Physiol Genomics
2: 143–147, 2000.

Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes

JANET A. WARRINGTON, ARCHANA NAIR, MAMATHA MAHADEVAPPA, AND MAYA TSYGANSKAYA
Affymetrix, Inc., Santa Clara, California 95051

Warrington, Janet A., Archana Nair, Mamatha Mahadevappa, and Maya Tsyganskaya. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiol Genomics* 2: 143–147, 2000.—Gene expression levels of about 7,000 genes were measured in 11 different human adult and fetal tissues using high-density oligonucleotide arrays to identify genes involved in cellular maintenance. The tissues share a set of 535 transcripts that are turned on early in fetal development and stay on throughout adulthood. Because our goal was to identify genes that are involved in maintaining cellular function in normal individuals, we minimized the effect of individual variation by screening mRNA pooled from many individuals. This information is useful for establishing average expression levels in normal individuals. Additionally, we identified transcripts uniquely expressed in each of the 11 tissues.

housekeeping genes; DNA chips; human gene expression; cellular maintenance genes

HOUSEKEEPING GENES, or maintenance genes, are those genes constitutively expressed to maintain cellular function (31). Previously, tens of genes have been reported as putative housekeeping genes, but no small or large scale studies have been reported in which the identification of housekeeping genes was the primary goal of the study. The genes previously reported were identified by conventional methods, and the putative housekeeping role of the gene product is an incidental observation (8, 13, 17, 19, 21–23, 27, 34). Only recently has it become practical to perform large quantitative surveys screening thousands of genes simultaneously to obtain expression information on a genomic scale (9, 10, 16, 24, 28, 30). The availability of sequence information made possible by the Human Genome Project and the ability to produce and read high-density oligonucleotide arrays are the two major developments making this type of large scale analysis possible. Because of the specificity and sensitivity of high-density probe arrays, we were able to simultaneously measure thousands of genes expressed at low, moderate, and high abundance (6, 16, 32). In an effort to identify the subset of genes required for cell maintenance, we measured expression levels of about 7,000 full-length genes in 11 different human tissues including adult heart, brain, lung, kidney, pancreas, uterus, and testis and fetal brain, lung, kidney, and liver. At the same time, we identified genes uniquely expressed in each of the 11 tissues, genes expressed in fetal tissues that are not detected in adult tissues, genes detected in adult tissues that are not detected in fetal tissues, genes uniquely expressed in a comparison of the 7 adult tissues, and genes uniquely expressed in a comparison of the 4 fetal tissues. Arguably, the uniquely expressed genes are critical for the specific functions that characterize and distinguish heart, brain, lung, liver, kidney, pancreas, uterus, and testis. Genes identified as expressed exclusively in fetal tissues may provide clues to developmental processes and are a candidate set for further analysis in disease studies.

Received 21 December 1999; accepted in final form 2 March 2000.
Article published online before print. See web site for date of publication (http://physiolgenomics.physiology.org).

MATERIALS AND METHODS

Sample preparation. All samples were prepared from pools of human poly(A) RNA purchased from Clontech (Palo Alto, CA). The tissues screened are listed followed by the number of tissues pooled and the Clontech catalog number in parenthesis. Adult samples: heart, 3 (6533–1); brain, 5 (6516–1); lung, 5 (6524–1); kidney, 8 (6538–1); pancreas, 10 (6539–1); uterus, 10 (6537–1); and testis, 19 (6535–1). Fetal samples: brain, 9 (6525–1); kidney, 27 (6526–1); lung, 7 (6528–1); and liver, 17 (6527–1). Poly(A) RNA was amplified and labeled with biotin following the procedure described by Wodicka et al. in 1997 (32). First-strand cDNA synthesis was carried out at 37°C for 60 min. The amplified cRNA (target) was purified on an affinity resin (RNeasy, Qiagen) and quantitated.

