ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Microarray-Based Gene Expression Analysis as a Process Characterization Tool to Establish Comparability of Complex Biological Products: Scale-Up of a Whole-Cell Immunotherapy Product

Min Wang,[1] Ryan S. Senger,[2,3] Carlos Paredes,[3] Gautam G. Banik,[4] Andy Lin,[4] Eleftherios T. Papoutsakis[1,2,3]

[1]Interdepartmental Biological Sciences Program, Northwestern University, Evanston, Illinois; e-mail: epaps@udel.edu

[2]Department of Chemical Engineering, Delaware Biotechnology Institute, University of Delaware, 15 Innovation Way, Newark, Delaware 19711; telephone: 302-831-8376; fax: 302-831-4841

[3]Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois

[4]Cell Genesys, Inc., South San Francisco, California

Received 23 January 2009; revision received 7 May 2009; accepted 5 June 2009

Published online 8 June 2009 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.22441

ABSTRACT: Whole-cell immunotherapies and other cellular therapies have shown promising results in clinical trials. Due to the complex nature of the whole cell product and of the sometimes limited correlation of clinical potency with the proposed mechanism of action, these cellular immunotherapy products are generally not considered well characterized. Therefore, one major challenge in the product development of whole cell therapies is the ability to demonstrate comparability of product after changes in the manufacturing process. Such changes are nearly inevitable with increase in manufacturing experience leading to improved and robust processes that may have higher commercial feasibility. In order to comprehensively assess the impact of the process changes on the final product, and thus establish comparability, a matrix of characterization assays (in addition to lot release assays) assessing the various aspects of the cellular product are required. In this study, we assessed the capability of DNA-microarray-based, gene-expression analysis as a characterization tool using GVAX cancer immunotherapy cells manufactured by Cell Genesys, Inc. The GVAX immunotherapy product consists two prostate cancer cell lines (CG1940 and CG8711) engineered to secrete human GM-CSF. To demonstrate the capability of the assay, we assessed the transcriptional changes in the product when produced in the presence or absence of fetal bovine serum, and under normal and hypoxic conditions, both changes intended to stress the cell lines. We then assessed the impact of an approximately 10-fold process scale-up on the final product at the transcriptional level. These data were used to develop comparisons and statistical analyses suitable for characterizing culture reproducibility and cellular product similarity. Use of gene-expression data for process characterization proved to be a reproducible and sensitive method for detecting differences due to small or large changes in culture conditions as might be encountered in process scale-up or unanticipated bioprocess failures. Gene expression analysis demonstrated that cell products of representative lots under the same production process and at the same production scale were statistically identical. Large process changes that resulted from the artificial stress conditions used (absence of FBS and induction of hypoxia) displayed profoundly different gene expression patterns. We propose the use of simple *t*-test analysis in combination with the herein introduced expression ratio with mean intensity (ERMI) analysis as useful tools for process characterization by global gene expression analysis.

Biotechnol. Bioeng. 2009;104: 796–808.

© 2009 Wiley Periodicals, Inc.

KEYWORDS: cellular vaccine; cancer immunotherapy; process characterization; microarrays; cell culture; genome scale analysis; scale up; quality control; statistical analysis

Min Wang's present address is Abbott Bioresearch Center, Worcester, MA 01605.
Ryan S. Senger's present address is Biological Systems Engineering Department, Virginia Polytechnic Institute and State University, Blacksburg, VA.
Carlos Paredes's present address is Cobalt Biofuels, 500 Clyde Avenue, Mountain View, CA 94043.
Correspondence to: E.T. Papoutsakis

Introduction

With recent advances in cell and molecular biology, the potential of whole-cell immunotherapy has been widely

© 2009 Wiley Periodicals, Inc.

recognized as an clinical application for the future (Copier and Dalgleish, 2006; Guinn et al., 2007; Ward and McNeel, 2007). Increasing attention and efforts have been invested in research and clinical trials of whole-cell immunotherapies (Guinn et al., 2008). Several of the proposed immunotherapies involve established cell lines (Copier and Dalgleish, 2006), while others involve patient-specific cells (Guinn et al., 2007). Because these therapies utilize the whole cells and not products produced by the cells, assessing product quality and process reproducibility constitutes a major challenge. For cellular immunotherapies, where the cell is a multivalent source of antigens, all types of proteins and other cellular components (secreted, surface, or intracellular) are likely important in stimulating a beneficiary immune response. Due to the complex nature of the whole cell product and since clinical potency often does not correlate directly with the proposed mechanism of action, these cellular immunotherapy products are generally not considered well characterized. Thus, the manufacturing process must be tightly controlled within established process-parameter ranges, because even small unintentional process changes (such as process-control failures resulting in process parameters outside set points, or changes in raw materials used for media preparation) may result in changes in the cellular properties, which could be of importance for the intended therapeutic outcomes. However, manufacturing changes are inevitable with increased manufacturing experience. These process changes seek to increase yield, establish a more robust process or a more commercially viable process. Some examples of such process changes are changes in manufacturing equipment, changes in process parameters, changes in raw-material vendors, scale-up or changes in manufacturing facility.

Although process changes are difficult to implement for biologicals that are not considered well characterized, the impact of specific process changes can be assessed. How does one quickly and predictably assess the impact of small or large changes in process parameters or scale on the bioprocess outcomes, especially when the product is the whole cell or complex cellular or viral components? A logical approach is to use panels of surface or intracellular proteins in the case of cellular immunotherapies based on the presumed role of these proteins in the intended immune response. However, such assays alone may not be sufficient for characterizing the product efficacy/quality since there may be limited correlation of clinical efficacy with the proposed mechanism of action. In order to comprehensively assess the impact of the process changes on the final product, and thus establish comparability, a matrix of characterization assays (in addition to lot release assays) assessing the various aspects of the cellular product are required. These assays would include assays characterizing the consistency of antigen presentation and assays characterizing the consistency of the immunological cascade.

Here, we examine the capability of DNA-microarray-based gene expression analysis as a characterization tool to evaluate the consistency of gene expression of a cellular product at the transcriptional level following process changes, most notably production scale. To this effect, we employed two cell lines of the GVAX whole-cell prostate cancer immunotherapy platform of Cell Genesys, Inc. (South San Francisco, CA). We made the logical hypothesis that, given a sensitive and reproducible assay for gene expression analysis, gene expression changes should effectively capture any significant changes in the cellular properties/composition due to process changes. For example, even if process changes may impact only post-transcriptional cellular process (such as changes in translational or post-translational processes in the cell), the sensitivity of transcriptional control would likely produce changes in transcript abundances as a result of such post-transcriptional alterations. This is of course an empirical conjecture at this point, but a logical one that is not contradicted by experimental evidence.

Gene expression profiling by DNA-microarray analysis is widely used to examine cellular responses to genetic or environmental changes. However, this technology has not yet been widely explored in the context of process characterization in producing cellular products or more generally in examining the impact of relevant bioprocess parameters in animal cell biotechnology. A few examples of using this technology comes from our own laboratory, whereby the impact of oxygen tension, serum and cytokines on T-cell expansion and cellular properties was investigated (Haddad et al., 2004; Ramsborg and Papoutsakis, 2007; Ramsborg et al., 2004). Microarray analysis has now reached the technological level where, in suitably able laboratories, it can be assumed to be a very accurate and reliable technology capable of quantitatively monitoring transcript abundance at virtually the full genome scale for any sequenced organism or virus.

