Background

Expression profiling is an emerging experimental method whereby RNA accumulation in cells and tissues can be assayed for many thousands of genes simultaneously in a single experiment. There are two common experimental platforms for expression profiling; redundant oligonucleotide arrays (Affymetrix GeneChips) [1], and spotted cDNA microarrays [2–4]. The Affymetrix GeneChips have the inherent advantages of redundancy, specificity, and transportability; there are typically 30–40 oligonucleotide...
Expression profiling has led to dramatic advances in understanding of yeast biology, where homogeneous cultures can be grown and exposed to timed environmental variables [7–12]. Such studies have led to the rapid assignment of function to a large number of anonymous gene sequences. Large-scale expression profiling studies of tissues from higher vertebrates are more challenging, due to the higher complexity of the genome, larger related gene families, and incomplete genomic resources. Nevertheless, DNA microarrays have been successfully applied in the analysis of aging and caloric restriction [13] and pulmonary fibrosis [14]. And many publications, particularly on cancer, have appeared [14–19]. Affymetrix has recently announced the availability of the U133 GeneChip series with 33,000 well-characterized human genes mined from genomic sequence. The nearly complete ascertainment of genes in the human genome should make expression-profiling studies of human tissues particularly powerful. However, identification of the sources of experimental variability, and knowledge of the relative contribution of variation from each source, is critical for appropriate experimental design in expression profiling experiments. Mills and Gordon recently studied the relative contribution of experimental variability of probe production on the reproducibility of microarray results using mixed murine tissue RNA on Affymetrix Mu11K GeneChips [20]. In their study, the same RNA preparation was used as a template for distinct cDNA/cRNA amplifications and hybridizations. An additional variable studied was the effect of different laboratories processing the same RNAs. The authors found relatively poor concordance between duplicate arrays, with an average of 12% increase/decrease calls between the same RNA processed in parallel and hybridized to two Mu11K-A microarrays. The authors concluded that there was substantial experimental variability in the experimental procedure, necessitating extensive filtering and large numbers of arrays to detect accurate gene expression changes (LUT: look-up tables) [20]. In our laboratory, we have processed over 1,200 Affymetrix arrays, and have found significantly higher experimental reproducibility ($R^2 = 0.979$ for new generation U74A version 2 murine arrays or human U95 series, see Result and Discussion). In addition, a recent publication of a single human patient, where RNA was prepared from two distinct breast tumors, and placed on duplicate U95A GeneChips (four chips total) found a very low degree of experimental variability between microarrays ($R^2 = 0.995$), and between the two tumors ($R^2 = 0.987$) [21]. The marked differences in experimental variability between laboratories could be due to different quality control protocols (see [http://microarray.cnmresearch.org]), newer more robust Affymetrix arrays now available (murine Mu11K versus U74A version 2 and new generation human U95 series), use of more recent algorithms for data interpretation, or due to more consistent processing of RNA, cDNA, and cRNA in the same laboratory.

The previous studies did not systematically address the reproducibility of GeneChip hybridization (e.g. the same biotinylated cRNA on two different microarrays). In addition to lingering questions concerning variability due to specific experimental procedures, there are other possible sources of variability that have not yet been investigated, specifically tissue heterogeneity and inter-individual variation. The latter two sources of variability are particularly important in human expression profiling studies. The study of human tissues often involves the use of tissue biopsies, where a relatively limited region of an organ is sampled. Tissue heterogeneity and sampling error might be expected to introduce significant variability in expression profiles. Second, tissues may derive from individuals from different ethnic backgrounds; humans are highly outbred, leading to the potential of significant polymorphic noise (herein called "SNP noise") between individuals unrelated to the disease or variable under study. SNP noise also exists between different inbred mouse strains, and some experiments have normalized this effect by breeding the same mutation on different strains, and profiling each individually [22]. Knowledge of the relative effect of each experimental, tissue, and patient variable on expression profiling results in humans is important, so that appropriate experimental designs can be employed.

We recently reported the design and production of a highly redundant oligonucleotide microarray for analysis of human muscle biopsies (Borup et al. submitted). This MuscleChip contains 4,601 probe sets corresponding to 3,369 distinct genes and ESTs expressed in human muscle. Each probe set contains between 16 to 40 oligonucleotides, such that the number of specific oligonucleotide probes on the array was 138,000.