Fragmentation, array hybridization, and scanning. Labeled target was fragmented by incubation at 94°C for 35 min in the presence of 40 mM Tris-acetate, pH 8.1, 100 mM potassium acetate, and 30 mM magnesium acetate. The hybridization solution consisted of 20 μg fragmented cRNA and 0.1 mg/ml sonicated herring sperm DNA, in 1× MES buffer (containing 100 mM MES, 1 M Na^+, 20 mM EDTA, and 0.01% Tween 20). The hybridization mixture was heated to 99°C for 5 min followed by incubation at 45°C for 5 min before injection of the sample into the probe array cartridge. All preparations and hybridizations were performed in duplicate and were carried out at 45°C for 16–17 h with mixing on a rotisserie at 60 rpm. Following hybridization, the solutions were removed, arrays were rinsed with 1× MES. Subsequent washing and staining of the arrays was carried out using the GeneChip Fluidics station protocol EukGE_WS2. The EukGE_WS2 protocol included two posthybridization washes, staining, and a poststain wash. The first wash consisted of 10 cycles of 2 mixes per cycle with nonstringent wash buffer (6× SSPE, 0.01% Tween 20, and 0.005% antifoam) at 25°C. The second wash consisted of 4 cycles of 15 mixes per cycle with stringent wash buffer (100 mM MES, 0.1 M Na^+, and 0.01%

Tween 20) at 50°C. The probe arrays were stained for 10 min in streptavidin-phycoerythrin solution (SAPE) [1× MES solution, 0.005% antifoam, 10 μg/ml SAPE (Molecular Probes, Eugene, OR), and 2 μg/μl acetylated BSA (Sigma, St. Louis, MO)] at 25°C. The poststain wash consisted of 10 cycles of 4 mixes per cycle at 25°C. The probe arrays were treated for 10 min in antibody solution [1× MES solution, 0.005% antifoam, 2 μg/μl acetylated BSA, 0.1 μg/μl normal goat IgG (Sigma Chemical), 3 μg/μl antibody (goat), and antistreptavidin, biotinylated (Vector Laboratories, Burlingame, CA)] at 25°C. The final wash consisted of 15 cycles of 4 mixes per cycle at 30°C. Following washing and staining, probe arrays were scanned twice (multiple image scan) at 3-μm resolution using the GeneChip System confocal scanner made for Affymetrix by Hewlett-Packard.

Probe arrays. The arrays were synthesized using light-directed combinatorial chemistry as described previously (9, 10). The HuGeneFL GeneChip probe arrays used for the current study contain probe sets representing 7,129 genes. The oligonucleotides are 25 bases in length. Probes are complementary and correspond to human genes registered in Unigene, GenBank, and The Institute for Genomic Research Database (TIGR). Each probe set has oligonucleotides that are identical to sequence in the gene and oligonucleotides that contain a homomeric (base transversion) mismatch at the central base position of the oligomer used for measuring cross hybridization. Probes are selected with a bias toward the 3′ region of each gene. Probe pairs representing human genes such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH), β-actin, transferrin receptor, and transcription factor ISGF-3 serve as internal controls for monitoring RNA integrity. In addition, the probe arrays contain oligonucleotides representing sequences of bacterial genes, BioB, BioC, and BioD, and one phage gene, Cre, as quantitative standards. Copy numbers are determined by correlating the known concentrations of the spiked standards with their hybridization intensities as described previously (16). Copies per cell are calculated based on the assumption that the average transcript length is 1 kb and there are 300,000 transcripts per cell.

Analysis. All samples were prepared and hybridized in duplicate. Only those transcripts detected as present in duplicate hybridizations or absent in duplicate hybridizations are reported. Of the transcripts present in duplicate hybridizations, the hybridization values were within twofold. The values from the duplicate hybridizations were averaged. GeneChip 3.0 software was used to scan and analyze the data. Microsoft Excel and Microsoft Access were also used for data analysis.

RESULTS

Identification of housekeeping/cellular maintenance genes. Using GeneChip probe arrays (DNA chips), we identified 535 genes that are expressed in each of 11 fetal and adult tissues. These genes are turned on early in fetal development and stay on throughout adulthood and therefore are likely candidates as the genes responsible for cellular maintenance also known as housekeeping genes. Forty-seven of the transcripts are detected at similar levels in each of the tissues and will be useful as a set of controls. For example, in each of the 11 tissues surveyed, the transcript for elongation factor EF-1-α (GenBank accession no. J04617) is detected at high abundance and the transcript for E2 ubiquitin (U39317) is detected in low abundance. Of the 535 genes, 288 of the transcripts vary in expression level by 5- to 10-fold, 134 transcripts vary by 11- to 19-fold, and 69 vary by greater than 19-fold. (For a list of the 535 genes expressed in all 11 tissues and the 47 transcripts expressed at the same levels, see Tables 1 and 2 of the supplementary material.[1])

