

# A DNA microarray survey of gene expression in normal human tissues

Radha Shyamsundar<sup>\*†</sup>, Young H Kim<sup>\*</sup>, John P Higgins<sup>\*</sup>, Kelli Montgomery<sup>\*</sup>, Michelle Jordan<sup>\*</sup>, Anand Sethuraman<sup>‡</sup>, Matt van de Rijn<sup>\*</sup>, David Botstein<sup>¶¶</sup>, Patrick O Brown<sup>†§</sup> and Jonathan R Pollack<sup>\*</sup>

Addresses: <sup>\*</sup>Department of Pathology, Stanford University School of Medicine, 269 Campus Drive, CCSR 3245A, Stanford, CA 94305-5176, USA. <sup>†</sup>Department of Biochemistry, Stanford University School of Medicine, 279 Campus Drive, Stanford, CA 94305-5307, USA. <sup>‡</sup>Department of Genetics, Stanford University, Stanford, CA 94305, USA. <sup>§</sup>Howard Hughes Medical Institute, Stanford University School of Medicine, 279 Campus Drive, Stanford, CA 94305-5307, USA. <sup>¶</sup>Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 80544, USA.

Correspondence: Patrick O Brown. E-mail: pbrown@cmgm.stanford.edu. Jonathan R Pollack. E-mail: pollack1@stanford.edu

Published: 14 February 2005

*Genome Biology* 2005, 6:R22

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2005/6/3/R22>

Received: 29 November 2004

Revised: 14 January 2005

Accepted: 18 January 2005

© 2005 Shyamsundar et al.; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** Numerous studies have used DNA microarrays to survey gene expression in cancer and other disease states. Comparatively little is known about the genes expressed across the gamut of normal human tissues. Systematic studies of global gene-expression patterns, by linking variation in the expression of specific genes to phenotypic variation in the cells or tissues in which they are expressed, provide clues to the molecular organization of diverse cells and to the potential roles of the genes.

**Results:** Here we describe a systematic survey of gene expression in 115 human tissue samples representing 35 different tissue types, using cDNA microarrays representing approximately 26,000 different human genes. Unsupervised hierarchical cluster analysis of the gene-expression patterns in these tissues identified clusters of genes with related biological functions and grouped the tissue specimens in a pattern that reflected their anatomic locations, cellular compositions or physiologic functions. In unsupervised and supervised analyses, tissue-specific patterns of gene expression were readily discernable. By comparative hybridization to normal genomic DNA, we were also able to estimate transcript abundances for expressed genes.

**Conclusions:** Our dataset provides a baseline for comparison to diseased tissues, and will aid in the identification of tissue-specific functions. In addition, our analysis identifies potential molecular markers for detection of injury to specific organs and tissues, and provides a foundation for selection of potential targets for selective anticancer therapy.

## Background

DNA microarrays [1,2] have been used to profile gene expression in cancer and other diseases. In cancer, for example, microarray profiling has been applied to classify tumors according to their sites of origin [3-5], to discover previously unrecognized subtypes of cancer [6-11], to predict clinical outcome [12-14] and to suggest targets for therapy [15,16]. However, the identification of improved markers for diagnosis and molecular targets for therapy will depend on knowledge not only of the genes expressed in the diseased tissues of interest, but also on detailed information about the expression of the corresponding genes across the gamut of normal human tissues.

At present there is relatively little data on gene expression across the diversity of normal human tissues [17-20]. Here we report a DNA microarray-based survey of gene expression in a diverse collection of normal human tissues and also present an empirical method for estimating transcript abundance from DNA microarray data.

## Results

### Hierarchical clustering of gene expression in normal tissues

To survey gene expression across normal human tissues, we analyzed 115 normal tissue specimens representing 35 different human tissue types, using cDNA microarray representing 26,260 different genes (see Materials and methods). To explore the relationship among samples and underlying features of gene expression, we applied an unsupervised two-way (that is, genes against samples) hierarchical clustering method using the 5,592 cDNAs (representing 3,960 different UniGene clusters [21]) whose expression varied most across samples (Figure 1a; also see Additional data file 2). Overall, tissue samples clustered in large part according to their anatomic locations, cellular compositions or physiologic functions (Figure 1b). For example, lymphoid tissues (lymph node, tonsil, thymus, buffy coat and spleen) clustered together, as did gastrointestinal tissues (stomach, gall bladder, liver, pancreas, small bowel and colon), muscular tissues (heart and skeletal muscle), secretory tissues (parathyroid, thyroid, prostate, seminal vesicle and salivary gland), and female genitourinary tissues (ovary, fallopian tube, uterus, cervix and bladder). Brain and testis were also found to cluster together, largely because genes encoding ribosomal proteins and lymphoid-specific genes were expressed at particularly low levels in both tissues, the latter possibly reflecting immunological privilege [22].