Here, we utilize this MuscleChip to investigate the relative significance of variables affecting expression profiling data and interpretation. Specifically, we studied the correlation coefficients of profiles considering the following variables: 1. variation due to probe production (same RNA); 2. variation due to the microarray itself (same cRNA on different GeneChips); 3. tissue heterogeneity (different regions of the same muscle biopsy); 4. inter-patient variability (SNP noise); 5. diagnosis (underlying pathological variable); and 6. patient age.
Table 1: Patient data and characteristics of 34 MuscleChip expression profiles.

| Patients/ Arrays | Individual or Mixed | Age (years) | Stage of histopathology | Scaling Factor | % Present Calls (%) | % Diff Calls Paired samples | Four comparisons > 2-fold changes relative to controls 1a, 1b |
|------------------|---------------------|-------------|--------------------------|----------------|---------------------|-----------------------------|-------------------------------------------------------------|
| 1a               | Ind                 | 5.5         | mid                      | 1.08           | 0.9                 | 51                          | 50                           | 5.6 302                                                   |
| 1b               |                     |             |                          |                |                     |                             |                              |                                                             |
| 2a               | Ind                 | 4.5         | early stage              | 1.93           | 3.28                | 38                          | 36                           | 3 427                                                   |
| 2b               |                     |             |                          |                |                     |                             |                              |                                                             |
| 3a               | Ind                 | 5           | early/mid stage          | 0.53           | 0.97                | 49                          | 44                           | 8.6 312                                                   |
| 3b               |                     |             |                          |                |                     |                             |                              |                                                             |
| 3a-dup           | Ind                 | 5           | early/mid stage          | 0.46           | 0.82                | 50                          | 46                           | 0.8 N/A                                                  |
| 3b-dup           |                     |             |                          |                |                     |                             |                              |                                                             |
| 4a               | Ind                 | 6           | mid                      | 0.78           | 1.02                | 51                          | 50                           | 18 324                                                   |
| 4b               |                     |             |                          |                |                     |                             |                              |                                                             |
| 5a               | Ind                 | 5           | early                    | 1.26           | 2.13                | 51                          | 40                           | 6.2 453                                                   |
| 5b               |                     |             |                          |                |                     |                             |                              |                                                             |
| 6a               | Ind                 | 12          | mid                      | 2.72           | 2.46                | 35                          | 38                           | 1.5 305                                                   |
| 6b               |                     |             |                          |                |                     |                             |                              |                                                             |
| 7a               | Ind                 | 11          | mid/moderate             | 0.79           | 0.9                 | 51                          | 49                           | 2.6 356                                                   |
| 7b               |                     |             |                          |                |                     |                             |                              |                                                             |
| 8a               | Ind                 | 10          | mid                      | 1.24           | 1.74                | 49                          | 50                           | 1.5 463                                                   |
| 8b               |                     |             |                          |                |                     |                             |                              |                                                             |
| 9a               | Ind                 | 11          | variable                 | 1.04           | 1.37                | 50                          | 46                           | 8.5 266                                                   |
| 9b               |                     |             |                          |                |                     |                             |                              |                                                             |
| 10a              | Ind                 | 10          | mid/late                 | 0.48           | 0.75                | 52                          | 48                           | 1.5 250                                                   |
| 10b              |                     |             |                          |                |                     |                             |                              |                                                             |
| 6-9mix-1a        | mix                 | 6 to 9      | 5 patients biopsies, cRNAs mixed | 0.82 | 0.84                | 49                          | 3.2                           | 389                                                   |
| 6-9mix-1b        | mix                 |             |                          |                |                     |                             |                              |                                                             |
| 5-6mix-1a        | mix                 | 5 to 6      | 5 patients biopsies, cRNAs mixed | 0.67 | 0.67                | 60                          | 0.5                           | 58 486                                                   |
| 5-6mix-1b        | mix                 |             |                          |                |                     |                             |                              |                                                             |
We have recently reported generation of expression profiling results using mixed patient samples [23]. Our hypothesis was that mixing of RNA samples from multiple regions of muscle biopsies, and from multiple patients matched for most variables (disease, age, sex), would effectively normalize both intra-patient variability (tissue heterogeneity), and inter-patient variability (SNP noise; e.g. normal human polymorphic variation unrelated to the primary defect). Here, we test this hypothesis directly, and show that sample mixing does indeed result in relatively high sensitivity and specificity for gene expression changes that would be detected by many individual expression profiles. Thus, sample mixing appears to be an appropriate first-pass method to obtain the most significant expression changes, while using small numbers of arrays.