The majority of the maintenance transcripts detected were present in moderate levels. The distributions of transcripts detected in all 11 tissues sorted by tissue type and abundance level are shown in Fig. 1. The subset of transcripts expressed in each of the tissues, the maintenance transcripts, sorted by tissue type and abundance level are shown in Fig. 2. Most transcripts detected in any one cell type are detected at low levels, ≤5 copies per cell. The majority of the maintenance transcripts detected are at moderate levels, 10–50 copies per cell. This may suggest that most maintenance transcripts are produced in excess and regulation occurs during translation or protein modification and/or delivery. Alternatively, because we chose to study whole organs, in which the variety of transcripts produced could be substantially complex, we may not be detecting all of the low-abundance messages.

Comparison of fetal and adult expression. In a comparison of genes expressed in fetal vs. adult tissues, we found ~400 genes expressed in fetal tissues not detected in any of the adult tissues. These genes are of interest as candidate disease-causing genes when activated in adult tissues and include stem cell leukemia product (GenBank accession no. M63589), faciogenital dysplasia protein (U11690), N-ras (X02751), B-myb (X13293), protooncogene c-myc (HG3523), and transcript CH138 originally reported as isolated from stomach cancer cell lines (S77393).

In a study of only adult tissues, 695 transcripts are expressed in all 7 tissues, with a subset of 241 genes expressed at the same level; 333 of the genes vary in expression level by 5- to 10-fold. Forty genes expressed in all 7 tissues differ in transcript levels by greater than 19-fold, and of these, 8 differ by more than 50-fold, including COX7A muscle isoform (GenBank accession no. M83186) varying by 52-fold (highest in heart, lowest in kidney, pancreas, and testis), lectin (J04456) varying by 58-fold (highest in uterus, lowest in kidney and pancreas), myosin heavy chain (AF001548) varying by 61-fold (highest in uterus, lowest in brain and pancreas), elongation factor-1δ (Z21507) varying by 69-fold (highest in pancreas, lowest in lung and kidney), RNA polymerase II elongation protein (Z47087) varying by 70-fold (highest in brain, lowest in pancreas), extracellular mRNA for glutathione peroxidase (D00632) varying by 78-fold (highest in kidney, lowest in brain, pancreas, and testis), 14-9-9 protein η-chain (D78577) varying by 81-fold (highest in brain, lowest in testis), and L-arginine:glycine amidinotransferase (S68805) varying by 133-fold (highest in pancreas, lowest in heart and lung). A set of genes frequently used

[1] Supplemental material to this article (Tables 1–5) is available online (**http://physiolgenomics.physiology.org/cgi/content/full/2/3/143**).

