J., Shannon, K. W., Lifkowitz, S. M., Ziman, M., Schelter, J. M., Meyer, M. R., Kobayashi, S., Davis, C., Dai, H., He, Y. D., Stepaniants, S. B., Cavet, G., Walker, W. L., West, A., Coffey, E., Shoemaker, D. D., Stoughton, R., Blanchard, A. P., Friend, S. H., and Lincoln, P. S. (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer, Nat. Biotechnol. 19, 342–347
10. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat. Biotechnol. 19, 946–951
11. Koyama, H., Boueres, J. K., Han, W., Metzger, E. J., Bergman, J. P., Gratals, D. F., Miller, D. J., Tolman, R. L., Machaul, K. L., Berger, J. P., Doebber, T. W., Leung, K., Moller, D. E., Heck, J. V., and Sahoo, S. P. (2003) 5-Aryl thiazolidine-2,4-diones as selective PPARγ agonists. Bioorg. Med. Chem. Lett. 13, 1801–1804
12. Unlu, M., Morgan, M. E., and Minden, J. S. (1997) Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. Electrophoresis 18, 2071–2077
13. Roberts, C. J., Nelson, B., Marton, M. J., Stoughton, R., Meyer, M. R., Bennett, H. A., He, Y. D., Dai, H., Walker, W. L., Hughes, T. R., Tyers, M., Boone, C., and Friend, S. H. (2000) Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science 287, 873–880
14. Stoughton, R. S., and Dai, H. (2002) Statistical combining of cell expression profiles. United States Patent No. 6,351,712
15. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 74, 5383–5392
16. Broudy, V. C. (1997) Stem cell factor and hematopoiesis. Blood 90, 1345–1364
17. Geisjen, N., Spaargaren, M., Raaijmakers, J. A., Lammers, J. W., Koenderman, L., and Coffer, P. J. (1999) Association of RACK1 and PKC δ-chain of the IL-5/IL-3/GM-CSF receptor. Oncogene 18, 5126–5130
18. de Heredia, M. L., Izquierdo, J. M., and Cuevza, J. M. (2000) A conserved mechanism for controlling the translation of β-F1-ATPase mRNA between the fetal liver and cancer cells. J. Biol. Chem. 275, 7430–7437
19. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992) Numerical Recipes in C, The Art of Scientific Computing, 2nd Ed., University of Cambridge Press, Cambridge, United Kingdom
20. Bogus, D. P., Hunter, W. G., and Hunt, J. S. (1976) Statistics for Experimenters, An Introduction to Design, Data Analysis and Model Building. John Wiley & Sons, Inc., New York.
21. Yang, D., Li, Y., Yuan, X., Matoney, L., and Yan, B. (2001) Regulation of rat cardiac cytochrome c oxidase expression by 2,3,7,8-tetrachlorodibenzo-para-dioxin (TCDD): A dose-dependent decrease in mRNA levels but a biphasic change in protein levels and activity. Toxicol. Sci. 64, 20–29
22. He, Y. D., Dai, H., Schadt, E. E., Cavet, G., Edwards, S. W., Stepaniants, S. B., Duenwald, S., Kleinhanz, R., Jones, A. R., Shoemaker, D. D., and Stoughton, R. B. (2003) Microarray standard data set and figures of merit for comparing data processing methods and experiment designs. Bioin-

968 Molecular & Cellular Proteomics 3.10
23. Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., and Davison, M. (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. Proteomics 1, 377–396

24. Zhou, G., Li, H., DeCamp, D., Chen, S., Shu, H., Gong, Y., Flaig, M., Gillespie, J. W., Hu, N., Taylor, P. R., Emmert-Buck, M. R., Liotta, L. A., Petricoin, E. F., 3rd, and Zhao, Y. (2002) 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. Mol. Cell. Proteomics 1, 117–124

25. Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M. J., Cramer, R., Waterfield, M. D., and Timms, J. F. (2002) Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. Mol. Cell. Proteomics 1, 91–98