**Results and discussion**
Fifty six (56) different RNA samples were prepared from different regions of muscle biopsies from 28 individuals (15 Duchenne muscular dystrophy (DMD) patients, 13 normal controls). The profiles of five of the DMD patients and the five controls have been previously reported using the Affymetrix HuFL microarray [23]; however, we re-tested these same samples on the custom MuscleChip (Borup et al. *submitted*) for comparison to the other patients here. All RNAs were converted to double-stranded cDNA, and then to biotinylated cRNA. The cRNAs were then hybridized to the MuscleChip either singly, in mixed groups, or both, as described below. In total, 34 hybridizations were performed, scanned, and the data statistically analyzed using Affymetrix Microarray Suite and Excel. Quality control criteria were as described on our web site ([http://microarray.cnmcresearch.org](http://microarray.cnmcresearch.org)), and included sufficient cRNA amplification, and adequate post-hybridization scaling factors. Scaling factors (normalization needed to reach a common target intensity) ranged from 0.46 to 3.28 (Table 1). All raw image files, processed image files, and difference analyses are posted on a web-queried SQL database interface to our Affymetrix LIMS Oracle warehouse (see [http://microarray.cnmcresearch.org](http://microarray.cnmcresearch.org): link to "programs in genomic applications", "data", "human").

Among the 4,601 probe sets on the Affymetrix custom muscle microarray, we found a consistent percentage of "present" calls for each of the 34 cRNA samples tested (Duchenne dystrophy, 28 arrays, 48.2% ± 6.1%; controls 6 arrays, 53.3% ± 1.4%). To test for inter-array variability, two different hybridization solutions were applied to duplicate arrays, and correlation coefficients determined. A high correlation coefficient was found in this analysis, suggesting that inter-array variability of the MuscleChip used was a relatively minor variable (Patient 3 a and 3a-duplicate R² = 0.96 and percent shared [No Change (NC)] calls by Microarray Suite software was 99%; Patient 3b and 3b-duplicate R² = 0.98 and percent NC was 98%; Table 1). The high reproducibility of Affymetrix array results is consistent with other data in our laboratory, and from previously published data [6,21,23,24], and shows that experimental variability associated with hybridization and scanning of highly redundant oligonucleotide GeneChips is not a major source of experimental variability.

Given the previous report suggesting that the conversion of RNA to biotinylated cRNA probe was a major source of variability in murine array experiments [20], we tested a series of murine RNA from different sources, using the

| Table 1: Patient data and characteristics of 34 MuscleChip expression profiles. (Continued) |
|----------------------------------|----------|-----------------|-----------------|----------|
| 10-12mix-1a mix                  | 10 to 12 | 5 patients biopsies, cRNAs mixed | 0.82 | 49 | 0.4 | 388 |
| 10-12mix-1b mix                  |         |                 | 1.16 | 55 |       |       |
| control 1a                       | mix     | 6 to 9          | 5 normal biopsies, cRNAs mixed | 0.71 | 53 | 1.5 | N/A |
| control 1b                       | mix     |                 | 0.48 | 54 |       |       |
| control 2a                       | mix     | 5 to 12         | 3 normal biopsies, cRNAs mixed | 0.86 | 54 | 0.8 | N/A |
| control 2b                       | mix     |                 | 0.78 | 53 |       |       |
| control 3a                       | mix     | 4 to 13         | 3 normal biopsies, cRNAs mixed | 0.95 | 51 | 1.1 | N/A |
| control 3b                       | mix     |                 | 0.78 | 55 |       |       |
Figure 1
Unsupervised hierarchical clustering of 42 murine U74Av2 Affymetrix arrays. Unsupervised hierarchical clustering of 42 murine U74Av2 Affymetrix arrays shows that probe synthesis and hybridization is not a major source of experimental variability. Expression profiles shown were from three different experimental groups; one using cultured murine myogenic cells (VSM samples), one using mouse spleens (KNagaraju samples), and one group from mouse skeletal muscle from normal and mdx mouse strains (FBooth samples). For the KNagaraju samples, the same spleen RNA was split prior to cDNA synthesis to create duplicate cDNA-cRNA-profile results; these duplicates show a very high correlation coefficient, and close relationship by Unsupervised clustering (low branches on dendrogram). The VSM cultured samples were each derived from different culture plates, and the FBooth samples from different murine muscles. The duplicate murine muscle samples are more closely related (high correlation coefficient) than the parallel cultures (VSM). Additional variables, such as male versus female (KNagaraju samples), and time after TSA treatment (VSM samples) are indicated, but are not relevant for this manuscript, and will be discussed in more detail elsewhere.
newer generation U74Av2 GeneChips. One series of samples was from murine spleens, where spleens from multiple animals for each variable under study were mixed, RNA isolated, RNA samples split, and duplicate cDNA, cRNA, and hybridizations processed in parallel for each RNA (Fig. 1, *KNagaraju* samples). We also compared RNAs processed from parallel murine myogenic cell cultures (Fig. 1, *VSM* samples), where each profile was from a different cell culture. Finally, we used a series of murine muscle tissues from normal and dystrophin-deficient mice, where each profile was from a different series of complete gastrocnemius muscles (Fig. 1, *FBooth* samples). The data from these 42 murine U74Av2 profiles were then analyzed by unsupervised clustering [25] to determine which profiles were most closely related to each other (Fig. 1). This analysis shows that the different sources of RNA cluster together, as expected. Importantly, the same RNA used as a template for two distinct cDNA/cRNA preparations and hybridizations showed a high correlation coefficient ($R^2 = 0.99$ for five of the six samples, with average $R^2 = 0.978$) (Fig. 1). The large muscle group profiles (*FBooth* samples) showed excellent correlation, both with respect to diagnosis; however here there was no sampling error as the entire muscle group was used rather than isolated biopsies. Finally, the parallel tissue culture experiments (VSM samples) showed greater variability between duplicates, suggesting that tissue culture conditions may be more subject to variability than *in vivo* tissues (Fig. 1). This murine data shows that variability from different cDNA-cRNA reactions is very low ($R^2 = 0.978$).