Materials and Methods

Samples and Cell Lines

The Cell Genesys, Inc. GVAX® whole-cell prostate cancer immunotherapy product consists of two human prostate cancer cell lines (PC-3 and LNCaP) engineered to secrete human GM-CSF. The GM-CSF modified PC-3 cells (Kaighn et al., 1979) are used to generate a product called CG1940, and the GM-CSF modified LNCaP cells (Horoszewicz et al., 1983) are used to generate a product called CG8711. A master and working cell bank was established for each cell line and was fully tested per ICH guidelines (The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH); http://www.ich.org/cache/compo/276-254-1.html). Each production lot of CG1940 and CG8711 was initiated from an aliquot of the working cell bank, which was thawed and serially passaged until the cells were inoculated in a 50 L-scale

perfusion bioreactor. The bioreactor was controlled for pH, temperature and DO. Upon reaching harvest cell density, cells were concentrated by continuous centrifugation, formulated in a cryoprotectant and stored in the vapor phase of liquid N_2. For scale-up, the cells were grown under identical environmental conditions in a perfusion 500 L bioreactor and formulated into the identical cryoprotectant.

RNA Extraction, Labeling, and Microarray Hybridization

Total RNA was extracted from frozen cells using the Total RNA Isolation Mini Kit (Agilent, Wilmington, DE). RNA samples were re-suspended in RNase-free water and stored at −80°C. RNA yield and purity were assessed at 260 and 280 nm (Biomate 3, Thermo Spectronic, Marietta, OH). RNA integrity was evaluated using a Bioanalyzer 2100 (Agilent). RNA samples with RNA Integrity Number (RIN) above 9 were processed with labeling and amplification reaction using Low Input Fluorescent Amplification Kit (Agilent, 5190-0444).

In the initial stages of this study, the Human 1A Oligo Microarray (V2) (Agilent) platform, which contained approximately 18,500 genes and transcripts, was used. Once available, the Whole Human Genome Microarray 44K (Agilent), representing approximately 41,000 genes and transcripts, was used. Hybridization, washing and scanning were carried out following the manufacturer's protocol. Feature Extraction (Agilent G2567AA, version 7.2 and 9.5) software was used to identify spot and feature outliers. Spot quality was assessed and, if not discarded, multiple spots of the same gene were averaged (geometric mean).

Microarray Analysis: Technical Replicates and Dye Swaps

The Agilent technology used is a two color dye-based (Cyanine 3 (Cy3) and Cyanine 5 (Cy5)) platform, which allows direct comparison of two RNA preparations on a microarray. To be able to compare multiple experimental samples in this study, a "reference" design (Ramsborg and Papoutsakis, 2007) was used for the microarray analysis. The reference RNA was constructed by pooling equal amounts of RNA from three similarly manufactured cell lots for each cell line as described below. Thus, the RNA from an experimental sample of a cell line was labeled with Cy3 and the pooled-reference RNA of the same cell line was labeled with Cy5. These two were hybridized onto the same microarray. In order to normalize for and thus factor out dye-incorporation bias, the same sample was also labeled in the opposite dye orientation; this is known as "dye swap." Two technical replicates for each dye orientation were generated for each sample, for a total of 4 microarray hybridizations.

Reference RNA Pools and Sample-Culture Lots

RNA from three lots of each cell line manufactured by the same 50 L process (identical culture conditions) was pooled in equal ratios to generate the reference RNA. For the CG1940 cell line, pooled reference RNA consisted of equal ratios of RNA from three 50 L lots: P50-1, P50-2, and P50-3. For the CG8711 cell line, the pooled reference RNA consisted of equal ratios of RNA from three 50 L lots: L50-1, L50-2, and L50-3. A fourth 50 L lot produced under identical culture conditions and scale was analyzed for each cell line: P50-4 for CG1940 and L50-4 for CG8711. The 500 L culture lots analyzed were designated P500 for CG1940 and L500 for CG8711. For the CG1940 line, stressed product (called P50-h) was generated by exposing the production culture to hypoxic conditions for 24 h. For CG8711, stressed product (called L50-s) was generated by growing the cells in the absence of fetal bovine serum (FBS) for 72 h. 50 L lots (one for each cell line) were also analyzed using the full-genome microarrays (Human Genome Microarray 44K (Agilent)): P50-2 for CG1940 and L50-2 for CG8711.

Microarray Data Analysis: Statistical Metrics and Gene Ontology Analysis

Signal intensities from microarray slides were normalized using the segmental nearest-neighbor logarithmic expression ratio of the mean (SNNLERM) algorithm (Yang et al., 2003) following initial processing with Agilent's Feature Extraction software. Several measures and statistical analyses were then developed in order to assess potential differences in global expression profiles of the analyzed samples. The first method of analysis is referred to as the Expression Ratio with Mean Intensity (ERMI) analysis. Normalized gene expression ratios (sample intensity/reference-RNA pool intensity) and spot intensities were averaged over all technical replicates and dye-swap configurations to remove (or minimize) the dye-bias. We then constructed ERMI plots that consist of averaged expression ratios plotted as a function of the averaged mean signal intensities. Genes with an expression ratio of $>1/2$ and <2 were declared to be within the 2-fold expression window. Likewise, a threefold expression window encompassed the range of expression ratios from 1/3 to 3. The numbers of genes falling outside these expression windows were used as potential metrics in order to assess expression differences or similarity/identity of samples. We also calculated the number of "erroneous probes" in this analysis. We defined an "erroneous probe" as a probe showing at least twofold "differential expression" (in the same direction) in all samples comprising the reference-RNA pool when each individual sample is hybridized against the reference-RNA pool.

For each probe, we also examined if the average of expression ratios from technical replicates and dye-swaps for any chosen sample was statistically equal (with a given confidence) from that of another sample. The hypothesis that two mean expression ratios were equal with variances

normally distributed about the mean was used for this analysis, and a two-tailed *t*-test was performed at the 95% confidence ($\alpha = 0.025$) level. The analysis was then repeated at the 99% confidence ($\alpha = 0.005$) level. The percentage of expression ratios determined to be "equal" was used as another metric in determining differences or similarity in gene expression between samples. Further, for genes determined to be statistically "different" at the specified confidence level, the number of genes with expression ratios falling outside the two- and threefold differential expression windows were used as an additional metric for comparison of samples.

Further analysis (hierarchical clustering and gene ontology analysis) was carried out using the "Multi-Experiment Viewer (MeV)" (MeV v4.3.02) from The Institute for Genomic Research (TIGR) (Saeed et al., 2003). One-way hierarchical clustering on lots/samples was carried out with the Euclidean-distance metric. The ordering optimization on sample leaf and average linkage were used for clustering. For cases in which a large number of genes were found to be differentially expressed, Gene Ontology (GO) analysis (Ashburner et al., 2000) was carried out using the EASE program (Hosack et al., 2003). The analysis allows the user to probe if sets of differentially expressed genes are linked by a physiological characteristic that could be traced back to cell cultivation conditions. The *GO Biological process*, *GO Molecular Function*, and *GO Cellular Location* were the Gene Ontology databases used in the analysis, and only EASE scores from 0 to 0.1 were considered.

Results

The Expression Pattern of "Same:" Experimental White Noise and Dye-Bias Effects

The capability, sensitivity, and reproducibility of the microarray approach for assessing product quality was tested by obtaining gene expression profiles of self–self hybridizations to test dye-bias effects. The self-self hybridization experiment involved labeling a CG1940 pooled RNA sample (later used as reference RNA) with Cy3 and Cy5 and hybridizing the labeled material to the microarrays, in four (4) technical replicates. The RNA was handled the same way for all hybridizations. Thus, only genes with a significant "dye-bias" will show an expression ratio (Cy3 channel intensity/Cy5 channel intensity) significantly different from 1. Dye bias could result from non-equal labeling of RNA with the two dyes, but also differences in dye-scanning detection sensitivity (Borden et al., 2005). As a result of the latter, at low signal intensities, the Cy5 channel intensity measurements will be accompanied by greater noise than low signals in the Cy3 channel (Borden et al., 2005), and the asymmetric noise of the data at the low intensity range (especially below 150) of Figure 1 likely reflects this fact. The CG1940 reference RNA consisted of equal ratios of RNA from three 50 L lots (P50-1, P50-2, and P50-3) (see Materials and Methods Section). The expression ratio with mean intensity (ERMI) plot of normalized microarray data from the 4 averaged replicates of the self–self comparison demonstrated a high degree of reproducibility, especially for mean intensity values >150 (Fig. 1). Of the 17,851 microarray elements (probes; representing genes and transcripts) with at least 3 (of 4 possible) readable data values, 15,912 (89.1%) had an expression ratio (Cy3 channel intensity/Cy5 channel intensity) between 0.9 and 1.1. A two-tailed *t*-test ($\alpha = 0.025$) was used to determine if the averaged expression ratio for each gene was statistically different from 1 (an expression ratio of 1 indicates no dye bias). A total of 9,003 genes (50.4%) showed an expression ratio indistinguishable from 1. However, for the 8,240 genes with an averaged mean intensity value >150, $>60\%$ had an expression ratio indistinguishable from 1 using the *t*-test, and 98.9% of these genes had expression ratios between 0.9 and 1.1. These results further enforced the expectation that there is some dye bias in this microarray technology, thus making the dye-swap experiments and technical replicates of arrays an absolute necessity. These results also suggested that only mean intensity values above 150 should be used for analysis. We have reported similar intensity cutoff values using a direct experimental approach based on a prokaryote (Paredes et al., 2007). We thus consider data with intensities below 150 as unsuitable for analysis, and such data were filtered out in the subsequent analyses in this work.