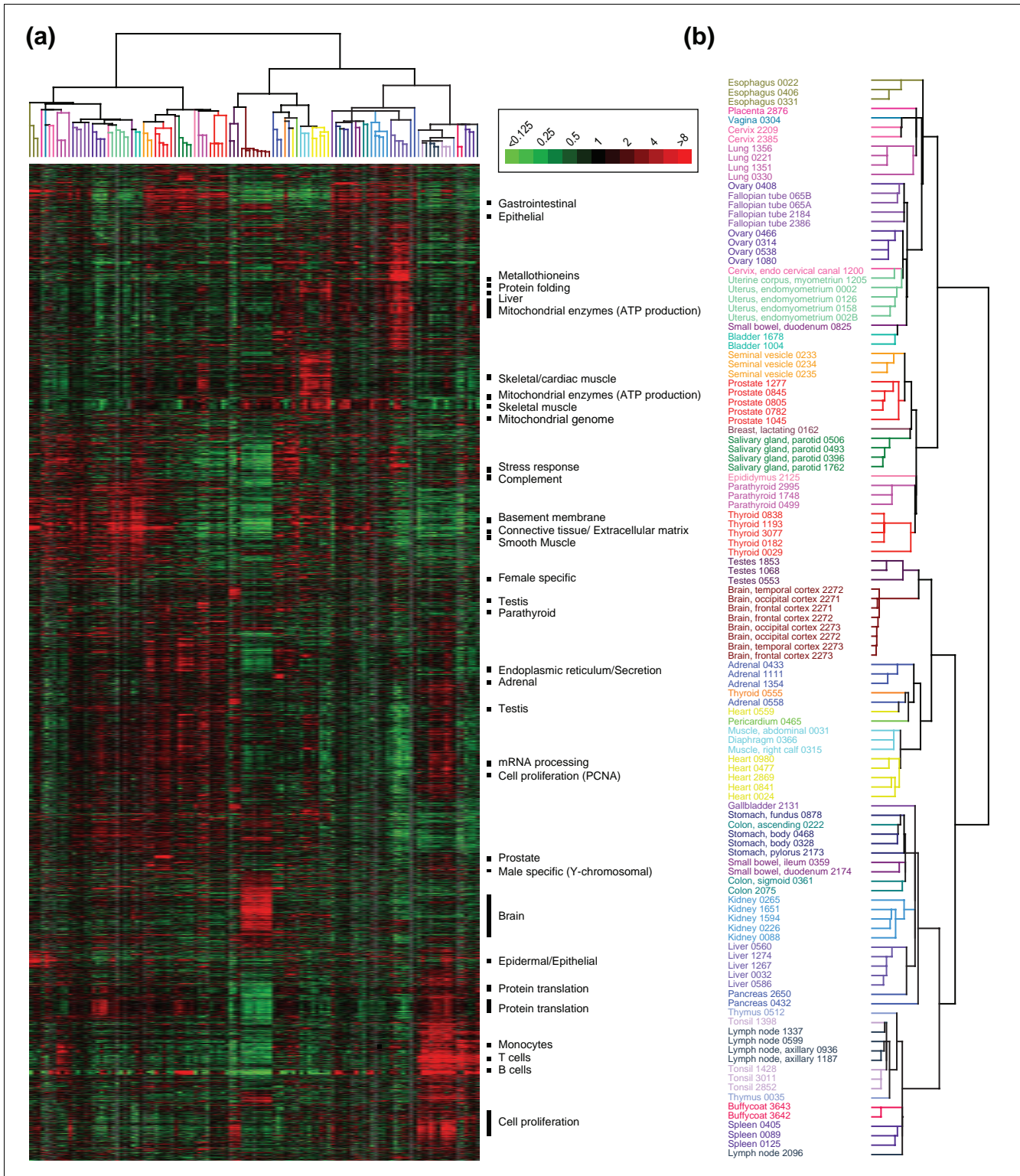
The two-way unsupervised analysis also identified clusters of coexpressed genes (annotated in Figure 1), which represented both tissue-specific structures and systems (discussed further below) and coordinately regulated cellular processes. For example, on the basis of the shared characteristics of well annotated genes in the clusters, we identified clusters repre-

sented cell proliferation [23], mitochondrial ATP production, mRNA processing, protein translation and endoplasmic reticulum-associated protein modification and secretion. Interestingly, proliferation, mitochondrial ATP production and protein translation were each represented by two distinct clusters of genes, suggesting that subsets of these functions might be differentially regulated among different tissues. One gene cluster corresponded to sequences on the mitochondrial chromosome [24]; we interpret this feature to reflect the relative abundance of mitochondria in each tissue sample.