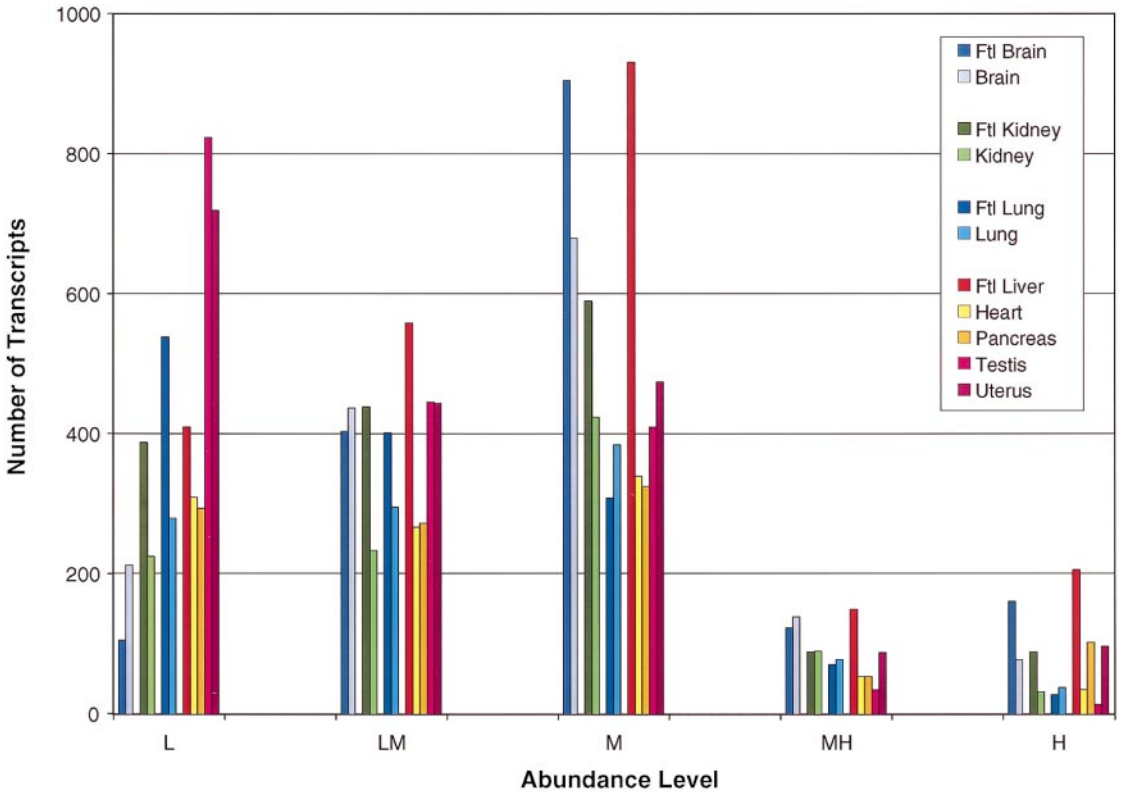

Fig. 1. Distribution of the expression levels for all of the transcripts detected in each tissue sorted by abundance level. The *x*-axis shows the range of abundance of the transcripts, binned from low to high; the *y*-axis shows the number of different transcripts in each bin. Abundance levels (L, low; M, moderate; H, high) in copies per cell are $L \leq 5$; $5 < LM \leq 10$; $10 < M \leq 50$; $50 < MH \leq 100$; $H > 100$. Ftl, fetal.

as controls in standard expression analysis were found to vary in expression level by 7- to 23-fold; these include β-actin (M10277) varying by 7-fold, with highest expression in brain and uterus and lowest expression in heart, and GAPDH (M33197) varying by 8-fold, with highest expression in brain, heart, and kidney and lowest in pancreas. Another form of β-actin (X00351) varies by 22-fold, with highest expression in uterus and lowest in pancreas. α-Actin (X13839) varies by 23-fold, and γ-actin (M19283) varies by 9-fold.

In fetal tissues, we found 767 transcripts expressed in all four tissues, 397 of which are expressed at the same level, 310 vary in expression level by 5- to 10-fold, 45 vary by 11- to 19-fold, and 15 vary by more than 19-fold. (See Tables 3 and 4 of the supplementary material for a list of the 695 shared adult transcripts and the 768 shared fetal transcripts.)

Tissue-specific transcripts. In the same experiments, we identified genes expressed uniquely in each of the tissues. For instance, in adult heart there were 3 transcripts not detected in the other 10 tissues, muscle glycogen synthase (GenBank accession no. J04501), MLC-1V/Sb isoform (M24248), and cytokine inducible nuclear protein (X83703). Not surprisingly, we found the greatest number of uniquely expressed genes in fetal tissues. Transcript numbers for brain were lower than anticipated, probably due to the difficulty of obtaining nondegraded brain RNA and the complexity of whole brain tissue. (See Table 5 of the supplementary material for a list of the genes uniquely expressed in a comparison of the tissues.)

DISCUSSION

Maintenance genes, a biologically relevant name for housekeeping genes. What is a housekeeping gene? About 35 years ago, housekeeping genes were simply defined as those genes that are always expressed (31). Today, with a better understanding of cellular processes, the housekeeping genes are defined as those genes critical to the activities that must be carried out for successful completion of the cell cycle. They are genes that play a key role in the maintenance of every cell. In light of an improved understanding of this subset of genes (no dusting or vacuuming genes have been identified) it may be time to replace the term "housekeeping gene" with a term that is biologically relevant, "maintenance gene." We have identified a subset of 535 genes expressed in 11 major fetal and adult tissues. These genes are turned on early in fetal development and stay on. We also identified a set of 47 transcripts that are expressed at the same level in fetal and adult tissues. This set will serve as a useful quantitative internal control in studies of normal adult and fetal gene expression.