A First Assessment of What Is "Different": Gene Expression Comparison of Two Different Cell Lines

The global gene expression profiles of the two different cell lines, CG1940 and CG8711, were directly compared. Six microarrays using the CG1940 reference-pool RNA (see Materials and Methods Section) and CG8711 reference-pool RNA (see Materials and Methods Section) were hybridized using the dye-swap design, with a total of three technical replicates for each dye configuration. Averaged ERMI results of normalized data are shown as Figure 2. A two-tailed *t*-test (α =0.025) was used to compare the mean expression ratios of each gene to the value 1 (no differentiable expression). A total of 16,106 genes (all of with contained at least 3 expression ratio values) were used in the analysis. Since we are comparing two different cell lines, a large number of genes would be expected to show expression ratios significantly different from 1 after normalizing for dye bias. Analysis of the results by the *t*-test showed that 7,175 probes (44.5% of the total) were statistically "different" with 95% confidence, and that 6,631 probes had an expression ratio indistinguishable from 1. The balance (14.4%) of genes was not found to be statistically "different" or indistinguishable from 1 given the confidence limits of the analysis. A total of 1,281 of the statistically "different" probes had expression ratios different by at least twofold, and 217 ($\sim$2.3%) of the probes with mean intensity values >150 had expression

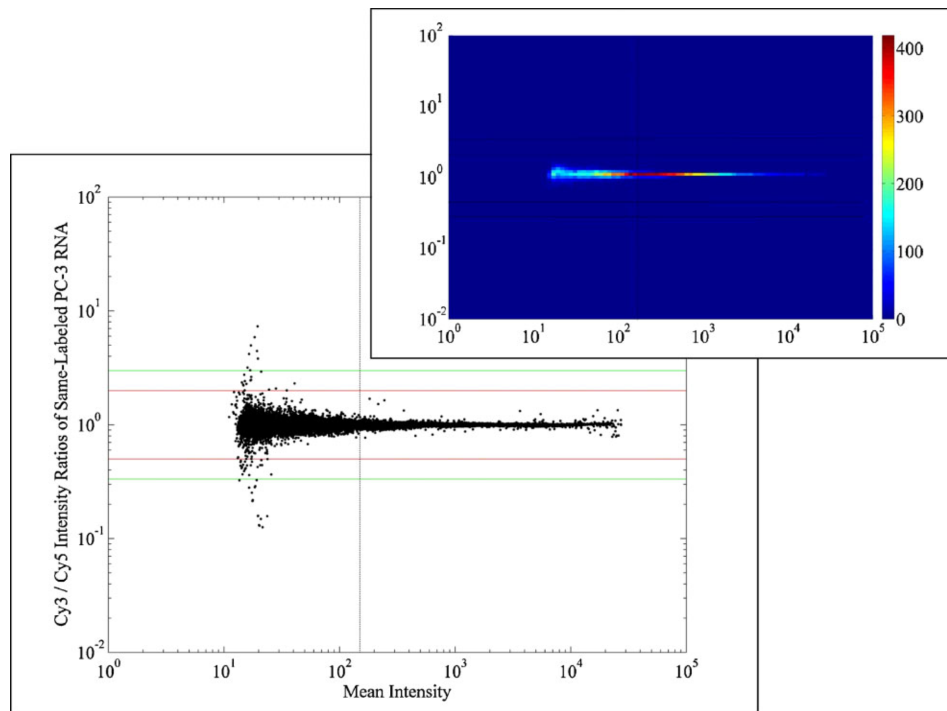

Figure 1. Expression ratios (*y* axis) versus mean intensity (*x* axis) averaged over four (4) self-self hybridizations (microarrays) for the reference-RNA pool of the CG1940 cell line. The heat map (upper right) displays the density of overlapping data points. The color scale (*z* axis) represents the number of data points contained in the grid section of the plot.

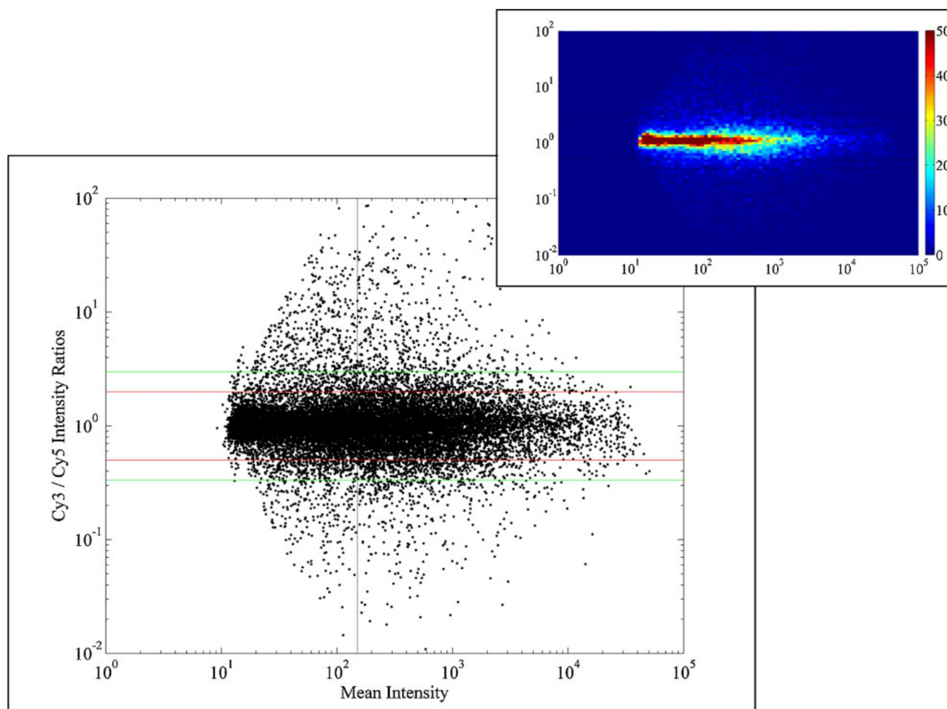

Figure 2. Expression ratios (*y* axis) (CG1877/CG1940) plotted as a function of mean expression intensity (*x* axis). For each cell line, the corresponding reference-RNA pool was used. The data represent the average of two dye-swap replicates, that is, a total of 4 microarrays. The heat map (upper right) displays the density of overlapping data points. The color scale (*z* axis) represents the number of data points contained in the grid section of the plot.

ratios >10 or <1/10. In addition, 215 of the "not different" probes (at the 95% confidence level) had expression ratios >2 or <1/2. These results provide a first assessment of the types of differences one might expect to observe for truly different biological samples.