To analyze the impact of intra-patient variability (tissue heterogeneity), inter-patient variability (polymorphic noise in outbred populations), and the effect of sample mixing on the sensitivity of detection of gene expression differences between patient groups, we conducted a series of individual and mixed profiling (Table 1). Muscle biopsies from five 4–6 yr old DMD patients, and five 10–12 yr old patients were selected, each biopsy split into two parts, and RNA isolated independently from each of the 20 biopsy fragments. For these ten DMD patients, the two different regions of the same biopsy were expression profiled both individually (20 profiles), and also mixed into four pools where each pool originated from distinct RNA samples (Table 1). The resulting profiles were also compared to previously reported mixed 6–9 yr old DMD patient cRNAs, and mixed 6–9 yr old control cRNAs [23], as mentioned above.

As an initial statistical analysis, we used Affymetrix software to define genes that showed expression changes (Increased, Decreased or Marginal) in expression levels between pairs of profiles (difference analyses). This method of data interpretation showed that some muscle biopsies showed very little variance between different regions of the same biopsy, while other patient biopsies showed considerable variability (see Fig. 2 for representative scatter graphs). Expressing this variance as a percentage of "Diff Calls" between the two regions of the same biopsy, as determined by Affymetrix default algorithms, we found considerable variability in the similarity of profiles, with values ranging from 1.5% to 18% of the 4,601 probe sets studied (4.99% ± 4.94%). This data suggests that tissue heterogeneity (intra-patient variability) can be a major source of variation in expression profiling experiments, even when using relatively large pieces (50 mg) of relatively homogeneous tissues (such as muscle).

The most common strategy for interpreting Affymetrix microarray data is to use two profile comparisons, with an arbitrary threshold for "significant fold-change" in expression levels. Typically, multiple arrays are compared, with those gene expression changes showing the most consistent fold changes prioritized, although other methods have been reported [13,22,26,27]. To study inter-patient variability, we defined the gene expression changes surviving four pairwise comparisons with mixed control samples, as we have previously described [23]. Briefly, four comparisons were done by Affymetrix software (eg. DMD 1a versus control 1a; DMD 1a versus control 1b; DMD1b versus control 1a; DMD1b versus control 1b). The four data sets were then compared, with only those gene expression changes that showed >2-fold change in all four comparisons (four comparison survival method). The number of surviving diff calls by this method ranged from 250 to 463 (355 ± 80) (Table 1). Interestingly, those patients showing considerable variation between different regions of the same biopsy did not show a corresponding decrease in the number of gene expression changes surviving the iterative comparisons to controls (Table 1). This suggests (but does not prove) the most significant changes might be shared, independent of tissue variability (see below).