Determining biologically relevant differences in expression levels. What is a biologically relevant difference in expression level? From a functional perspective, pro-

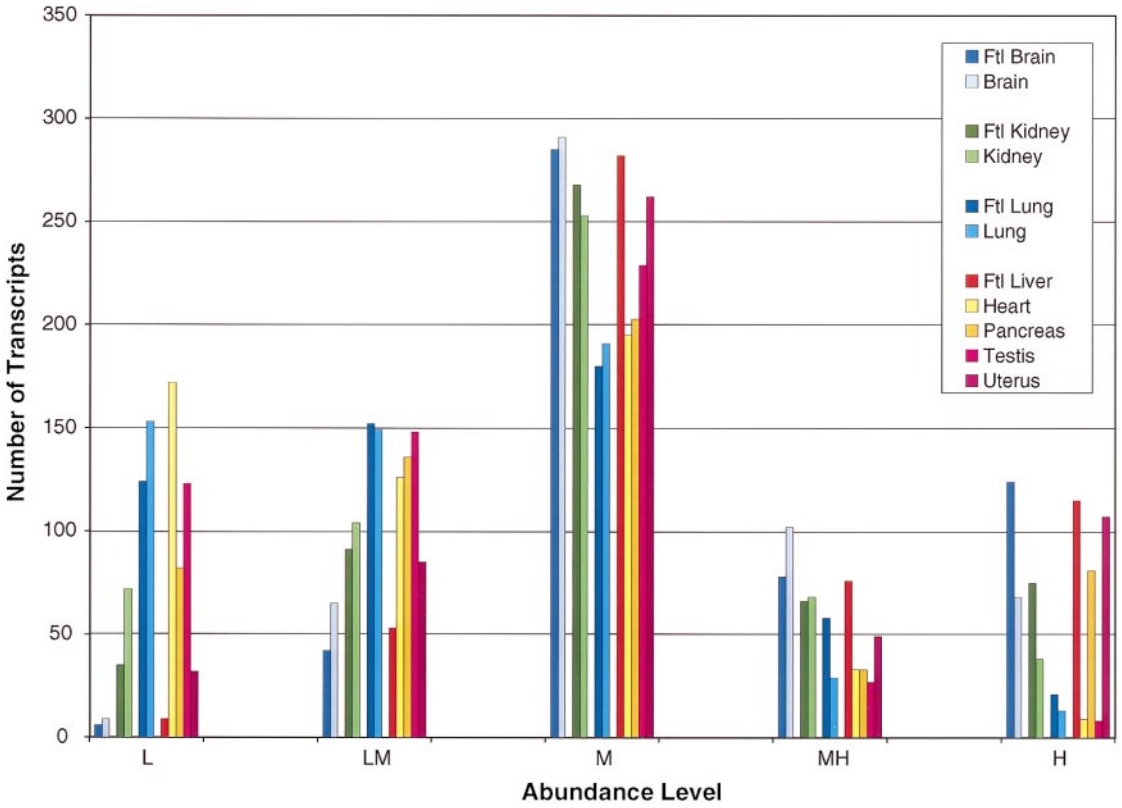

Fig. 2. Distribution of expression levels for the 535 transcripts detected in all 11 tissues (maintenance genes) sorted by abundance level. The *x*-axis shows the range of abundance of the transcripts, binned from low to high; the *y*-axis shows the number of different transcripts in each bin. Abundance levels in copies per cell are L $\leq$ 5; 5 $<$ LM $\leq$ 10; 10 $<$ M $\leq$ 50; 50 $<$ MH $\leq$ 100; H $>$ 100.