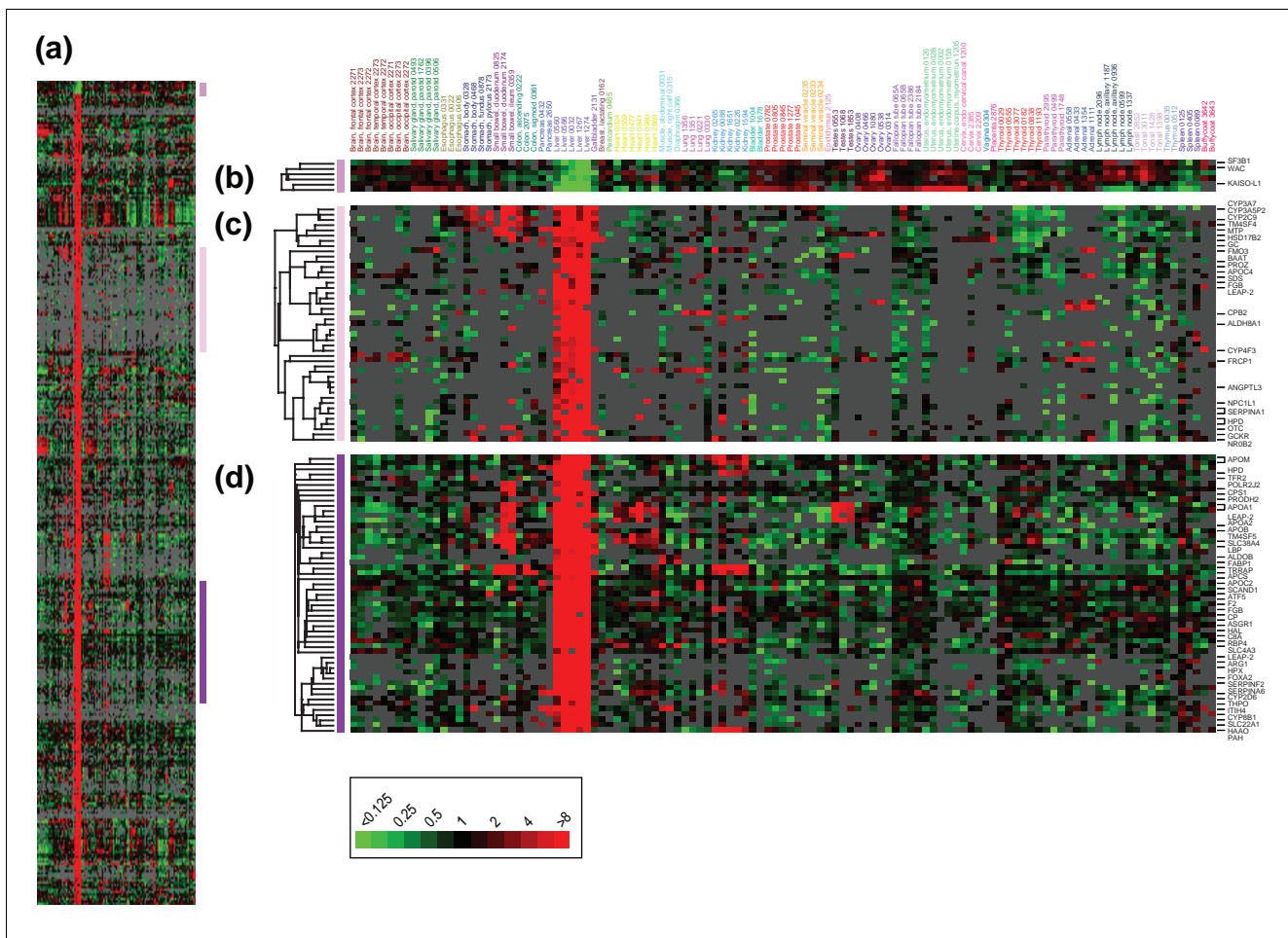
### Identifying tissue-specific gene expression

While tissue-specific gene expression features were apparent in the hierarchical cluster, in order to identify tissue-specific genes more systematically we performed supervised analyses using the significance analysis of microarrays (SAM) method ([25], see Materials and methods). Tissue-specific genes were identified for all tissues analyzed, and included named genes with known tissue-specific functions, as well as named genes and anonymous expressed sequence tags (ESTs) that had not been previously characterized as having tissue-specific functions. For example, while the set of liver-specific genes (Figure 2) included, as expected, genes encoding blood-clotting factors (for example, *F2*, *F7*), complement components (*C1R*, *C2*), lipid (*APOB*, *APOE*) and metal transport proteins (*TF*, *CP*), and proteins for detoxification (*CYP2D6*, *CYP3A7*), amino acid metabolism (*PAH*, *HAL*) and carbohydrate metabolism (*G6PT1*, *GYS2*), other intriguing genes, for example, *WRNIP1* (Werner helicase interacting protein 1), *BIRC5* (survivin), *ANGPTL3* (angiopoietin-like 3), and *CNTNAP1* (contactin associated protein 1), were also identified as selectively expressed in liver. The new connections these results might make between our knowledge of the gene and its product on the one hand, and our knowledge of the physiological functions, cellular characteristics and pathologies of a specific organ, on the other, are a step towards better understanding of both the genes and the organs. Interestingly, we also identified a smaller number of genes displaying selectively decreased expression in some organs, for example, splicing factor *SF3B1* in the liver (Figure 2b): we speculate that the decreased expression of such genes may have a role in regulating cellular/tissue differentiation. Tissue-specific genes characteristically expressed in each of the tissues we examined are viewable in Additional data file 6 (see also Additional data file 3).