Assessment of "Very Small" Differences: Comparison of Cell Lots Produced Under Similar Culture Conditions

"Normal" or "random" biological variation is expected among cells cultured in separate batches under the same bioprocessing conditions even under the strictest quality control. Thus, we desired to establish gene-expression differences, in the statistical sense, that would be measured by this microarray technology when comparing RNAs extracted from cells produced under the same bioprocessing conditions at the 50 L scale for each of the two cell lines, that is, to determine the intra-lot variability within the 50 L scale process. For each cell line, each of the three lots used to generate the reference RNA and a fourth 50 L lot (P50-4 for CG1940 and N50-4 for CG8711) were independently hybridized against the corresponding reference RNA of each cell line. Comparing each of the lots used to produce the reference RNA would be expected to result in somewhat smaller expression differences than comparing an independent 50 L lot against the reference RNA.

The ERMI plots for each lot hybridized with the reference-pool are given as Figure 3A–D for the CG1940 and Figure 4A–D for the CG8711. These plots demonstrate that very few probes with a mean intensity >150 have absolute values of expression ratios larger than 2. In particular for CG1940, as shown in Table I, of the approximately 16,000 probes analyzed, 0.13–0.40% of the probes with an average mean intensity >150 had expression ratios exceeding twofold. The percentage of probes exceeding a threefold expression ratio value ranged between 0% and 0.02%. Similar results were obtained for CG8711 (Table II): up to 0.19% of the probes analyzed exceeded twofold in the absolute value of the expression ratio, and 0.02% exceeded the threefold value. These very small percentages of "different" probes reflect both the inherent biological variability discussed above, as well as unavoidable random error in the experimental measurements. Two important conclusions can be drawn from these data. First, that there is cell-line dependence in microarray analysis outcomes in the statistical sense, namely that CG8711 consistently results in lesser variability in microarray-analysis outcomes compared to CG1940, as can be seen from the data in Tables I and II. Second, the 4th lot of each cell line, which was not used in the preparation of the reference RNA, appears indistinguishable from the three lots that were used in the preparation of the reference RNA. This suggests that the quality control of the 50 L processes is very robust and produces cells that are "virtually identical" when assessed by gene expression analysis.

Assessment of "Large" Differences: Hypoxia and Serum-Withdrawal Process Perturbations

The ability of the assay to detect significant biological changes was further assessed by exposing the CG1940 cells to hypoxic excursion for 24 h and the CG8711 cells to serum starvation excursion for 72 h. These two conditions were selected because transcriptional changes in response to such stresses have been described in the literature (Comerford et al., 2002; Leonard et al., 2003; Ramsborg et al., 2004; Tan et al., 2003), and such information may provide perspective in the interpretation of our own data. Compared to the CG1940 reference RNA, over 10% of the probes in the hypoxic lot of CG1940 cells fell outside the twofold differential-expression window and over 3% were outside the threefold window (Table I and Fig. 3E). In the serum-starved CG8711 cells, over 3% of genes were outside the twofold window, and 0.61% outside the threefold window based on ERMI analysis compared to the CG8711 reference RNA (Table II and Fig. 4E). The significantly larger number of differentially expressed genes and large ratios of differential expression in the stressed samples demonstrates the ability of the microarray technology to identify changes in bioprocessing conditions, and therefore the potential of the technology for process/product characterization and even quality control. We analyzed these limited sets of data by Gene Ontology analysis as described in Materials and Methods Section, and found them to be ontologically similar to previous literature reports (data not shown).

Comparison of Cells From 50 L Versus 500 L Processes by Microarray Analysis

Thus far, our studies have focused on assessing expression patterns that characterize the "same" sample, very similar samples, and very different cell samples. Our second goal was to assess the impact of the scale of the bioprocess: comparing cells cultured in 500 L bioreactors to cells cultured in 50 L bioreactors, manufactured under similar conditions. For the 500 L CG1940 cells (P500), ERMI analysis showed that 0.59% of probes fell outside the twofold differential-expression window, and 0.08% fell outside of the threefold window (Table I and Fig. 3F). For the CG8711 L500 cell lot, ERMI analysis showed 0.29% and 0.04% of probes fell outside the two- and threefold windows, respectively (Table II and Fig. 4F). For both cell lines, these fractions are extremely small, but somewhat higher than the fractions of "different" probes found in the comparison of the corresponding 50 L cell lots. Thus, one could argue that the microarray analysis does detect slightly different cells due to the use of different process scales, but in order to assign a biological significance to such differences, these gene expression changes must be repeatable in several sets of biological experiments. Such data are very hard to come by due to the limited availability of many large-scale manufacturing lots. If such data become available and microarray analysis shows a consistent pattern of differentially

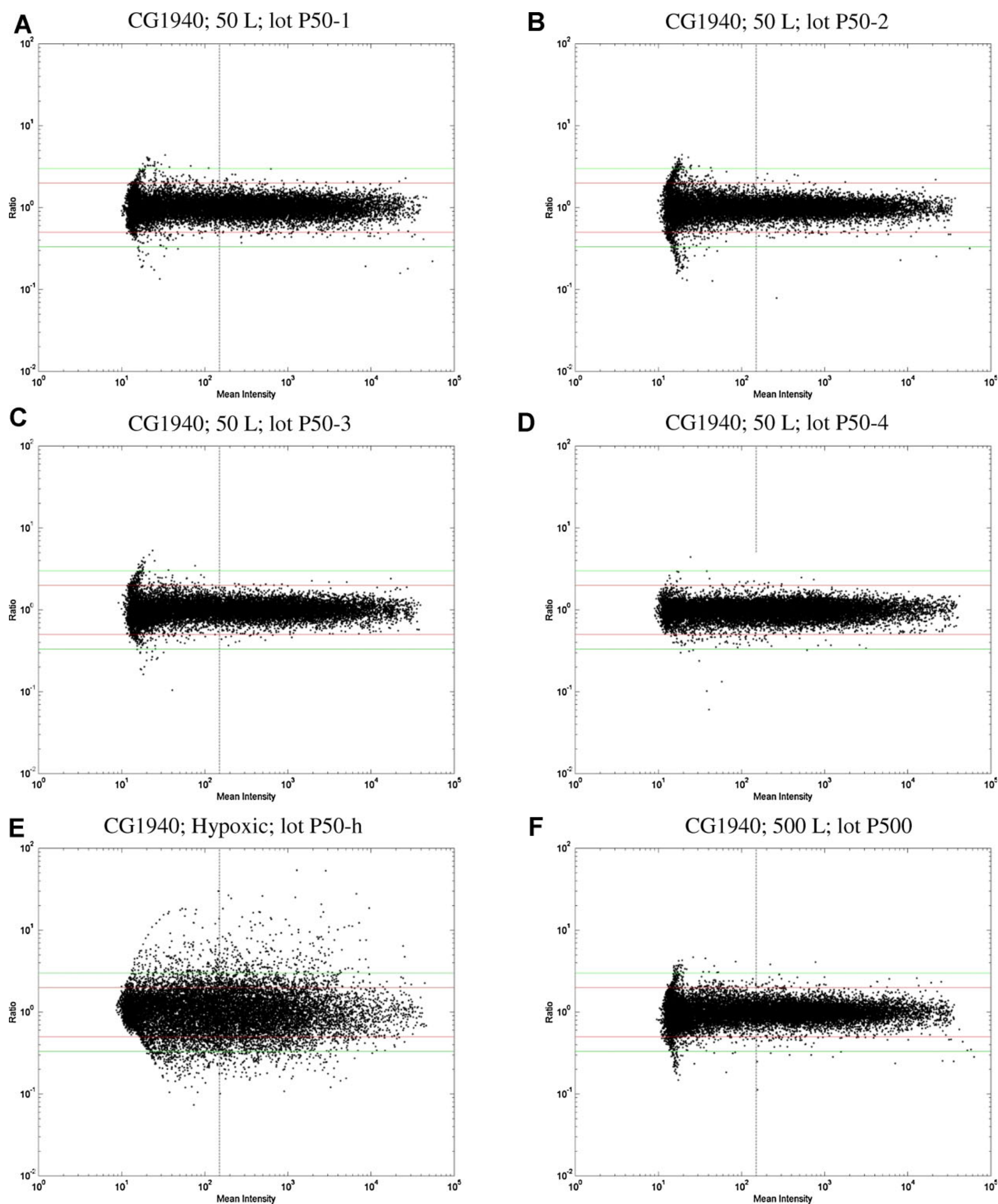

Figure 3. ERMI plots for the CG1940 cell line. Averaged expression ratios (*y* axis) as a function of the mean intensity (*x* axis) for all dye-swaps and technical replicates (i.e., 4 microarrays total for each plot). The twofold expression window is shown by horizontal lines at 1/2 and 2. The threefold window is shown by horizontal lines at 1/3 and 3. The mean intensity cut-off of 150 is shown by a dashed vertical line. **A**: Lot P50-1, 50 L; (**B**) lot P50-2, 50 L; (**C**) lot P50-3, 50 L; (**D**) lot P50-4, 50 L; (**E**) lot P50-h, hypoxic; and (**F**) lot P500, 500 L. [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