tein activity is the most critical measure of biological significance. We know that biological systems are complex and regulation of gene expression is quite removed from the ultimate destiny of the gene product, protein activity. Transcription, posttranscriptional modification, translation, posttranslational modification, and transport processes are not 100% efficient (1, 4, 11, 12, 20, 25, 29, 33). The cell must be able to tolerate and compensate for processing inefficiencies. The system must be flexible and, in most cases, probably produce an excess of transcript. Our data suggest that this may be the case for the maintenance genes. In a comparison of the abundance levels of all of the transcripts detected in all of the tissues with the abundance levels of the maintenance transcripts alone, we found that the majority of transcripts are expressed in low abundance, less than five copies per cell, whereas the maintenance transcripts are present in moderate levels, 10–50 copies per cell. Of course, some genes must be tightly controlled at the transcription step, but for the group of proteins responsible for basic cellular maintenance and survival, tight regulation at the transcription level is probably too risky. Studies of *Saccharomyces cerevisiae* and *S. pombe* support this line of reasoning. Twenty percent of genes in *S. cerevisiae* show noisy oscillations throughout the cell cycle, and in *S. pombe* it has been demonstrated that transcription is present in the absence of cell cycle progression and cellular concentrations of transcripts vary by two- to fourfold (3, 15). Here, we report genes as expressed at the same level if they are expressed in all 11 tissues at levels within fourfold. For most genes, differences less than fourfold are probably not biologically significant, but there is not sufficient data to conclude that a five- or sixfold difference is more biologically significant than a three- or fourfold difference (5, 14).

Until recently the technical challenge of accurately measuring small differences in gene expression has been practically insurmountable; consequently, there is little evidence to support the importance of small differences. For a subset of genes, it is likely that small differences have biological relevance, such as the genes encoding proteins that function differently when bound to high-affinity vs. low-affinity receptors or gene products triggering cellular cascades (2, 7, 18, 26). What is a biologically significant fold difference at the mRNA level? With so few data, it is difficult to know what a biologically significant difference in expression level is, but with the increase in sensitivity made possible by array technology and the development of other competing methods, we are surely about to find out.

We gratefully acknowledge L. Stryer and T. Gingeras for words of encouragement, S. Venkatapathy for technical assistance, S. Carter and A. Lau for assistance in manuscript preparation, and M. Durst for statistical assistance. We thank the people of Vicoforte, Italy, for their kindness and hospitality during the preparation of this manuscript.

Address for reprint requests and other correspondence: J. A. Warrington, Affymetrix, Inc., 3380 Central Expressway, Santa Clara, CA 95051 (E-mail: janet_warrington@affymetrix.com).