Recent efforts by the Gene Ontology (GO) Consortium have resulted in the systematic annotation of genes, ascribing genes to specific biological processes, cellular components and molecular functions [26]. This annotation system, while rudimentary, facilitates the systematic exploration of the expression of genes reflecting specific biological processes, cellular components and molecular functions in these normal tissues. For example, the gene sets encoding tyrosine kinase, G-protein-coupled receptor and transcription factor functions, as well as components of the extracellular matrix and



**Figure 1**  
 Hierarchical cluster analysis of normal tissue specimens. **(a)** Thumbnail overview of the two-way hierarchical cluster of 115 normal tissue specimens (columns) and 5,592 variably-expressed genes (rows). Mean-centered gene expression ratios are depicted by a  $\log_2$  pseudocolor scale (ratio fold-change indicated); gray denotes poorly-measured data. Selected gene-expression clusters are annotated. The dataset represented here is available as Additional data file 2. **(b)** Enlarged view of the sample dendrogram. Terminal branches for samples are color-coded by tissue type.



**Figure 2**  
 Liver-specific gene expression. **(a)** Thumbnail overview of a hierarchical cluster of 115 normal tissue specimens and 353 variably expressed genes identified using the SAM method (see Materials and methods) as selectively expressed in liver (false discovery rate = 0.12%). Genes are hierarchically clustered, while samples are grouped by tissue type and ordered according to anatomical location/function. Mean-centered gene-expression ratios are depicted by a log<sub>2</sub> pseudocolor scale (indicated); samples are color-coded by tissue type. **(b-d)** Selected gene-expression clusters (locations indicated by vertical colored bars). Because of space limitations, only named genes (and not expressed sequence tags (ESTs)) are indicated. Tissue-specific genes identified for other tissues are available as Additional data files 3 and 6.

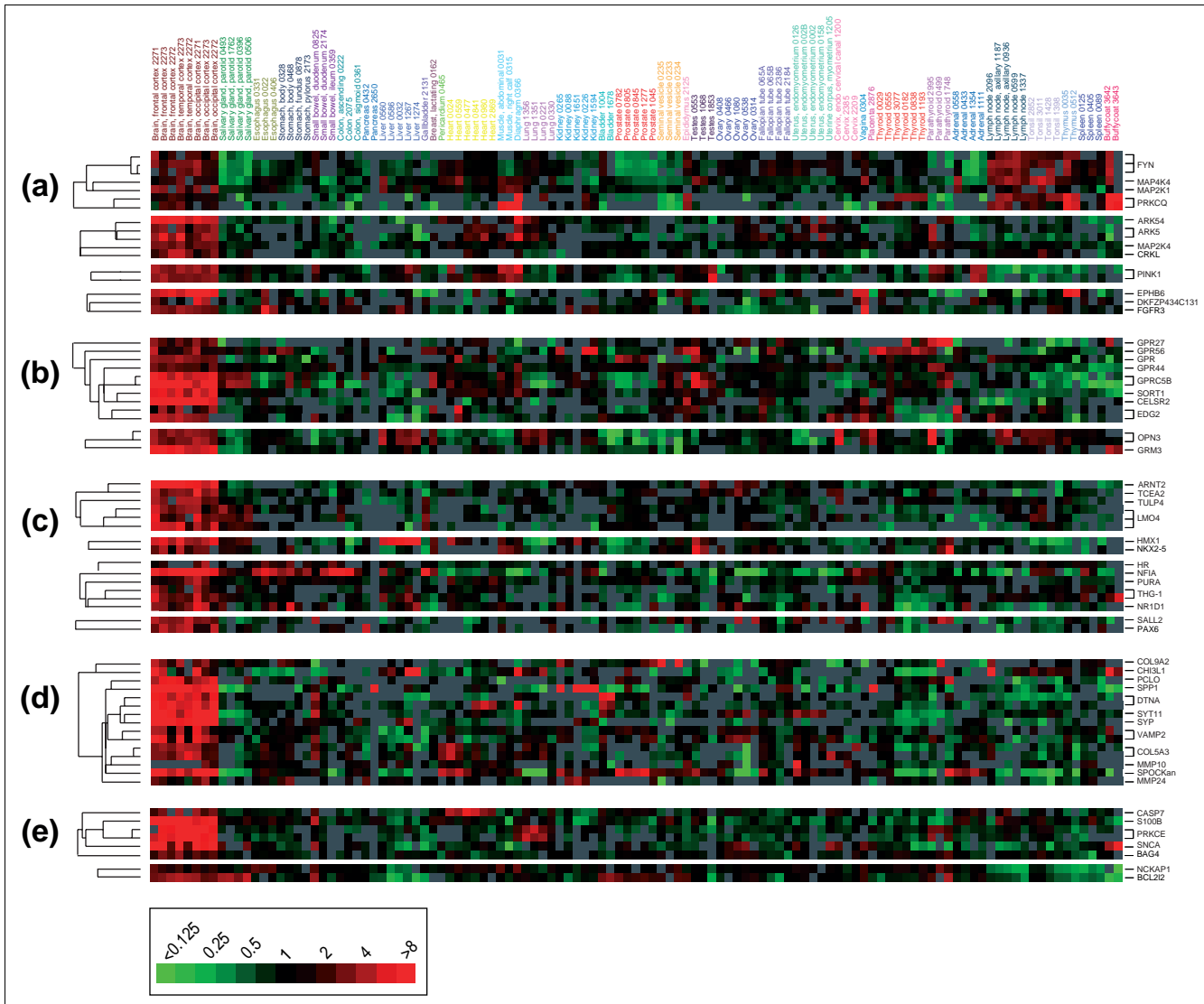
the process of programmed cell death, each demonstrate tissue-specific patterns of expression (Figure 3; see also Additional data files 4 and 7).