A different statistical method to determine the effect of the different variables under study is to perform hierarchical cluster analysis using nearest neighbor statistical methods [25]. Here, we subjected all profiles to unsupervised cluster analysis, as a means of determining which variables had the greatest effect (e.g. intra-patient variability [different regions of biopsy], versus diagnosis [DMD vs control], versus inter-patient variability [DMD patients in same age group], versus age of patient). For this analysis, we used the fluorescence intensity of each probe set (Average difference), after data scrubbing to remove genes that showed expression levels near background ("Absent Calls") for all profiles (Fig. 3). This analysis shows that duplicate profiles of the same cRNA hybridization solution are the most highly related (Patient 3 a and duplicate (3a-d); 3b and duplicate (3b-d)), consistent with the high correlation found by the comparisons using Affymetrix
Figure 2
Different regions of the same tissue specimen can give highly similar or highly discordant expression profiles. Shown are scatter plots of the expression profiles of two different regions of the same muscle biopsy from patient 6 (Panel A), and patient 4 (Panel B). Only "present" calls are shown (~2,000 of the 4,600 probe sets studied). An example of one patient showing very high concordance between two different biopsy regions is shown (panel A), and an example of a second patient showing very poor correlation between the two biopsy fragments (panel B). The solid lines indicate two-, three-, ten- and thirty-fold difference thresholds.
Microarray Suite software described above. Again, this reflects the low amount of combined experimental variability intrinsic to the laboratory processing of RNA, cDNA, cRNA and hybridization.

When comparing two different regions of the same biopsy [intra-patient variability], we found widely varying results, depending on the patient studied (Fig. 3). For example, some individual patients showed very closely related profiles that approached the similarity of duplicate arrays on the same cRNA (Fig. 2; profiles 6a, 6b; 10a, 10b). On the other hand, some patients showed very distantly related profiles for two regions of the same biopsy (Fig. 3; profiles 1a, 1b; 4a, 4b; 9a, 9b). Importantly, the variation caused by intra-patient tissue variation often overshadowed all other variables. For example, a profile from DMD patient 9 (9a) clustered with the normal controls, rather than with the other DMD patients (Fig. 3). The histopathology of this patient was noted as being unusually variable in severity prior to expression profiling. Also, unsupervised clustering was unable to group patients of similar ages, despite DMD showing a progressive clinical course. We conclude that intra-patient tissue heterogeneity is a major source of experimental variability in expression profiling, and must be considered in experimental design.

The above findings suggested that both intra-patient variability (tissue heterogeneity) and inter-patient variability (polymorphic noise) had major effects on the expression profiles. One method to control for these sources of noise is to analyze large numbers of profiles, both on multiple patients, and on multiple regions of tissue from each patient. This would allow determinations of p values and statistical significance for a single controlled variable under study (e.g. DMD vs controls). An alternative method is to experimentally normalize these variables through mixing of samples from patient groups; such mixing would be expected to average out both intra- and inter-patient variation. The expectation is that the most significant and dramatic gene expression changes would still be identified, while using many less profiles (and thus a substantial reduction in cost of the analyses).

To test for the relative sensitivity of interpretation of sample mixing versus individual profiles, we mixed together the 10 cRNAs for the two different age groups of DMD patients (samples 1a - 5b; samples 6a - 10b). For this analysis, we also generated expression profiles for two additional groups of control individuals. One was a second set of five normal male biopsies ages 5–12 yrs (controls 2a, 2b), and the third control set was three normal
age-matched female biopsies ages 4–13 yrs (controls 3a, 3b) (Fig. 4). As with the original male control group (control 1a, 1b), two different regions of each biopsy were processed independently through the biotinylated cRNA step, and then equimolar amounts of cRNA mixed for hybridization to the MuscleChip. All 34 profiles (both individual and mixed samples) were again analyzed by unsupervised hierarchical clustering (Fig. 4) [25]. As described above, we scrubbed the profiles to eliminate all genes showing expression levels consistently at or below background hybridization intensities by requiring each gene to show a “Present Call” in one or more of the 34 profiles.

As above, duplicate profiles using the same cRNA hybridization solution on different arrays, whether mixed or individual samples, showed very highly correlated results (very low branch on dendrogram) (Fig. 4; mix 5–6 yrs, mix 10–12 yrs; patient 3a/3a-d; patient 3b/3b-d). As above, this indicates that experimental variability from laboratory procedures or different arrays is a relatively minor factor in interpretation of results. Mixed samples from different regions of the same biopsies showed the same, or only slightly more variation (mixed controls c1, c2, and c3, mixed DMD 6–9 yrs). This showed that sample mixing does indeed average out tissue heterogeneity (intra-patient variability), as well as inter-patient variability. We noted that all of the controls (both male and female) clustered in the same branch of the dendrogram, while the four of the six mixed DMD profiles clustered just one level away from the controls, separately from the other DMD profiles. This analysis suggests that there is considerable variability in the progressive tissue pathology induced by dystrophin deficiency, both within a patient, and between patients.