REFERENCES

1. **Ayoubi TA and Van De Ven W.** J regulation of gene expression by alternative promoters. *FASEB J* 10: 453–460, 1996.
2. **Chen WP, Chang YC, and Hsieh STJ.** Trophic interactions between sensory nerves and their targets. *J Biomed Sci* 6: 79–85, 1999.
3. **Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrielian AE, Landsman D, Lockhart DJ, and Davis RW.** A genome-wide transcriptional analysis of the miotic cell cycle. *Mol Cell* 2: 65–73, 1998.
4. **Conaway JW, Bradsher JW, Tan S, and Conway RC.** Transcription factor SIII. A novel component of the RNA polymerase ii elongation complex. *Cell Mol Biol Res* 39: 323–329, 1993.
5. **Creanor J and Mitchinson JM.** Nucleoside diphosphokinase, an enzyme with step changes in activity during the cell cycle of the fission yeast *Schizosaccharomyces pombe. J Cell Sci* 207–215, 1986.
6. **De Saizieu A, Certa U, Warrington JA, Gray C, Keck W, and Mous J.** Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays. *Nat Biotechnol* 16: 45–48, 1998.
7. **Dirks RP and Bloemers HP.** Signals controlling the expression of PDGF. *Mol Biol Rep* 22: 1–24, 1995–96.
8. **Duhig T, Ruhrberg C, Mor O, and Fried M.** The human surfeit locus. *Genomics* 52: 72–78, 1998.
9. **Fodor SPA, Rava RP, Huang XC, Pease AC, Holmes CP, and Adams CL.** Multiplexed biochemical assays with biological chips. *Science* 364: 555–556, 1993.
10. **Fodor SPA, Read JL, Pirrung MC, Stryer L, Lu AT, and Solas D.** Light-directed, spatially addressable parallel chemical synthesis. *Science* 251: 713–844, 1991.
11. **Haddad MM, Xu W, and Medrano EE.** Aging in epidermal melanocytes: cell cycle genes and melanins. *J Invest Dermatol Symp Proc* 3: 36–40, 1998.
12. **Hampsey M.** Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* 62: 465–503, 1998.
13. **Kagawa Y and Ohta S.** Regulation of mitochondrial ATP synthesis in mammalian cells by transcriptional control. *Int J Biochem* 22: 219–229, 1990.
14. **Klevecz RR.** *The Scientist* 13: 22–24, 1999.
15. **Klevecz RR, Kaufman SA, and Shymko RM.** Cellular clocks and oscillators. *Int Rev Cytol* 86: 97–128, 1984.
16. **Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, and Brown EL.** Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14: 131675–1680, 1996.
17. **May BK, Bhasker CR, and Cox TC.** Molecular regulation of 5-amniolevulinate synthase diseases related to heme biosynthesis. *Mol Biol Med* 7: 405–421, 1990.
18. **Merchav S.** The haematopoietic effects of growth hormone and insulin-like growth factor-I. *J Pediatr Endocrinol Metab* 11: 677–685, 1998.
19. **Milner CM and Campbell RD.** Genes, genes and more genes in the human major histocompatibility complex. *Bioessays* 14: 565–571, 1992.
20. **Nakamura YJ, Koyama K, and Matsushima M.** VNTR sequences as transcriptional, translational, or functional regulators. *Hum Genet* 43: 36–40, 1998.
21. **Rifkind RA, Marks PA, Bank A, Terada M, Maniatis GM, Reuben FE, and Fibach E.** Erythroid differentiation and the cell cycle: some implications from murine fetal and erythroleukemic cells. *Ann Immunol* 127: 887–893, 1976.
22. **Roberston HA.** Immediate-early genes, neuronal plasticity, and memory. *Biochem Cell Biol* 70: 729–737, 1992.
23. **Russo-Marie F.** Macrophages and the glucocorticoids. *J Neuroimmunol* 40: 281–286, 1992.
24. **Schena M, Shalon D, Davis RW, and Brown PO.** Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467–470, 1995.
25. **Shilatifard A.** The RNA polymerase II general elongation complex. *Biol Chem* 379: 27–31, 1998.
26. **Skerry TM.** Identification of novel signaling pathways during functional adaption of the skeleton to mechanical loading: the role of glutamate as a paracrine signaling agent in the skeleton. *J Bone Miner Metab (Japan)*17: 66–70, 1999.
27. **Strehler BL and Freeman MR.** Randomness, redundancy and repair: roles and relevance to biological aging. *Mech Aging Dev* 14: 15–38, 1980.
28. **Takahashi N, Hashida H, Zhao N, Misumi Y, and Sakaki Y.** High-density cDNA filter analysis of the expression profiles of the genes preferentially expressed in human brain. *Gene* 164: 219–227, 1995.
29. **Tate WP, Pool ES, Dalphin ME, Major LL, Crawford DJ, and Mannering SA.** The translational stop signal: codon with a context, or extended factor recognition element? *Biochimie* 78: 945–952, 1996.
30. **Velculescu VE, Zhang L, Vogelstein B, and Kinzler KW.** Serial analysis of gene expression. *Science* 270: 484–487, 1995.
31. **Watson JD, Hopkins NH, Roberts JW, Steitz JA, and Weiner AM.** The functioning of higher eucaryotic genes. In: *Molecular Biology of the Gene*, 1965, vol. 1, chapt. 21, p. 704.
32. **Wodicka L, Dong H, Mittman M, Ho MH, and Lockhart DJ.** Genome-wide expression monitoring in *Saccharomyces cerevisiae. Nat Biotechnol* 15: 1359–1367, 1997.
33. **Wolffe AP and Meric F.** Coupling transcription to translation: a novel site for the regulation of eukaryotic gene expression. *Int J Biochem Cell Biol* 28: 247–257, 1996.
34. **Yamamoto T, Matsui Y, Natori S, and Obinata M.** Cloning of a housekeeping-type gene (MER5) preferentially expressed in murine erythroleukemia cells. *Gene* 80: 337–343, 1989.