**Estimating transcript abundance**

DNA microarray experiments are often performed as comparative two-color hybridizations, permitting precise quantification of the ratio of each gene's expression between two samples. In the experiments reported here, each tissue sample was compared by hybridization to the same 'common reference' mRNA (see Materials and methods), a standard experimental design permitting the comparison of expression across all samples [27]. Therefore, the primary measurements give us a precise picture of the variation in relative levels of each gene's expression among the samples. While this information is sufficient for many purposes, a quantitative comparison of the expression levels of transcripts of different

genes is also of interest, for example in selecting especially highly expressed genes for potential diagnostic markers or therapeutic targets. Single-channel fluorescence intensities can provide a crude estimate of the relative transcript abundance of different genes, but do not control for the variable quantities of spotted DNA.

To estimate transcript levels for our dataset, we used microarray hybridization to compare the common reference mRNA against normal female genomic DNA. We reasoned that, for each gene on the microarray, the ratio of mRNA to genomic DNA should reflect the relative level of transcript in the common reference compared to normal genomic DNA (for which each gene is present in two copies per cell). For each tissue sample in our study, the ratio of expression for each gene in that sample versus common reference mRNA, multiplied by the ratio for that gene in common reference mRNA versus

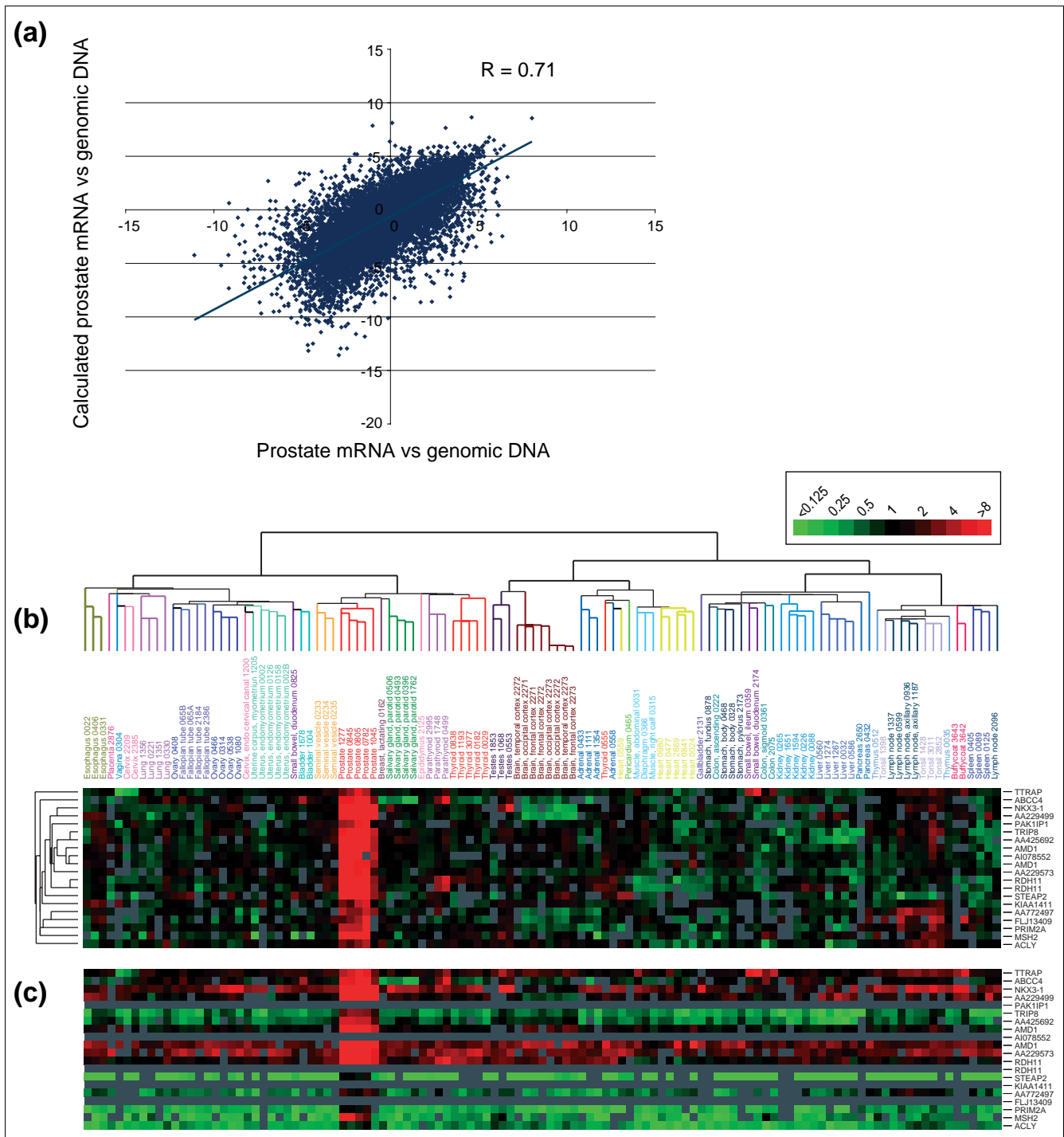


**Figure 3** Brain-selective expression of functionally annotated gene sets. Hierarchical cluster of 115 normal tissue specimens and annotated gene sets representing the following examples of (a-c) specific molecular functions (a) tyrosine kinase, (b) G-protein-coupled receptor, (c) transcription factor, (d) cellular components (extracellular matrix) or (e) biological processes (programmed cell death). Samples are ordered as in Figure 2. Genes are ordered by hierarchical clustering. For gene selection, we considered genes that were well measured in at least 50% of samples; no ratio-fold cutoff was applied. Only features representing brain-specific expression are shown here; the complete clusters are available as Additional data files 4 and 7.

normal genomic DNA, would then approximate transcript abundance. To test our approach, we compared our estimates of transcript levels for a single prostate specimen, calculated either indirectly using the common reference mRNA versus genomic DNA ratios, or calculated through a direct hybridization comparison of prostate sample mRNA versus normal female genomic DNA. Our results show high concordance for the prostate sample (Figure 4a); comparable results were obtained in a similar analysis using liver, breast, heart and kidney specimens (data not shown).

The utility of this approach is illustrated for the cluster of prostate-specific genes (derived from the hierarchical cluster

in Figure 1), and is evident on comparing results depicting the relative level of each gene's expression in different samples (Figure 4b), and the relative levels of transcripts for different genes (Figure 4c). While all genes within the prostate-specific cluster were expressed at relatively increased levels in prostate compared with other tissues, estimates of transcript abundance indicated that only a subset of these genes was highly expressed in the prostate (Figure 4c). For example, *RDH11* was highly expressed in prostate and was expressed at lower levels in other tissues, while *STEAP2* was expressed at low levels in prostate and displayed very little or no expression in other tissues. For each of the tissue types, transcripts identified as both highly abundant and tissue specific are