To test the sensitivity and specificity of sample mixing versus individual profiling, we defined differentially expressed genes using a two group t-test (GeneSpring [28,29]), comparing all 6 mixed control profiles and the 10 individual 5–6 yr old DMD profiles. Genes were retained that met specific p value thresholds between the two sets of profiles. In parallel, we compared the two corresponding mixed 5–6 yr old DMD profiles to the 6 mixed controls.

Comparison of 10 individual 5–6 yr Duchenne dystrophy profiles to 6 mixed controls revealed 1,498 genes showing differential expression with p < 0.05 (Fig. 5). Comparison of the two mixed Duchenne dystrophy profiles to the 6 mixed controls showed 1,350 genes with p < 0.05 (Fig. 5A). Comparison of the two gene lists showed that 61% of differentially regulated genes detected by the 10 individual profiles were also detected by the two mixed profiles. This suggests that the sensitivity and specificity of using mixed samples is approximately half that of individual profiles. However, there was a rapid shift in specificity and sensitivity as stringency of the analysis was increased. Raising the statistical threshold to p < 0.0001 for individual profiles, while keeping the threshold for mixed profiles at p < 0.05 as required by the small number of data points (Fig. 5B), resulted in a sensitivity of 86% for mixed samples (351 of 408 genes p < 0.0001 detected). In conclusion, mixing detected about two thirds of statistically significant changes (p < 0.05). Mixing was a relatively sensitive method of detecting the most highly significant changes (p < 0.0001) (86% of changes detected), however it was not very specific; as many as one third of gene ex-

![Figure 5](image-url)
expression changes showing p < 0.05 in mixed samples were not confirmed by individual profiles.

Use of t-test measurements is expected to contain significant amounts of noise, due to the very large number of comparisons involved in array studies; a value of p = 0.05 means that as many as 5% of gene expression changes are expected to be identified by "chance", and thereby not reflect true differences between samples. We have previously reported a very simple, yet potentially more stringent method for data analysis of small numbers of expression profiles, using duplicate profiles for control and experimental samples, and then identifying those genes that show consistent changes >2-fold in the four possible pair-wise data comparisons (four comparison survival method) [23]. A similar pair-wise comparison method, using a less stringent average fold-change analysis, was recently reported for muscle from aging and calorie-restricted mouse muscle [13].

To investigate the validity of this approach we compared the sensitivity and specificity of t-test detection of expression changes versus the four-pairwise survival method. Two sample t-test of the 10 individual Duchenne dystrophy profiles compared to the 6 mixed control profiles revealed 1,498 genes showing p < 0.05 as above. In parallel, the mixed DMD duplicate profiles were compared to a single pair of mixed control sample profiles (c1a, c1b), using the pairwise comparison survival method [23]. Briefly, four comparisons were done (DMD 1a versus control 1a; DMD1b versus control 1a; DMD 1a versus control 1b; DMD1b versus control 1b), and only those genes retained which showed >2-fold change in all four comparisons. This method was indeed considerably more specific in identifying significant (p < 0.05) gene expression changes (Fig. 6A) with 85% of gene expression changes in the mixed profiles verified by individual profiles (p < 0.05). The sensitivity of this method depended on the p-value threshold for the individual profiles, but only reached a maximum of 49% sensitivity at p < 0.0001 (Fig. 6B).

The results above suggested that analysis of mixed samples using t-test methods was relatively sensitive but nonspecific, while analysis of the same mixed profiles by 2-fold survival method was relatively specific but insensitive. To confirm this conclusion, we directly compared the sensitivity and specificity of the four pairwise comparison method to more standard t-test methods (Fig. 7A). We found that the pairwise survival method was indeed highly specific, with 97% of changes identified by this method also detected by t-test. However, as predicted, it was not very sensitive, with only 30% of the expression changes with p < 0.05 identified by t-test being detected by the pairwise survival method. Comparison of all three analysis methods showed that many (349) genes expression changes were detected by all three methods (Fig. 7B).