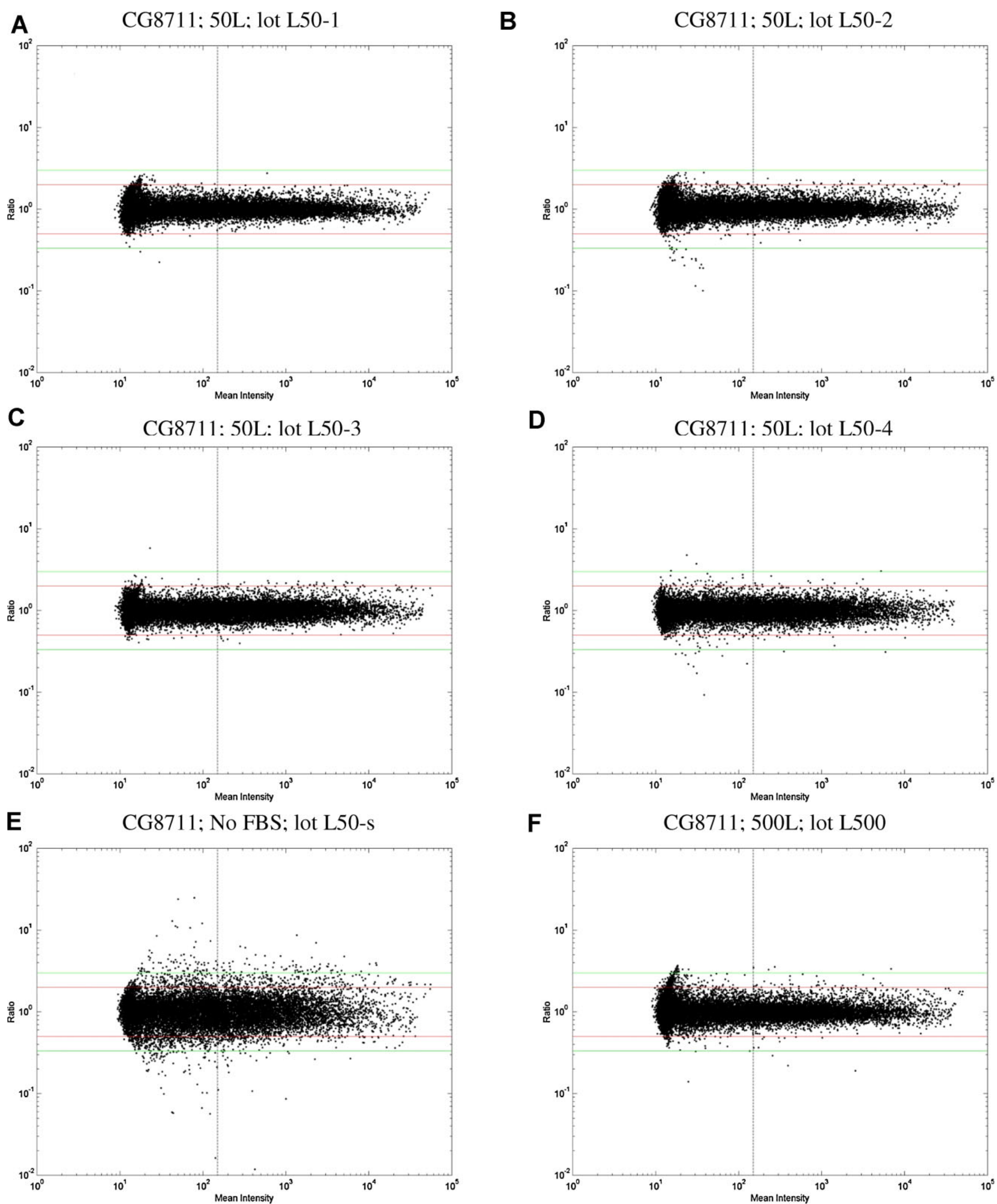

Figure 4. ERMI plots for the CG8711 cell line. Averaged expression ratios (*y* axis) as a function of the mean intensity (*x* axis) for all dye-swaps and technical replicates (i.e., 4 microarrays total for each plot). The twofold expression window is shown by horizontal lines at 1/2 and 2. The threefold window is shown by horizontal lines at 1/3 and 3. The mean intensity cut-off of 150 is shown by a dashed vertical line. **A**: Lot L50-1, 50 L; (**B**) lot L50-2, 50 L; (**C**) lot L50-3, 50 L; (**D**) lot L50-4, 50 L; (**E**) lot L50-s, No FBS; (**F**) lot L500, 500 L. [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

Table I. ERMI analysis of lots of the CG1940 cell line hybridized against its own reference RN pool.

Lot	Description	Normalized probes in all slides	Probes with averaged mean intensity >150	Inside 1/2—2 window	Outside 1/2—2 window	Inside 1/3—3 window	Outside 1/3—3 window	Erroneous probes outside 1/2—2 window	Erroneous probes outside 1/3—3 window
P50-1	50 L	16,635	8,949	8,883	66 (0.40%)	8,945	4 (0.02%)	99	50
P50-2	50 L	16,401	8,479	8,435	44 (0.27%)	8,475	4 (0.02%)	99	50
P50-3	50 L	16,645	8,860	8,838	22 (0.13%)	8,860	0 (0%)	99	50
P50-4	50 L	15,710	8,065	8,040	25 (0.16%)	8,064	1 (0.01%)	99	50
P50-h	Hypoxia	15,051	6,854	5,351	1,503 (10.00%)	6,391	463 (3.10%)	99	50
P500	500 L	16,490	8,520	8,423	97 (0.59%)	8,507	13 (0.08%)	100	52
P50-2	50 L, 44 K arrays	35,182	19,963	19,963	0 (0%)	19,963	0 (0%)	2	1

A gene with an expression ratio (sample mean intensity/reference mean intensity) of >0.5 but <2 is considered "Inside the 1/2—2 (expression) Window." Likewise, genes with expression ratios <0.5 or >2 are counted as "Outside the 1/2—2 Window." Data are also shown for an expression window spanning from 1/3 to 3. "Erroneous probes" are genes either all up-regulated or all-down regulated at least twofold in all samples of the reference-pool when individually hybridized against the reference-pool.

expressed genes, follow up studies would be warranted to assess the impact of the changes through protein level and functional assays.

Assessing a Possible Impact of the Microarray Size (Probe Numbers)

Agilent updated the "22K Human 1A Oligo Microarray" which contained probes for approximately 18,500 genes and transcripts with the "44K Whole Human Genome Microarray," which represents over 41,000 genes and transcripts. Two 50 L lots previously analyzed with the smaller microarrays, one each of the CG1940 and CG8711, were analyzed again against the corresponding reference RNA using the 44K microarray. In the 44K experiments, RNA of all the samples was extracted and prepared fresh from the cells of the corresponding lots. The same or updated commercial kits for RNA extraction, hybridization and microarray experiments were used, strictly following the manufacturer's protocols, and the results from the ERMI analyses are shown in Tables I and II and Figure 5A and B. Despite the fact that the number of probes (spots) per array doubled, not a single probe (gene) for either cell line fell outside the twofold differential-expression window. In view of the fact that these experiments were done 17 months after the previous set of experiments, with different RNA extractions from the same cell lots, these data suggest that the outcomes from this analysis are very robust, and thus suitable for process characterization.