**Figure 4** Estimating relative transcript abundance. **(a)** Comparison of transcript levels estimated either directly by hybridization of prostate sample mRNA versus normal female genomic DNA, or indirectly by multiplying the ratio of prostate sample mRNA vs common reference mRNA by the ratio of common reference mRNA vs normal female genomic DNA. The correlation value ( $R$ ) is indicated. **(b)** Prostate-specific gene-expression cluster, extracted from the hierarchical cluster shown in Figure 1a, is displayed as mean-centered relative gene expression (ratio-fold change scale indicated). **(c)** The same gene-expression feature as in (b), is now displayed as transcript abundance (relative to the average transcript level for all expressed genes), calculated indirectly using the common reference mRNA versus normal female genomic DNA hybridization data.

displayed in Additional data files 5 and 8 (for the transcript levels of all variably expressed genes, see Additional data file 2).

## Discussion

The main objective of our study was to survey variation in gene expression across a diverse set of normal human tissue types. We have reported here a cDNA microarray gene-expression dataset profiling approximately 26,000 human genes across 115 human tissue specimens representing 35 different tissue types. An unsupervised, two-way hierarchical clustering of the genes whose expression varied most across samples showed that at the level of gene expression, the relationship among tissues was in large part based on their anatomic locations, cellular compositions and physiologic functions. Tissue-specific features of gene expression were readily discernable in the hierarchical cluster, as were gene-expression features related to specific cellular processes (as inferred from the named genes within these features). Of particular importance, the function of uncharacterized ESTs might be deduced by virtue of their inclusion in one of these clusters. Supervised analysis also identified genes selectively expressed in each of the tissues types studied, and the analysis of functionally annotated gene sets provided information on the tissue distribution of specific biological processes, cellular components and molecular functions.

We have also reported here the application of mRNA versus genomic DNA hybridizations for estimating transcript abundances for expressed genes. Knowledge of transcript abundance should prove useful in prioritizing candidate genes for use as diagnostic markers or therapeutic targets, for which more highly expressed genes might be more tenable candidates. It is worth pointing out that our approach for estimating absolute transcript levels should be applicable to any cDNA microarray study incorporating a common reference mRNA.

While many investigators have been using DNA microarrays to profile gene expression in cancer and other human diseases, scant data exist on profiles of gene expression across the diversity of normal human tissues. Our cDNA-microarray-based survey of gene expression in normal human tissues provides a publicly accessible dataset which can be used in future analyses aimed at better understanding the physiology of various normal tissues; developing a baseline for comparison to diseased tissues, including cancer; identifying tissue-specific diagnostic markers that signify tissue injury; discovering tissue-specific therapeutic targets (for example, for treatment of prostate cancer); and identifying tumor-specific diagnostic markers and therapeutic targets, for which minimal expression in the collection of normal adult human tissues is desirable.

## Conclusions

We have used cDNA microarrays to survey gene expression across a diverse set of normal human tissues. Using unsupervised and supervised analyses, we have identified tissue-specific patterns of gene expression. Furthermore, by comparative hybridization to normal genomic DNA, we were able to estimate transcript abundances and identify the subsets of abundantly expressed tissue-specific genes. Our dataset provides a baseline for comparison to diseased tissues, as well as a basis for identifying molecular markers of injury to specific organs and tissues, and for anticancer therapy.