Conclusions

Microarray data analyses have been criticized as being "quite elusive about measurement reproducibility" [30]. This is largely the consequence of the large number of uncontrolled or unknown variables, and the prohibitive cost of isolating and investigating each variable. Here, we report the systematic isolation and study of most variables in microarray experiments using Affymetrix oligonucleotide arrays and human tissue biopsies. We found that all sources of experimental variability were quite minor (microarray R² = 0.98–0.99; probe synthesis + microarray R² = 0.98–0.99). On the other hand, tissue heterogeneity

![Figure 6](http://www.biomedcentral.com/1471-2105/3/4)

**Figure 6**

**Use of >2-fold survival method provides relatively specific, but insensitive detection of significant gene changes.** Shown in green is a representation of number of genes surviving four pair-wise comparisons to two mixed control profiles, with retention of only those genes showing fold changes > 2-fold in the four pair-wise comparisons. Shown in red are differentially expressed genes for 10 individual DMD 5–6 y profiles versus 6 mixed controls at the indicated p-value thresholds. This fold-change survival method shows good specificity at p < 0.05 for individual profiles (85%), however it is relatively insensitive.
would have been detected by larger numbers of individual patient profiles. Our results suggest that stringent yet robust data can be generated by mixing a small number of individuals with a defined condition \((n = 5)\), preferably using different regions of tissue for duplicate arrays. Controls should be similarly processed. The resulting four arrays (2 controls, 2 experimental datasets) should then be subjected to the >2-fold survival method, as previously described [23]. This will yield a stringent set of expression changes that are likely to be verified by larger studies with individual arrays, but at low cost as only four arrays are employed. The preliminary data from just four mixed profiles (two experimental and two control) can then be used to generate functional clusters and pathophysiological models. These preliminary models can then direct more hypothesis-driven experiments, or more extensive expression profiling studies.

**Materials and methods**

**Expression profiling**

Human muscle biopsy samples were diagnostic specimens flash-frozen immediately after surgery in isopentane cooled in liquid nitrogen, with storage in small, airtight, humidified tubes at \(-80^\circ\text{C}\) until RNA isolation. Duchenne muscular dystrophy patient samples were all shown to have complete lack of dystrophin by immunostaining and/or immunoblot analysis, and were shown to have excellent morphology and preservation of tissue. Controls included groups of males and female (age described in text) that showed no histopathological abnormality, normal dystrophy proteins, and normal serum creatine kinase levels. Biopsy sizes ranged from 50 mg to 2 grams, with approximately 20–30 mg used for RNA isolation (~10–15 micrograms of total RNA). As described in the text, all biopsies had two different regions of the same biopsy expression profiled separately.

Details concerning the murine profiles will be published elsewhere. In this report, we used the murine profiles simply to test the sources of variation during sample preparation prior to hybridization to oligonucleotides.

RNA isolation (Trizol, Gibco BRL), RNA purification (RNAeasy, Qiagen), cDNA synthesis and biotinylated cRNA were all done as per standard protocols provided by Affymetrix Inc. Quality control methods are described on our web site (http://microarray.cnmcresearch.org/pga.htm), with cRNA amplifications of between 5- and 13-fold for each of the samples. Ten micrograms of gel-verified fragmented biotinylated cRNA were hybridized to each MuscleChip or U74A v2 array, and scanning done after biotin/avidin/phycoerythrin amplification. Details on the specific patients studied, and details for each GeneChip (scaling factors, number of present calls, percentage difference calls between each duplicate sample, number...
of difference calls surviving four pair-wise comparisons of duplicate chips) is provided (Table 1). All profiling data presented here is available on our web site (http://microarray.cnmcresearch.org); data link), as image (.dat), absolute analysis (.cph), and ASCII text conversions of .cph (.txt) for each individual profile (see http://microarray.cnmcresearch.org/pga.htm) for file descriptions and use.

Bio-informatic methods

Absolute analysis (average difference determinations for each probe set) was done using Affymetrix default parameters. As described in the text, data was analyzed using a variety of methods, including unsupervised nearest-neighbor hierarchical clustering analyses (GeneSpring [28,29] [Silicon Genetics], and Cluster [25] [Stanford University]), t-test (GeneSpring) and four-comparison survival method [23]. The Cluster and Tree View software were downloaded from [http://rana.lbl.gov] and installed on an NT workstation.

Acknowledgements

We would like to thank Brian S. Tseng and Frank W. Booth (Department of Physiology University of Missouri, Columbia, MO) and Simona lezzi and Vittorio Sartorelli (Muscle Biophysics Section, Laboratory of Physical Biology NIAMS/NIH Bethesda, MD) for allowing to use their data in our analysis.

Drs. Balay and Chen were supported by post-doctoral fellowships from the Stichting-Porticus Foundation. Supported in part by grants from the National Institutes of Health (5RO1 NS29525-10; and a “Programs in Genomic Medicine” with DNA microarrays.