Can Standard Clustering Techniques be Used to Visualize the Similarity or Differences Among Tested Lots/Samples?

There are many clustering methods used for analyzing and visualizing microarray data, such as hierarchical, self-organizing maps (SOMs) and K-means, all of which we have extensively used in our lab (e.g., Fuhrken et al., 2007; Tomas et al., 2003). Objectively, these methods would not be able to accurately segregate the similar lots (the normal 50 L lots or even the 500 L lots) simply because the differences are statistically minute as shown in Tables I and II, although they should be able to identify as outliers the "aberrant" lots P50-h (Table I) and L50-s (Table II). Furthermore, based on the principles underlying all these clustering methods, clustering is likely to pick up differences among lots related to how many useful probes each lot has rather than any differences among the common-probe values. This was confirmed (data not shown) by computational experiments using all data points for all lots without

Table II. ERMI analysis of lots of the CG8711 cell line hybridized against the reference-pool.

Lot	Description	Normalized probes in all slides	Probes with averaged mean intensity >150	Inside 1/2—2 window	Outside 1/2—2 window	Inside 1/3—3 window	Outside 1/3—3 window	Erroneous probes outside 1/2—2 window	Erroneous probes outside 1/3—3 window
L50-1	50 L	16,206	7,259	7,258	1 (0.01%)	7,259	0 (0%)	2	0
L50-2	50 L	16,601	7,445	7,417	28 (0.17%)	7,445	0 (0%)	2	0
L50-3	50 L	16,942	7,734	7,702	32 (0.19%)	7,734	0 (0%)	2	0
L50-4	50 L	15,008	6,767	6,742	25 (0.17%)	6,764	3 (0.02%)	2	0
L50-s	No FBS	15,129	6,340	5,838	502 (3.32%)	6,248	92 (0.61%)	2	0
L500	500 L	16,241	7,194	7,147	47 (0.29%)	7,187	7 (0.04%)	2	0
L50-1	50 L, 44 K arrays	35,966	20,224	20,224	0 (0%)	20,224	0 (0%)	31	5

A gene with an expression ratio (sample mean intensity/reference mean intensity) of >0.5 but <2 is considered "Inside the 1/2—2 (expression) Window." Likewise, genes with expression ratios <0.5 or >2 are counted as "Outside the 1/2—2 Window." Data are also shown for an expression window spanning from 1/3 to 3. "Erroneous probes" are genes either all up-regulated or all-down regulated at least twofold in all samples of the reference-pool when individually hybridized against the reference-pool.

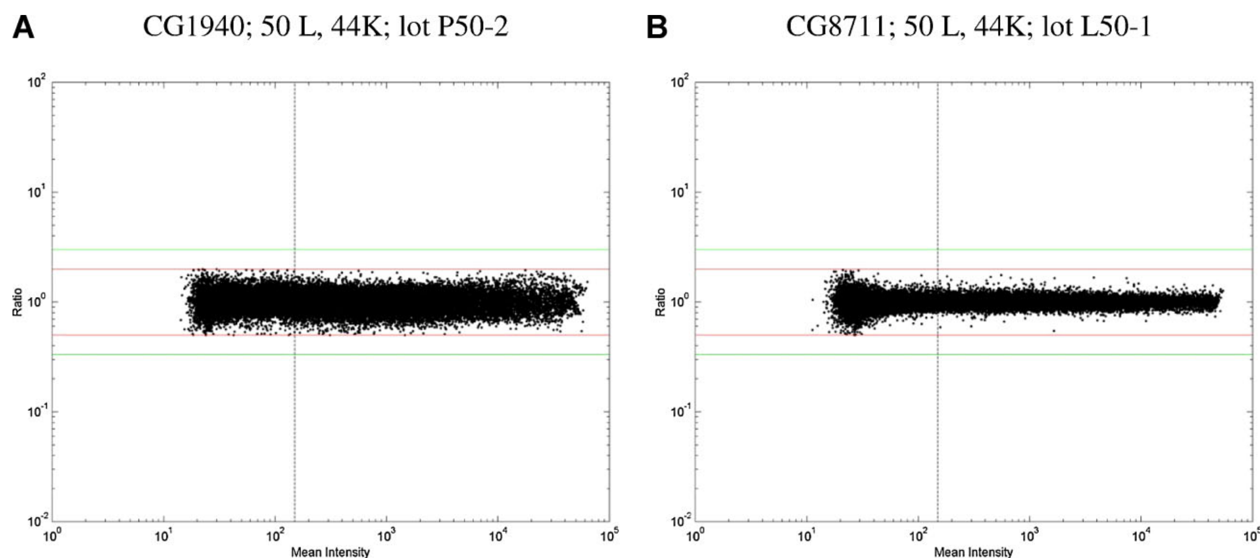

Figure 5. ERMI plots for the CG1940 and CG8711 cell lines using the Agilent 44K Whole Human Genome Microarrays. Averaged expression ratios (*y* axis) are plotted as a function of the mean intensity (*x* axis) for all dye-swaps and technical replicates (i.e., 4 microarrays total for each plot). The twofold expression window is shown by horizontal lines at 1/2 and 2. The threefold window is shown by horizontal lines at 1/3 and 3. The mean intensity cut-off of 150 is shown by a dashed vertical line. **A**: Lot P50-2, 50 L; and (**B**) lot L50-1, 50 L. [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

regard for how many common values for the microarray probes each lots has. When we carried out hierarchical clustering of the data separately for the CG1940 (PC-3) cells line samples (Table I), and for the CG8711 (LNCaP) cell line samples (Table II), clustering was able to segregate nicely sample P50-h, but it also segregated the normal, good sample P50-4, simply because of the smaller number (compared to the rest of the normal lots) of probes with available data in this sample (P50-4). Similarly for the L lots of Table II. Thus, we carried out a one-way hierarchical clustering (for lot #) of the microarray data separately for each cell line after removing the data on probes that had intensity <150 as shown in Tables I and II. Furthermore, in order to eliminate the clustering bias due to the different number of probes with usable values, we only used the probes/genes for which values exist for all lots clustered for each cell line. Thus, there were 4,105 usable probes/genes in the CG1940 (PC-3) cell line (among at least over 6,800 probes with values over 150 in each lot; Table I) and 3625 probes in CG8711 (LNCaP) cell line (among at least over 6,300 probes with values over 150 in each lot; Table II). The data are shown in Figure 6. Clustering of the genes is not informative for the purpose of this manuscript and is not shown. The dendrograms of the two clusters (Fig. 6) show that the "aberrant" lots (P50-h and L50-s) are indeed classified as outliers. Lots 50-1, 50-2, and 50-3 in each cell line used to prepare the reference RNA cluster close together, but so do the large scale lots (P500 or L500), and in fact these large-scale lots seem to cluster closer to the aforementioned corresponding reference-RNA lots than the other small scale lot P50-4 or L50-4. Statistically, once one imposes the requirement that all lots/samples contain values for the probes/genes used for clustering, all "normal" lots are so similar that they look virtually identical according to this clustering algorithm. Thus, although some are classified as a little closer than others, the classification does not capture the more refined assessment of "similar" versus "different" that is captured by the previous analysis as summarized in Tables I and II.