## Materials and methods

### Tissue specimens

Normal human tissue specimens were obtained from surgery (for example, the uninvolved regions of resected tumors) or from autopsy, with institutional review board approval. Specimens were frozen on dry ice within 30 minutes of surgical removal or procurement and stored at  $-80^{\circ}\text{C}$ . Histological evaluations were performed by H&E staining of frozen sections, and a pathologist (J.H. and/or M.vdR.) reviewed all slides to confirm the anatomical site of origin and histological normalcy (that is, to rule out inflammation, infection, necrosis, malignancy). In total, we selected for study 115 tissue samples representing 35 different human tissues (Additional data file 1). Total RNA was isolated from tissues using TRIzol Reagent (Invitrogen) according to the manufacturer's instruction, and RNA quality was assessed by the integrity of rRNA bands following gel electrophoresis. The poly(A)<sup>+</sup> mRNA fraction was then isolated from total RNA using FastTrack2.0 kit (Invitrogen), and quantified by UV spectrophotometry.

### Expression profiling

Gene-expression profiling was performed essentially as reported previously [8], and detailed protocols for array fabrication and hybridization are available online [28]. Briefly, Cy5-labeled cDNA was prepared using 2  $\mu\text{g}$  mRNA from normal tissue samples, and Cy3-labeled cDNA was prepared using 1.5  $\mu\text{g}$  mRNA common reference, pooled from 11 established human cell lines [8]. For each experimental sample, Cy5- and Cy3-labeled samples were co-hybridized to a cDNA microarray containing 39,711 human cDNAs, representing 26,260 different genes (UniGene clusters [21]). For the common reference mRNA (Cy5) versus genomic DNA (Cy3) comparisons, normal female genomic DNA was labeled as described [24]. Following hybridization, microarrays were imaged using an Axon GenePix 4000 scanner (Axon Instruments). Fluorescence ratios for array elements were extracted using GenePix software, and uploaded onto the Stanford Microarray Database (SMD) [29] for subsequent analysis. The complete microarray dataset is accessible from SMD [30], or from the Gene Expression Omnibus [31] (accession number GSE2193).

## Data analysis

Fluorescence ratios were normalized by mean-centering genes for each array (that is, 'global' normalization), and then by mean centering each gene across all arrays. We included for analysis only well-measured genes whose expression varied, as determined by: signal intensity over background more than twofold in either test or reference channels in at least 75% of samples; and a fourfold or more ratio variation from the mean in at least two samples (unless otherwise indicated). Hierarchical clustering was performed and displayed using Cluster and TreeView software [32]. Tissue-selective genes were identified using the two-class (each tissue versus all other tissues) significance analysis of microarrays (SAM) method [25], which utilizes a modified *t*-test statistic and sample-label permutations to evaluate statistical significance. The false-discovery rate (FDR), an estimate of the fraction of falsely called tissue-selective genes, varied by tissue, but in all cases was less than 5% (specific FDRs are listed in Additional data file 3). For tissue-selective genes, only tissue types with two or more samples were considered for analysis, and we only considered genes that were well-measured in more than 50% of the samples for the selected tissue type analyzed. GO annotations were assigned to arrayed genes using the AmiGO browser [33] to select relevant GO annotations, and the 'loc2go' file [34] to identify the corresponding sets of genes. Transcript abundance was estimated by multiplying (for each gene) the ratio of tissue sample mRNA versus common reference mRNA by the ratio (average ratio from triplicate experiments) of common reference mRNA versus normal female genomic DNA. Highly-abundant tissue specific transcripts were defined for each tissue type as the top (capped at 50 genes) tissue specific transcripts, identified using the SAM method, from the 1,000 most abundantly expressed transcripts in the full dataset.

## Additional data files

The following additional data are available with the online version of this paper. Additional data file 1 is a table listing the normal tissue specimens included in microarray analysis. Additional data file 2 is a table listing the variably expressed genes. Additional data file 3 is a table listing tissue-specific transcripts. Additional data file 4 is a table listing functionally annotated gene sets. Additional data file 5 is a table listing highly abundant tissue-specific transcripts. Additional data file 6 is a figure showing tissue-specific gene expression. Additional data file 7 is a figure showing expression of functionally annotated gene sets. Additional data file 8 is a figure showing highly abundant tissue-specific gene expression.

## Acknowledgements

We thank Ash Alizadeh and the members of the Pollack and Brown labs for helpful suggestions. We also thank Janet Mitchell and the Stanford Tissue Bank for collection of tissues, Mike Fero and the staff of the Stanford Functional Genomics Facility for providing high-quality cDNA microarrays, and Gavin Sherlock and Catherine Ball of the Stanford Microarray Database group for providing outstanding database support. This work was sup-

ported by a grant from the National Cancer Institute. P.O.B is an investigator of the Howard Hughes Medical Institute.