References

1. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ: High density synthetic oligonucleotide arrays. Nature Genetics 1999, 21:20-24
2. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM: Expression profiling using cDNA microarrays. Nature Genetics 1999, 21:10-14
3. Brown PO, Botstein D: Exploring the new world of the genome with DNA microarrays. Nature Genetics 1999, 21:33-37
4. Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G: Making and reading microarrays. Nature Genetics 1999, 21:15-19
5. McGall G, Labadie J, Brock P, Wallraff G, Nguyen T, Hinsberg W: Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. Proc. Natl. Acad. Sci. USA 1996, 93:13555-13560
6. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. 1996, 14:1675-1680
7. Winzeler EA, Richards DR, Conway AR, Goldstein AL, Kalman S, McCullough MJ, McCusker JH, Stevens DA, Wodicka L, Lockhart DJ, Davis RW: Direct allelic variation scanning of the yeast genome. Science 1998, 281:194-1197
8. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B: Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol Biol Cell 1998, 9:3273-3297
9. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO: Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 2000, 11:4241-4257
10. Sudarsanam P, Iyer VR, Brown PO, Winston F: Whole-genome expression analysis of snf1/swi mutants of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 2000, 97:12369-12374
11. Cavalieri D, Townsend JP, Hardt DL: Manifold anomalies in gene expression in a vineyard isolate of Saccharomyces cerevisiae revealed by DNA microarray analysis. Proc. Natl. Acad. Sci. USA 2000, 97:13269-13274
12. Ogawa N, DeRisi J, Brown PO: New components of a system for phosphate accumulation and polyphosphate metabolism in Saccharomyces cerevisiae revealed by genomic expression analysis. Mol Biol Cell 2000, 11:4309-4321
13. Lee CK, Kloppe RG, Weindruch R, Prolla TA: Gene expression profile of aging and its retardation by caloric restriction. Science 1999, 285:1390-1393
14. Kaminski N, Allard JD, Pizzet JF, Zuo F, Griffiths MJ, Morris D, Huang X, Sheppard D, Heller RA: Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. Proc. Natl. Acad. Sci. USA 2000, 97:1778-1783
15. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999, 286:531-537
16. Perou CM, Jeffrey SS, van de Rijn M, Ross CA, Eisen MB, van de Rijn M, Jeffery SS, Wang Q,桂芳 S, Liu Y, Roberts C, Lashkari D, Shalon D, Brown PO, Botstein D: Distinctive gene expression patterns in human mammary epithelial cell lines and breast cancers. Proc. Natl. Acad. Sci. USA 1999, 96:9221-9227
17. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P, Borresen-Dale AL: Gene expression patterns of breast cancer distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA 2001, 98:10869-74
18. Bittner M, Meltzer P, Chen Y, Jiang Y, Seflor E, Hendrix M, Ramdacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Luers J, Glatzel F, Pollock P, Carpent J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D, Sondak V: Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature 2000, 406:536-540
19. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, Levine AJ: Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Natl. Acad. Sci. USA 1999, 96:6745-6750
20. Mills JC, Gordon JL: A New approach for filtering noise from high-density oligonucleotide microarray datasets. Nucleic Acids Res 2001, 29:E72-2
21. Unger MA, Rishi M, Clemmer VB, Hartman JL, Keiper EA, Greshock JD, Chodos LA, Liebman MN, Weber BL: Characterization of adjacent breast tumors using oligonucleotide microarrays. Breast Cancer Res 2001, 3:336-341
22. Nadler ST, Stoehr JP, Schuler KL, Taniimoto G, Yandell BS, Attie AD: The expression of adipogetic genes is decreased in obesity and diabetes mellitus. Proc. Natl. Acad. Sci. USA 2000, 97:13711-11376
23. Chen Y, Zhao P, Borup R, Hoffman EP: Expression profiling in the muscular dystrophies: identification of novel aspects of molecular pathology. J. Cell Biol. 2000, 151:1321-1336
24. Kamb A, Ramaswami M: A simple method for statistical analysis of intensity differences in microarray-derived gene expression data. BMC Biotechnol 2001, 1:8
25. Eisen MB, Spellman PT, Brown PO, Botstein D: Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 1998, 95:14863-14868
26. Der SD, Zhou A, Williams BR, Silverman RH: Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. Proc. Natl. Acad. Sci. USA 1998, 95:15625-15629
27. Hooper LV, Wong MH, Thelin A, Hasson L, Falk PG, Gordon JL: Molecular analysis of commensal host-microbial relationships in the intestine. Science 2001, 291:881-884
28. GeneSpring User Manual, version 4.0, March Silicon Genetics 2001
29. GeneSpring Advanced Analysis Techniques, version 4.0, March Silicon Genetics 2001
30. Claverie JM: Computational methods for the identification of differential and coordinated gene expression. Hum Mol Genet 1999, 8:1821-1832