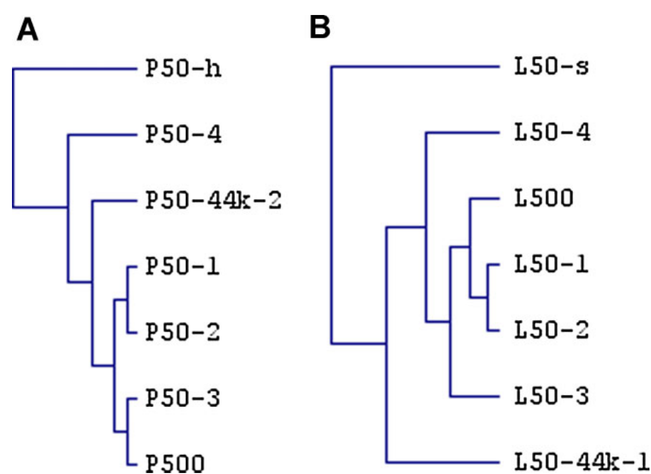

Figure 6. One way hierarchical clustering of culture lots based on the microarray data separately for each cell line. **A**: CG1940 (PC-3) cells line lots (Table I). **B**: The CG8711 (LNCaP) cell line lots (Table II). [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

A Flow Chart Summarizing the Proposed Process Characterization Process

The proposed microarray-based characterization of individual samples and/or cell lines is shown in Figure 7. The process is initiated through the identification of dye bias and background effects in the microarray system used for the analysis (track 1). A reference pool consisting of samples of the same cell line from the range of normal operating conditions is required, and this procedure can be performed for one or multiple cell lines. Experimental protocols (e.g., labeling and hybridization) and data analysis techniques (e.g., SNNLERM, ERMI, and *t*-test) have been presented above. The resulting values related to dye bias and background intensity are systematically fed into all data analyses to follow (tracks 2 or 3). The process in track 1 enables accuracy in the direct comparison of cell lines (track 2 of Fig. 7), and/or the identification of lot/sample similarity or outliers (track 3 of Fig. 7). For these applications, it is first determined if the lots or samples compared are the same, similar or different. If they are assessed as different, lists of differentially expressed genes can be generated and can be further analyzed such as by Gene Ontology analysis. These results may be used to reveal differentially expressed cellular processes, pathways, or genes delegated to specific organelles.

Discussion

The culture environment imparts complex and poorly understood effects on cells. It is now established that even sublethal mechanical forces in bioreactors may trigger alterations in the expression of surface proteins (McDowell and Papoutsakis, 1998; McDowell et al., 1998), while small changes in other process parameters affected by mixing, such as oxygen tension and pH, may alter cell proliferation

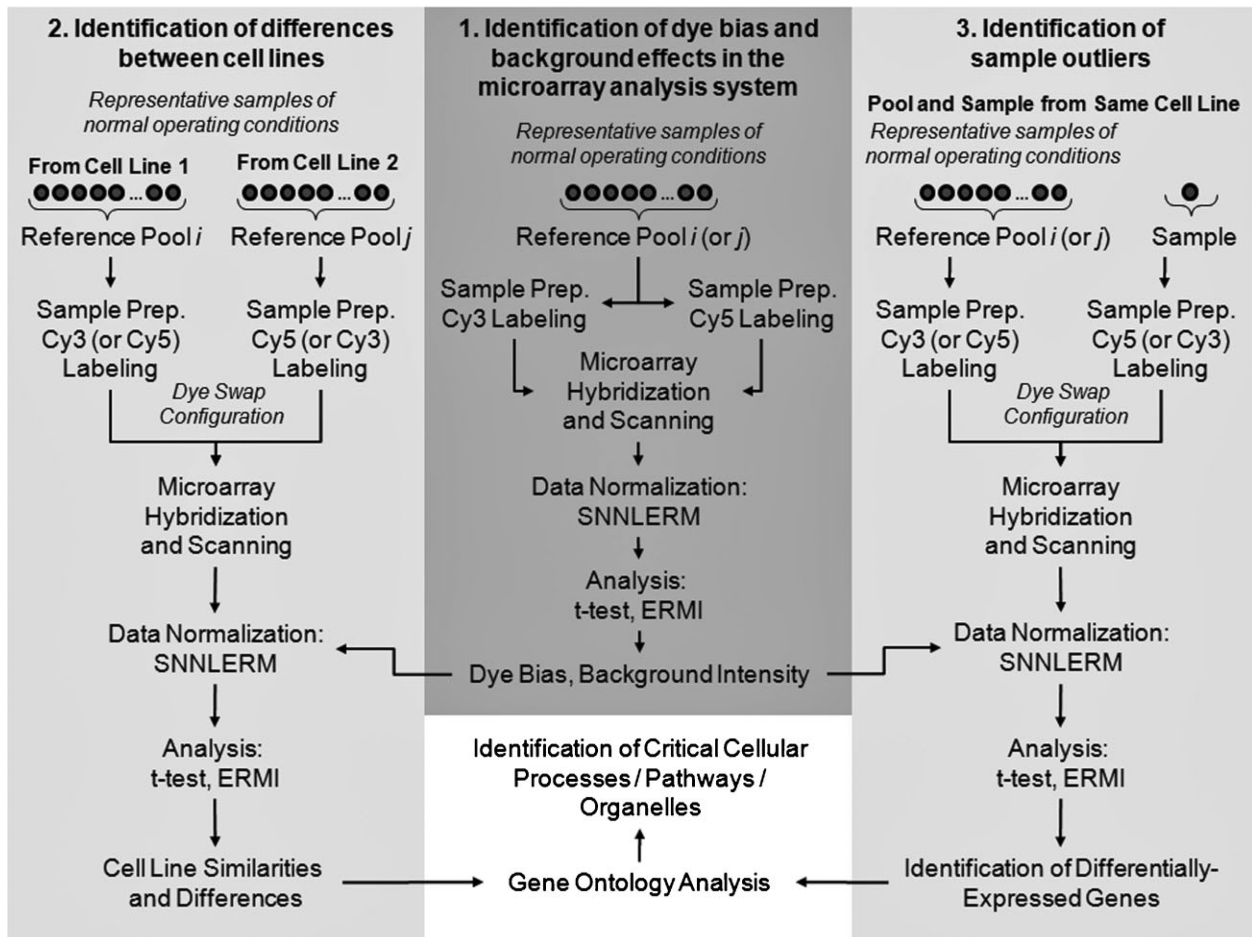

Figure 7. Flow diagram for the microarray-based procedure in order to identify significant differences in culture lots or samples. Individual lots/samples are represented as shaded circles. There are three parallel at the top tracks, which are linked at the bottom. In track 1, the process is initiated with the identification of dye bias and background effects, and this information is fed into tracks 2 and 3. Track 2 is to identify differences between different cell lines or types. Track 3 is the core track for identifying lot/sample differences or outliers as in the case of lots produced under different manufacturing conditions.

and differentiation (e.g., Hevehan et al., 2000; McAdams et al., 1997; Mostafa et al., 2001). DNA-microarray analysis is very well suited for detecting the impact of such and other bioprocessing effects in cell culture biotechnology due to its exquisite sensitivity, reproducibility, and global genome-scale nature.

In this study, we assessed the usefulness and sensitivity of microarray-based gene expression analysis in product characterization and scale-up validation for cell-therapy products using two different human cell lines that are being evaluated as prostate cancer immunotherapies. Gene expression profile demonstrated the exceptional capability and high sensitivity of microarrays in distinguishing samples that are the "same," "very similar," "very different," and "similar." Representative cell lots produced at the same scale and under the same bioprocessing conditions were deemed virtually identical, with <0.40% "differentially expressed" probes (genes) (Tables I and II) among over 18,500 probes. The differences between the two cell lines were profound (Fig. 2), and so were the differences between the hypoxically stressed CG1940 cells and the non-stressed cells. The differences between the FBS-starved CG8711 cells and non-stressed cells were also large, but smaller than those that capture the impact of hypoxia on the CG1940 cells. Furthermore, Gene Ontology analysis identified cellular programs and themes consistent with known response to hypoxia stress and serum starvation respectively.