## References

- Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**:467-470.
- 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.
- Su AI, Welsh JB, Sapinoso LM, Kern SG, Dimitrov P, Lapp H, Schultz PG, Powell SM, Moskaluk CA, Frierson HF Jr, Hampton GM: **Molecular classification of human carcinomas by use of gene expression signatures.** *Cancer Res* 2001, **61**:7388-7393.
- Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, et al.: **Multiclass cancer diagnosis using tumor gene expression signatures.** *Proc Natl Acad Sci USA* 2001, **98**:15149-15154.
- Bloom G, Yang IV, Boulware D, Kwong KY, Coppola D, Eschrich S, Quackenbush J, Yeatman TJ: **Multi-platform, multi-site, microarray-based human tumor classification.** *Am J Pathol* 2004, **164**:9-16.
- Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, et al.: **Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.** *Nature* 2000, **403**:503-511.
- Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, et al.: **Molecular classification of cutaneous malignant melanoma by gene expression profiling.** *Nature* 2000, **406**:536-540.
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al.: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747-752.
- Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, et al.: **Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses.** *Proc Natl Acad Sci USA* 2001, **98**:13790-13795.
- Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, van De Rijn M, Rosen GD, Perou CM, Whyte RI, et al.: **Diversity of gene expression in adenocarcinoma of the lung.** *Proc Natl Acad Sci USA* 2001, **98**:13784-13789.
- Lapointe J, Li C, Higgins JP, Van De Rijn M, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, Bergerheim U, et al.: **Gene expression profiling identifies clinically relevant subtypes of prostate cancer.** *Proc Natl Acad Sci USA* 2004, **101**:811-816.
- van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, et al.: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 2002, **415**:530-536.
- Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, et al.: **Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning.** *Nat Med* 2002, **8**:68-74.
- Leung SY, Chen X, Chu KM, Yuen ST, Mathy J, Ji J, Chan AS, Li R, Law S, Troyanskaya OG, et al.: **Phospholipase A2 group IIA expression in gastric adenocarcinoma is associated with prolonged survival and less frequent metastasis.** *Proc Natl Acad Sci USA* 2002, **99**:16203-16208.
- Armstrong SA, Kung AL, Mabon ME, Silverman LB, Stam RW, Den Boer ML, Pieters R, Kersey JH, Sallan SE, Fletcher JA, et al.: **Inhibition of FLT3 in MLL. Validation of a therapeutic target identified by gene expression based classification.** *Cancer Cell* 2003, **3**:173-183.
- Stegmaier K, Ross KN, Colavito SA, O'Malley S, Stockwell BR, Golub TR: **Gene expression-based high-throughput screening(GE-HTS) and application to leukemia differentiation.** *Nat Genet* 2004, **36**:257-263.
- Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M: **Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes.** *Physiol Genomics* 2000, **2**:143-147.
- Hsiao LL, Dangond F, Yoshida T, Hong R, Jensen RV, Misra J, Dillon



- W, Lee KF, Clark KE, Haverty P, et al.: **A compendium of gene expression in normal human tissues.** *Physiol Genomics* 2001, **7**:97-104.
19. Saito-Hisaminato A, Katagiri T, Kakiuchi S, Nakamura T, Tsunoda T, Nakamura Y: **Genome-wide profiling of gene expression in 29 normal human tissues with a cDNA microarray.** *DNA Res* 2002, **9**:35-45.
  20. Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, Orth AP, Vega RG, Sapinoso LM, Moqrich A, et al.: **Large-scale analysis of the human and mouse transcriptomes.** *Proc Natl Acad Sci USA* 2002, **99**:4465-4470.
  21. Schuler GD: **Pieces of the puzzle: expressed sequence tags and the catalog of human genes.** *J Mol Med* 1997, **75**:694-698.
  22. Filippini A, Riccioli A, Padula F, Lauretti P, D'Alessio A, De Cesaris P, Gandini L, Lenzi A, Ziparo E: **Control and impairment of immune privilege in the testis and in semen.** *Hum Reprod Update* 2001, **7**:444-449.
  23. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC, et al.: **Distinctive gene expression patterns in human mammary epithelial cells and breast cancers.** *Proc Natl Acad Sci USA* 1999, **96**:9212-9217.
  24. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO: **Genome-wide analysis of DNA copy-number changes using cDNA microarrays.** *Nat Genet* 1999, **23**:41-46.
  25. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.
  26. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al.: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
  27. Eisen MB, Brown PO: **DNA arrays for analysis of gene expression.** *Methods Enzymol* 1999, **303**:179-205.
  28. **Pat Brown's Lab Protocols** [<http://brownlab.stanford.edu/protocols.html>]
  29. Gollub J, Ball CA, Binkley G, Demeter J, Finkelstein DB, Hebert JM, Hernandez-Boussard T, Jin H, Kaloper M, Matese JC, et al.: **The Stanford Microarray Database: data access and quality assessment tools.** *Nucleic Acids Res* 2003, **31**:94-96.
  30. **Stanford Microarray Database** [<http://smd.stanford.edu>]
  31. **Gene Expression Omnibus (GEO)** [<http://www.ncbi.nlm.nih.gov/geo/>]
  32. **EisenLab Software** [<http://rana.lbl.gov/EisenSoftware.htm>]
  33. **AmiGO** [<http://www.godatabase.org/cgi-bin/go.cgi>]
  34. **NCBI loc2go** [<ftp://ftp.ncbi.nih.gov/refseq/LocusLink/loc2go>]