In the comparison of the two different process scales (50 L vs. 500 L), gene expression analysis showed very small changes (<0.59% differentially expressed genes), which are of the same order of magnitude as the changes assessed for the 50 L lots (<0.40% differentially expressed genes), for both cell lines. One would conclude that this process has been successfully scaled from 50 L to 500 L without apparently altering the properties of the cells as assessed by over 18,000 gene probes. Finally, the similar results obtained when using either the 22K or the 44K microarrays further validated the robustness of this microarray technology in the characterization and process validation of these two whole-cell products.

An obvious question to raise is if this approach can be extended to characterizing other cell culture processes such as those employed for the production of therapeutic proteins, and if so what type of arrays might one use for that case, given the fact that most such processes employ Chinese Hamster Ovary (CHO) cells, which lack for now a genome sequence. To the extent that process characterization at the level of accuracy examined in this study is necessary for CHO processes, then a similar approach would be immediately applicable based on the expectation that if there are any cellular changes that would impact product quality, then such changes would be picked up by the sensitive microarray analysis. In the absence of full genomic CHO microarrays, one could use genomic-scale mouse, rat or even human microarrays since in this case all one cares about is if the hybridization pattern is "identical" or "different" between samples to be compared rather than identifying specific genes which might be differentially expressed.

Conclusion and Recommendation

Based on the data and analysis presented, and since it is not possible to determine a priori the acceptance criteria for all genes, the results from DNA microarray characterization study were used to identify any statistically significant transcriptional changes after process scale up. For the DNA microarray assay, differentially expressed genes were defined by expression ratios outside of the assay precision range of 0.5- to 2.0-fold of reference and confirmed to be statistically significant by a two sided *t*-test with unequal variance. The differentially expressed genes were identified for each lot produced by the standard method and for each lot produced by the scaled-up method. If genes were/are identified that are differentially expressed in one lot but not in the others, then they should be attributed to lot-to-lot variability of product and the random variability of the overall assay. Only if genes were/are identified that are consistently differentially expressed in the scaled-up production lots, would the data suggest that some changes in the transcription level of the product took place upon scale up or other process change. The biological significance of these genes and its effect on product quality could then be further explored by other methods, such as activity assays, protein-level assays, etc. Here we argue that the exquisite and global, at the genome scale, microarray assay can be used a frontline assay to detect potentially important differences in the product upon scale up (or upon other implemented process modifications).

We acknowledge the assistance by Dr. Yili Chen for the data clustering studies.

References

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25(1):25–29.

Borden JR, Paredes CJ, Papoutsakis ET. 2005. Diffusion, mixing, and associated dye effects in DNA-microarray hybridizations. Biophys J 89(5):3277–3284.

Comerford KM, Wallace TJ, Karhausen J, Louis NA, Montalto MC, Colgan SP. 2002. Hypoxia-inducible factor-1-dependent regulation of the multidrug resistance (MDR1) gene. Cancer Res 62(12):3387–3394.

Copier J, Dalgleish A. 2006. Overview of tumor cell-based vaccines. Int Rev Immunol 25(5–6):297–319.

Fuhrken PG, Chen C, Miller WA, Papoutsakis ET. 2007. Comparative, genome-scale transcriptional analysis of CHRF-288-11 and primary human megakaryocytic cell cultures provides novel insights into lineage-specific differentiation. Exp Hematol 35(3):476–489.

Guinn BA, Kasahara N, Farzaneh F, Habib NA, Norris JS, Deisseroth AB. 2007. Recent advances and current challenges in tumor immunology and immunotherapy. Mol Ther 15(6):1065–1071.

Guinn B, Casey G, Collins S, O'Brien T, Alexander Y, Tangney M. 2008. Tripartite meeting in gene and cell therapy: Irish society for gene and

cell therapy, British society for gene therapy & international society for the cell and gene therapy of cancer 2008. Hum Gene Ther 19(10):967–978.

Haddad H, Windgassen D, Ramsborg CG, Paredes CJ, Papoutsakis ET. 2004. Molecular understanding of oxygen tension and patient-variability effects on ex vivo expanded T cells. Biotechnol Bioeng 87(4):437–450.

Hevehan DL, Papoutsakis ET, Miller WM. 2000. Physiologically significant effects of pH and oxygen tension on granulopoiesis. Exp Hematol 28(3):267–275.

Horoszewicz JS, Leong SS, Kawinski E, Karr JP, Rosenthal H, Chu TM, Mirand EA, Murphy GP. 1983. Lncap model of human prostatic-carcinoma. Cancer Res 43(4):1809–1818.

Hosack DA, Dennis G, Jr., Sherman BT, Lane HC, Lempicki RA. 2003. Identifying biological themes within lists of genes with EASE. Genome Biol 4(10):R70.

Kaighn ME, Narayan KS, Ohnuki Y, Lechner JF, Jones LW. 1979. Establishment and characterization of a human prostatic-carcinoma cell-line (PC-3). Investig Urol 17(1):16–23.

Leonard MO, Cottell DC, Godson C, Brady HR, Taylor CT. 2003. The role of HIF-1 alpha in transcriptional regulation of the proximal tubular epithelial cell response to hypoxia. J Biol Chem 278(41):40296–40304.

McAdams TA, Miller WM, Papoutsakis ET. 1997. Variations in culture pH affect the cloning efficiency and differentiation of progenitor cells in ex vivo haemopoiesis. Br J Haematol 97(4):889–895.

McDowell CL, Papoutsakis ET. 1998. Increased agitation intensity increases CD13 receptor surface content and mRNA levels, and alters the metabolism of HL60 cells cultured in stirred tank bioreactors. Biotechnol Bioeng 60(2):239–250.

McDowell CL, Carver RT, Papoutsakis ET. 1998. Effects of Methocel A15LV, polyethylene glycol, and poryvinyl alcohol on CD13 and CD33 receptor surface content and metabolism of HL60 cells cultured in stirred tank bioreactors. Biotechnol Bioeng 60(2):251–258.

Mostafa SS, Papoutsakis ET, Miller WM. 2001. Oxygen tension modulates the expression of cytokine receptors, transcription factors, and lineage-specific markers in cultured human megakaryocytes. Exp Hematol 29(7):873–883.

Paredes CJ, Senger RS, Spath IS, Borden JR, Sillers R, Papoutsakis ET. 2007. A general framework for designing and validating oligomer-based DNA microarrays and its application to *Clostridium acetobutylicum*. Appl Environ Microbiol 73(14):4631–4638.

Ramsborg CG, Papoutsakis ET. 2007. Global transcriptional analysis delineates the differential inflammatory response interleukin-15 elicits from cultured human T cells. Exp Hematol 35(3):454–464.

Ramsborg CG, Windgassen D, Fallon JK, Paredes CJ, Papoutsakis ET. 2004. Molecular insights into the pleiotropic effects of plasma on ex vivo-expanded T cells using DNA-microarray analysis. Exp Hematol 32(10):970–990.

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. 2003. TM4: A free, open-source system for microarray data management and analysis. Biotechniques 34(2):374–378.

Tan PK, Downey TJ, Spitznagel EL, Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC. 2003. Evaluation of gene expression measurements from commercial microarray platforms. Nucleic Acids Res 31(19):5676–5684.

Tomas CA, Alsaker KV, Bonarius HPJ, Hendriksen WT, Yang H, Beamish JA, Paredes CJ, Papoutsakis ET. 2003. DNA array-based transcriptional analysis of asporogenous, nonsolventogenic Clostridium acetobutylicum strains SKO1 and M5. J Bacteriol 185(15):4539–4547.

Ward JE, McNeel DG. 2007. GVAX: An allogeneic, whole-cell, GM-CSF-secreting cellular immunotherapy for the treatment of prostate cancer. Expert Opin Biol Ther 7(12):1893–1902.

Yang H, Haddad H, Tomas C, Alsaker K, Papoutsakis ET. 2003. A segmental nearest neighbor normalization and gene identification method gives superior results for DNA-array analysis. Proc Natl Acad Sci USA 100(3):1122–1127.